



Brasil

Alfabetizado: experiência de campo de 2004

Série Avaliação nº 2



ipea

Ministério
da Educação





Brasil Alfabetizado: experiência de campo de 2004

Organização:
Ricardo Henriques
Ricardo Paes de Barros
João Pedro Azevedo

Brasília, dezembro de 2006

Presidente da República

Luis Inácio Lula da Silva

Ministro da Educação

Fernando Haddad

Secretário-Executivo

José Henrique Paim Fernandes

Secretário de Educação Continuada, Alfabetização e Diversidade

Ricardo Henriques

Presidente do Instituto de Pesquisa Econômica Aplicada – IPEA

Luiz Henrique Proença Soares

Chefe de Gabinete

Persio Marco Antonio Davison

Diretoria de Estudos Macroeconômicos

Paulo Mansur Levy

SECAD – Secretaria de Educação Continuada, Alfabetização e Diversidade

Esplanada dos Ministérios, BI L, sala 700

Brasília, DF, CEP: 70097-900

Tel: (55 61) 2104-8432

Fax: (55 61) 2104-8476

Organização das Nações Unidas para a Educação, a Ciência e a Cultura

Representação no Brasil

SAS, Quadra 5, Bloco H, Lote 6, Ed. CNPq/IBICT/UNESCO, 9º andar

70070-914 – Brasília - DF – Brasil

Tel.: (55 61) 2106-3500

Fax: (5561) 3322-4261

Site: www.unesco.org.br

E-mail: grupoeditorial@unesco.org

IPEA – Instituto de Pesquisa Econômica Aplicada

IPEA – Brasília

SBS, Quadra 1, Bloco J, Ed. BNDES

70076-900 – Brasília - DF – Brasil

IPEA – Rio de Janeiro

Av. Presidente Antônio Carlos, 51

20020-010 – Rio de Janeiro - RJ – Brasil



Brasil Alfabetizado: experiência de campo de 2004

Organização:
Ricardo Henriques
Ricardo Paes de Barros
João Pedro Azevedo



ipea

Ministério
da Educação



© 2006. Secretaria de Educação Continuada, Alfabetização e Diversidade (Secad), Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco) e Instituto de Pesquisa Econômica Aplicada (IPEA)

Conselho Editorial da Coleção Educação para Todos

Adama Ouane

Alberto Melo

Célio da Cunha

Dalila Shepard

Osmar Fávero

Ricardo Henriques

Coordenação Editorial

Felipe Leitão Valadares Roquete

Revisão

Angélica Torres

Diagramação

ITECH – Instituto de Tecnologia do Habitat

Tiragem: 5.000 exemplares

Dados Internacionais de Catalogação na Publicação (CIP)

Brasil alfabetizado : experiência de campo de 2004 / organização, Ricardo Henriques, Ricardo Paes de Barros, João Pedro Azevedo. – Brasília : Secretaria de Educação Continuada, Alfabetização e Diversidade, 2006.

148 p. : il. – (Coleção Educação para Todos, Série Avaliação ; n. 2, v. 19)

ISBN 85-98171-63-8

1. Alfabetização. 2. Avaliação da aprendizagem. 3. Desenvolvimento congitivo. I. Henriques, Ricardo. II. Barros, Ricardo Paes de. III. Azevedo, João Pedro. IV. Secretaria de Educação Continuada, Alfabetização e Diversidade. V. Organização das Nações Unidas para a Educação, a Ciência e a Cultura. VI. Instituto de Pesquisa Econômica Aplicada. VII. Série.

CDU 372.415(81)

Os autores são responsáveis pela escolha e apresentação dos fatos contidos nesse livro, bem como pelas opiniões nele expressas, que não são necessariamente as da UNESCO e do Ministério da Educação, nem comprometem a Organização e o Ministério. As indicações de nomes e a apresentação do material ao longo deste livro não implicam a manifestação de qualquer opinião por parte da UNESCO e do Ministério da Educação a respeito da condição jurídica de qualquer país, território, cidade, região ou de suas autoridades, nem tampouco a delimitação de suas fronteiras ou limites.

Apresentação

A efetividade das políticas públicas sociais, entendida como o impacto produzido na vida dos beneficiários pelo investimento de recursos públicos na área social, depende de fatores como a eficiência na gestão dos programas, a eficácia das ações desenvolvidas e o quanto se está garantindo de atendimento às reais necessidades do público-alvo.

Por essa razão, o desenho adequado de um programa social, orientado para garantir o máximo possível de efetividade, depende primordialmente, além da identificação precisa do fenômeno sobre o qual se quer atuar, de três variáveis: i) elaboração de diagnóstico detalhado das causas desse fenômeno; ii) construção de conhecimento profundo tanto dos impactos gerados por programas similares aplicados a outras populações quanto das boas práticas de implementação e gestão; e iii) desenvolvimento de sistema de monitoramento e avaliação permanente, que oriente o aperfeiçoamento sistemático do desenho do programa.

Partindo desse entendimento, a fim de contribuir para o fortalecimento de uma cultura institucional comprometida com a avaliação das políticas públicas, foi instituído, no âmbito da Secretaria de Educação Continuada, Alfabetização e Diversidade do Ministério da Educação (Secad/MEC), o Departamento de Avaliação e Informações Educacionais (DAIE).

As avaliações levadas a cabo por esse Departamento têm por objetivo subsidiar os processos de aperfeiçoamento e redesenho dos programas. Para tanto, a utilização de metodologias quantitativas (visando refinar e organizar os objetivos específicos e as ações em indicadores mensuráveis) e de abordagens qualitativas (buscando delinear o contexto dos programas e a situação dos beneficiários) instrumentalizam as equipes das áreas finalísticas da Secad/MEC, permitindo, além disso, que se garanta a qualidade das ações, com utilização mais eficiente dos recursos disponíveis.

Essa forma inovadora de fazer política torna mais criterioso o processo de revisão das ações implementadas a cada ano e proporciona aos gestores e agentes uma oportunidade para repensarem dinamicamente suas atuações, frente aos objetivos postos pelos diferentes programas.

Além de dar transparência e compartilhar as experiências na área de avaliação obtidas pela Secad/MEC, a *Série Avaliação* é uma linha editorial concebida para dialogar com as demais obras da *Coleção Educação para Todos*, que problematizam, do ponto de vista conceitual, a realidade que se pretende alterar por meio de programas e ações. Espera-se, ainda, disseminar as metodologias utilizadas e promover o debate sobre sua adequação aos programas implementados, estimulando o leque de sugestões disponíveis para o aprimoramento tanto das avaliações realizadas pela Secad/MEC quanto daquelas realizadas por outras instituições.

Finalmente, a divulgação dos resultados alcançados, assim como das críticas construtivas feitas aos programas, contribui para um maior grau de controle social e indica o caminho a ser trilhado para a melhoria permanente das políticas públicas. Espera-se que essas publicações sejam úteis não apenas aos gestores dos programas e projetos analisados como também a todos os que se debruçam sobre iniciativas – governamentais ou da sociedade civil – voltadas para a transformação da realidade brasileira, em direção a um País de todos e para todos, com igualdade e equidade.

Ricardo Henriques

Secretário de Educação Continuada,
Alfabetização e Diversidade

Sumário

Introdução	11
------------------	----

Experiência do pré-teste: um processo de aprendizagem

Matriz de Referência:

proposta para o *Programa Brasil Alfabetizado* – Língua escrita

Antônio Augusto Gomes Batista, Ceris S. Ribas da Silva, Maria Lúcia Castanheira e Vera Masagão	17
Pressupostos	18
A aplicação do teste	20
Matriz de referência	21
Referências	30

Matriz de Referência:

proposta para o *Programa Brasil Alfabetizado* – Matemática

Maria da Conceição Ferreira Reis Fonseca, Maria da Penha Lopes, Maria Laura Magalhães Gomes e Vera Masagão	31
Pressupostos teóricos e suas implicações para a elaboração da matriz de avaliação	32
Matriz de referência	37

Síntese dos Relatórios de Aplicação dos Instrumentos de Avaliação Cognitiva: conhecimentos matemáticos, língua escrita e perfil socioeconômico

Maria Madalena Torres	43
-----------------------------	----

Qualidade psicométrica dos itens da avaliação educacional dos alunos egressos do *Programa Brasil Alfabetizado*

Anamara Ferreira Ribeiro e Josemberg M. de Andrade	49
Método	51
Resultados	52
Considerações Finais	59
Referências	60

Desafios da amostragem

Estudo sobre o desenho da amostra de alunos e turmas do Programa Brasil Alfabetizado

Mauricio Teixeira Leite de Vasconcellos	63
Objetivos da amostra do SBA 2004	64
Desenho da amostra do SBA 2004.....	64
Descrição do arquivo da amostra do SBA 2004.....	67

Estudo sobre o desenho da amostra de alunos e turmas do Programa Brasil Alfabetizado

Mauricio Teixeira Leite de Vasconcellos	71
Introdução	71
Objetivos da amostra do SBA 2005	72
População alvo e população de pesquisa	72
Estratificação a alocação do tamanho da amostra do SBA 2005	72
Seleção da amostra do SBA 2005	76
Descrição do arquivo da amostra do SBA 2005	81
Referências	88
Anexos	89

Estudo para estratificação e desenho da amostra utilizada para avaliação do impacto dos projetos aprovados no âmbito da Resolução de 2005 do Programa Brasil Alfabetizado

Mauricio Teixeira Leite de Vasconcellos	95
Introdução	95
Geração e exploração do cadastro de seleção	95
Objetivos, estratificação a alocação da amostra do SBA 2005.....	108
Referências	113

**Estudo sobre o desenho da amostra para análise
da demanda por cursos de Alfabetização
de Jovens e Adultos**

Pedro Luis do Nascimento Silva, Mauricio Teixeira Leite de Vasconcellos e Nuno Duarte da Costa Bittencourt	115
Introdução	115
Desenho da amostra da AIBF	116
Cadastros utilizados	117
Referências	128

Anexo I

Nota técnica 3: reflexões sobre a amostra para o componente da demanda do <i>Programa Brasil Alfabetizado</i>	129
--	------------

Anexo II

Quadros e Tabelas	135
--------------------------------	------------

Anexo III

Algoritmo para amostragem seqüencial de Poisson	143
--	------------

Introdução

O Plano de Avaliação do *Programa Brasil Alfabetizado* – cujos princípios, desenhos e desdobramentos espriam-se pelos diversos volumes dessa Série Avaliação – é o resultado de um longo e enriquecedor processo levado a cabo pela Secretaria de Educação Continuada, Alfabetização e Diversidade (Secad). Construção coletiva, fruto da contribuição de uma miríade de atores e parceiros, o Plano de Avaliação, para além de seus objetivos imediatos de avaliar as diversas dimensões do *Programa Brasil Alfabetizado*, representou uma oportunidade ímpar de aprendizado organizacional e institucional.

Sua gênese pode ser encontrada na confluência de intenções do Departamento de Avaliação e Informações Educacionais (DAIE) e do Departamento de Educação de Jovens e Adultos (DEJA), que buscavam aferir – então, cada um a seu modo e fundados em princípios idiossincráticos – os resultados alcançados pelo *Programa Brasil Alfabetizado*. O surgimento dessas demandas por iniciativas de avaliação – naturais quando se considera a busca dos gestores pelo real impacto das ações que empreendem, mas incomuns na prática cotidiana da administração pública – fossem elas destinadas à avaliação da gestão, da efetividade do Programa ou do desempenho cognitivo dos alfabetizandos, mobilizou as equipes de ambos Departamentos, em um longo e produtivo debate sobre as diversas alternativas disponíveis aos gestores do Programa.

No segundo semestre de 2004, a Secad colheu o primeiro fruto do processo de avaliação do *Programa Brasil Alfabetizado*, expressão transparente das contradições, embates e divergências – que não prescindiram da lealdade e do respeito à diferença – que fundamentaram a concepção das primeiras iniciativas concretas de avaliação. Ainda que a visão sobre os princípios primevos de uma política pública de alfabetização de jovens e adultos pudesse, aos olhos de observadores desavisados, parecer irreconciliável, importava o fato de haver compartilhamento efetivo em relação aos princípios metodológicos que embasavam a construção e implementação das iniciativas de avaliação.

Deflagram-se, então, duas frentes simultâneas de trabalho. A primeira, em parceria com o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio

Teixeira (Inep), visando à elaboração de itens de avaliação que comporiam, utilizando a mesma metodologia das provas do Sistema de Avaliação Básica (Saeb), um teste cognitivo exclusivamente voltado à avaliação do desempenho cognitivo, em Leitura/Escreita e Matemática, de alfabetizando jovens e adultos. Nesse ponto, cabe ressaltar o ineditismo desta iniciativa: a construção de itens, formando um “banco de itens” de caráter público, que permitissem aferir as competências e habilidades que jovens e adultos – então excluídos da escola e, muitas vezes, sem trajetória escolar prévia – desenvolveram durante o curso de alfabetização. A concepção metodológica, bem como a elaboração estrito senso dos itens de avaliação de Leitura/Escreita e Matemática, ficaram sob a responsabilidade da equipe do Centro de Alfabetização, Leitura e Escrita (Ceale), vinculado à Faculdade de Educação da Universidade Federal de Minas Gerais (FaE/UFMG), instituição com reconhecida experiência na seara da avaliação, pedagógica e cognitiva, de políticas e projetos de alfabetização de jovens e adultos.

A segunda frente, empreendida em parceria com a Sociedade Científica da Escola Nacional de Ciências Estatísticas (Science) e o Instituto de Pesquisa Econômica Aplicada (Ipea), visou, primordialmente, ao desenho das amostras de alfabetizando e turmas, bem como à construção de um instrumento de pesquisa que possibilitasse colher informações sobre o perfil socioeconômico dos beneficiários do *Programa Brasil Alfabetizado*, abarcando dimensões relacionadas à motivação, histórico familiar, educação, renda, condições habitacionais, usos do tempo, saúde preventiva e acesso ao mercado de trabalho.

Dessa forma, em dezembro de 2004, o processo de avaliação do *Programa Brasil Alfabetizado* deu seus primeiros passos – expressão translúcida da parceria, ora consolidada, entre o DAIE e o DEJA – rumo à implementação de iniciativas concretas para aferição do desempenho cognitivo dos alfabetizando e, ademais, para coleta de informações socioeconômicas sobre o público-atendido. O presente volume apresenta parte fundamental dessa trajetória.

Os dois primeiros capítulos da Seção I, “Experiência do pré-teste: processo de aprendizagem”, apresentam as Matrizes de Referência que fundamentaram a construção dos testes cognitivos de Leitura/Escreita, de um lado, e Matemática, de outro, ambos aplicados – em dezembro de 2004 – junto aos alfabetizando jovens e adultos atendidos pelo *Programa Brasil Alfabetizado*. Partindo dos pressupostos teóricos que embasaram as concepções de “aprendizado de Leitura/Escreita” e de “aprendizado de Matemática”, os textos voltam-se para a descrição dos conhecimentos e competências que serão possíveis avaliar por intermédio dos itens que

compõem os testes cognitivos. De forma detalhada e exaustiva, são apresentadas as instrumentalizações de cada uma das competências (descritores) e a exigência relativa de cada item (nível de dificuldade). Tal metodologia permitia aferir, por meio da aplicação da Teoria de Resposta ao Item (TRI), o nível de proficiência alcançado por cada um dos alfabetizandos avaliados: desde os níveis elementares de aprendizado da leitura, passando pela aquisição das competências de decifração e fluência em leitura e desenvolvimento da habilidade de codificação, para, ao final, alcançar a utilização plena dessas habilidades e competências nos procedimentos básicos de leitura e escrita. Detalhe de fundamental importância, cabe-nos ressaltar, refere-se à forma de aplicação de ambos os testes, Leitura/Escrita e Matemática: a equipe do Ceale optou pela *aplicação individual*. Tal opção metodológica – cujos impactos sobre a logística de campo serão abordados em capítulo posterior – implicava a necessidade de o entrevistador aplicar o teste, em separado, junto a cada um dos alfabetizandos, utilizando procedimentos que concentravam – na maioria das questões, excetuando-se aquelas relacionadas à escrita de palavras ou números – várias tarefas nas mãos do aplicador (leitura em voz alta dos comandos, apresentação das opções de resposta em um cartaz e preenchimento das respostas) e deixava aos alfabetizandos apenas a função de ditar ou apontar uma das alternativas de resposta.

O capítulo seguinte traz um breve relato da experiência de campo, listando as principais percepções relacionadas à aplicação, em dezembro de 2004, dos testes cognitivos e do instrumento socioeconômico. Ademais, é perceptível a louvável intenção de apresentar recomendações de melhorias nos instrumentos utilizados, visando à continuidade do processo de avaliação.

Finalmente, o quarto capítulo apresenta-nos a primeira tentativa de tratamento dos dados oriundos da correção dos testes cognitivos de Leitura/Escrita e Matemática. A partir da aplicação de técnica psicométrica – a Teoria Clássica dos Testes – foram verificados os parâmetros de “discriminação” e “dificuldade” de todos os itens que compunham os testes cognitivos, permitindo que, com base nos resultados desta análise, fossem iniciados os procedimentos para aperfeiçoamento dos itens e redesenho dos testes.

A Seção II, “Desafios da Amostragem”, compõe-se de quatro capítulos que abordam estudos, desenvolvidos pela equipe da Science, relacionados à estratificação e desenho de amostras –tanto para os alfabetizandos e turmas do *Programa Brasil Alfabetizado*, quanto para uma pesquisa exploratória que buscava estimar a demanda por cursos de alfabetização.

Ainda que, em alguns momentos, os textos possam parecer incipientes aos olhos de especialistas e pesquisadores mais exigentes, é possível percebermos nas entrelinhas – principalmente aqueles atores e parceiros que participaram ativamente da elaboração do Plano de Avaliação do *Programa Brasil Alfabetizado* – a germinação de um enriquecedor processo de aprendizagem e construção coletivas. Os primeiros frutos, ainda que temporãos, foram colhidos em dezembro de 2004: e são eles que lhes oferecemos agora.

Timothy Denis Ireland

Diretor do Departamento de Educação
de Jovens e Adultos da Secad/MEC

Experiência
do pré-teste:
um processo de
aprendizagem

Matriz de Referência: proposta para o *Programa Brasil Alfabetizado* – Língua escrita

Antônio Augusto Gomes Batista
Ceris S. Ribas da Silva
Maria Lúcia Castanheira¹
Vera Masagão²

Este documento apresenta diretrizes para a avaliação de competências relacionadas ao domínio da língua escrita de alfabetizandos do *Programa Brasil Alfabetizado*. As diretrizes foram organizadas com base nas *Orientações Gerais do Programa* (MEC, 2004), na Resolução nº 9 do FNDE, bem como em pressupostos sobre o aprendizado da língua escrita, especialmente por jovens e adultos, público-alvo do *Programa Brasil Alfabetizado*.

A avaliação, tendo em vista sua escala, privilegiará, no aprendizado da língua escrita, o desenvolvimento de competências e habilidades de *leitura*, embora (i) apresente também condições para a medição de competências mais gerais, relacionadas tanto ao domínio da leitura quanto da escrita, e (ii) permita a mensuração – ainda que de um número reduzido – de competências ligadas ao processo de escrita.

Apresentam-se, a seguir, pressupostos que sustentaram a elaboração do documento, condições que devem ser levadas em conta na aplicação de um instrumento de avaliação para alfabetizandos e, por fim, uma matriz de referência para orientar a construção de itens de avaliação.

¹ Equipe do Centro de Alfabetização, Leitura e Escrita – Ceale/FaE/UFMG

² Coordenadora de programas da Ação Educacional

Pressupostos

Assume-se aqui uma concepção de aprendizado da língua escrita que envolve quatro grandes dimensões cognitivas:

- (i) o domínio de competências que tendem a contribuir para o *processo inicial de apropriação do sistema de escrita*, servindo de base tanto ao desenvolvimento do processo de leitura quanto ao de escrita;
- (ii) o desenvolvimento da capacidade de *decifração*, quer dizer, de transformar sinais gráficos ou grafemas em fonemas, com maior ou menor nível de fluência;
- (iii) o desenvolvimento da capacidade de escrita, especificamente as capacidades muito iniciais de escrever palavras memorizadas e codificar palavras simples;
- (iv) o desenvolvimento do processo de *compreensão* de textos;

Segundo distinção proposta por Magda Soares (2004), as capacidades relacionadas a (i), (ii) e (iii) designam apropriadamente a *alfabetização*; vale dizer, o aprendizado do sistema de escrita alfabético-ortográfico e o desenvolvimento de capacidades cognitivas e motoras envolvidas no uso e na manipulação de instrumentos e equipamento de escrita – por exemplo, no caso da leitura, decodificar textos, sentenças e palavras, ler com fluência, posicionar-se adequadamente para ler, saber como se faz a seqüenciação de um texto num determinado suporte, conhecer a organização gráfica da escrita na página, dentre outras.

Ainda de acordo com Magda Soares (2004), as capacidades relacionadas a (iv) tendem a designar habilidades de *uso* do sistema de escrita e de seus instrumentos e equipamentos em práticas sociais; vale dizer, na *compreensão* (e na produção) de textos em situações sociais, tendo em vista finalidades sociais. Para Magda Soares, essas capacidades designam, em parte, o letramento: as práticas sociais de uso da língua escrita, assim como as competências necessárias para participar dessas práticas.

Pressupõe-se que tanto o letramento quanto a alfabetização se caracterizam como um *contínuo*, que não pode ser reduzido – tendo em vista as características do Programa que se avalia, tratadas mais à frente – a dicotomias como alfabetizado/analfabeto ou letrado/iletrado.³ Isto ocorre porque os dois processos consistem de

“(...) um grande número de diferentes habilidades, competências cognitivas e metacognitivas, aplicadas a um vasto conjunto de materiais de leitura e gêneros de escrita, e refere-se a uma variedade de usos da leitura e da escrita, praticadas em contextos sociais diferentes.” (SOARES, 1998, p.107)

³ Embora, evidentemente, de acordo com o contexto de medição, seja necessário – e mesmo desejável – utilizar essa distinção, ainda que com certo grau de arbitrariedade. Cf. a respeito Soares (1998).

Para evitar a arbitrariedade do processo de discretização desse contínuo, ou de seu “desagregamento” (SCRIBNER, 1984, p.18), considera-se que são “os propósitos e o contexto da avaliação” (SOARES, 1998, p.109) os fatores que permitem estabelecer critérios com base nos quais promover essa discretização.

Considerando-se que:

- é objetivo geral da avaliação “gerar informações sobre as práticas pedagógicas e os resultados de aprendizagem gerados pelo *Programa Brasil Alfabetizado*, de modo a subsidiar a ação de todos os agentes diretamente envolvidos e/ou interessados no pleno cumprimento de suas metas” (RIBEIRO, 2003, p.2);
- constitui um dos objetivos específicos da avaliação “estabelecer parâmetros comuns para avaliar as aprendizagens realizadas pelos alfabetizados, de modo a possibilitar a análise comparada das diversas práticas e a identificação de fatores que potencializam os resultados” (RIBEIRO, 2003, p.2);
- as ações realizadas no âmbito do *Programa Brasil Alfabetizado* desenvolvem-se num curto período de tempo – de seis a oito meses – e são tomadas como pontos de partida para uma inclusão mais duradoura dos sujeito na escola;
- outros estudos (BATISTA & RIBEIRO, 2004) mostram que mais de 30% dos jovens e adultos que não concluíram a 4ª série do ensino fundamental estão em situação de analfabetismo e 50% deles não conseguem ultrapassar o nível mais elementar de alfabetismo funcional; pode-se esperar, desse modo, que os sujeitos atendidos pelo Programa tenham – em razão da duração de sua participação no Programa, de sua escolarização em geral restrita ou inexistente – níveis muito iniciais de leitura,

optou-se por desagregar o contínuo que caracteriza o domínio da língua escrita de modo a apreender:

- variações ou estágios em níveis iniciais ou elementares de aprendizado da leitura, enfatizando a emergent literacy⁴ – (i);

⁴ A expressão *emergent literacy* compreende competências cujo desenvolvimento tende a concorrer para a alfabetização, especificamente para o domínio do princípio alfabético, quer dizer, para a compreensão de que o sistema de notação representa sons ou fonemas, caracterizando-se, assim, como um sistema de escrita alfabético-ortográfico. Para uma abordagem equilibrada de comportamentos e competências que assinalam a “emergência” da alfabetização, ver Cunningham & Allington (2003, p.26-68).

- a alfabetização propriamente dita, aqui identificada, de modo privilegiado, como a aquisição das competências de decifração e fluência em leitura e, de modo secundário, como o desenvolvimento da habilidade de codificação – (ii) e (iii);
- a utilização dessas habilidades e competências tendo em vista procedimentos básicos de leitura (a compreensão de palavras; a extração de um número reduzido de informações; o uso de gêneros simples e próximos do cotidiano dos sujeitos atendidos pelo Programa⁵ – bilhetes mais ou menos simples; a assinatura, no caso da escrita) – (iv).

É também em razão dessas considerações que se optou por discretizar, em maior pormenor, as competências relacionadas à *alfabetização*, quer dizer, aquelas que tendem a colaborar para a emergência da alfabetização, e que identificam o domínio inicial da língua escrita, caracterizado pela decifração e pela aquisição de fluência em leitura.

A aplicação do teste

Tendo em vista as características dos sujeitos cujo rendimento será avaliado, é necessário que:

- as tarefas a serem realizadas sejam propostas oralmente por um entrevistador, que lerá as instruções e, a não ser no caso de itens que supõem a escrita ou são relacionados a competências mais avançadas, buscará obter uma resposta comunicada oralmente ou por gesto;
- tendo em vista a pouca familiaridade dos sujeitos com situações de testagem, seja evitada, na elaboração do teste e em sua aplicação, a utilização de processos discursivos característicos de gêneros do discurso escolar;
- os textos utilizados para leitura autônoma pelos sujeitos (com exceção das questões ligadas a descritores que visam apreender exatamente o domínio de diferentes tipos de letra) estejam disponíveis em letra de imprensa (apenas maiúscula; maiúscula e minúscula) e em cursiva (maiúscula e minúscula, evidentemente).⁶

⁵ Teria sido possível utilizar uma gama mais ampla de gêneros de textos, mas a acentuada dispersão regional dos alfabetizados atendidos pelo Programa e a grande presença de sujeitos pertencentes a meios rurais leva a supor um domínio, por parte do público do Programa, relativamente restrito de referências que ultrapassam seu universo cultural de origem. Ao lado da limitada duração do Programa, esse fator conduziu à redução desse espectro.

⁶ Infelizmente, não existe, no Brasil, um consenso a respeito do tipo de letra mais adequado a ser utilizado nos momentos iniciais da alfabetização. Embora muitos educadores estejam de acordo com o pressuposto de que a imprensa maiúscula seria o melhor tipo de letra a ser empregado nesses momentos, materiais e práticas tendem a fazer um número muito variado de escolhas no que diz respeito a esse aspecto.

Matriz de referência

Conhecimentos	Competências	Descritores	Detalhamento
<p>Características da tecnologia da escrita</p>	<p>C1. Dominar conhecimentos que concorrem para a apropriação da tecnologia de escrita</p>	<p>D1. Compreender o emprego de termos ligados à alfabetização</p>	<p>Trata-se de uma capacidade de natureza metacognitiva. Espera-se que o alfabetizando compreenda o <i>emprego</i> de termos utilizados no processo de alfabetização, como alfabeto, sílaba, letra, palavra, sentença ou frase, de tal modo que manifeste conseguir <i>operar</i> com eles. Evidentemente, não se trata de solicitar do alfabetizando definições desses termos. Por exemplo: num pequeno texto, solicitar aos alfabetizandos que marquem uma ou mais letras que conhecem; mostrem a primeira ou a última sílaba de uma palavra simples; mostrem onde termina e onde começa uma palavra e uma frase.</p>
		<p>D2. Identificar letras do alfabeto</p>	<p>O alfabetizando deve manifestar conhecer as letras do alfabeto. Pode-se solicitar que as recite (de cor ou com apoio escrito) ou que identifique determinadas letras em palavras ou numa seqüência de letras. Deve-se procurar alternar consoantes e vogais e atentar para os modos diferenciados de designação dos nomes das letras no País (“mê”, “nê”, por exemplo, em oposição a “eme” e “ene”)</p>

Conhecimentos	Competências	Descritores	Detalhamento
Características da tecnologia da escrita (cont.)	C1. Dominar conhecimentos que concorrem para a apropriação da tecnologia de escrita (cont.)	D3. Conhecer as direções da escrita	<p>Avalia-se aqui o domínio das diferentes direções da escrita, na linha (da esquerda para a direita), na página (do alto para baixo), num folheto, livro ou jornal (movimento de “abrir” um desses suportes, indo da direita para a esquerda). É importante, no caso da direção na linha, que se marque a escrita sempre em relação às margens (a sua esquerda, direita e entre as margens), pois é em relação a elas que o alfabetizando tende a definir a direção. Um exemplo: Para isso, uma atividade relevante consistiria numa tarefa de cópia de uma palavra em diferentes posições em relação à margem. Manipular um jornal, revista ou folheto pode dar base a questões sobre a direção desses suportes.</p>
		D4. Diferenciar letras de outros sinais gráficos, como os números, sinais de pontuação ou outros sistemas de representação	<p>Verificar se o alfabetizando distingue letras de números e de outros sinais gráficos, como pontuação, acentuação, aspas etc. É também importante verificar se distingue “escritas” de outras formas de representação (como o desenho, logomarcas, placas de trânsito, por exemplo)</p>
		D5. Identificar, ao ouvir uma palavra, o número de sílabas	<p>Trata-se de uma importante competência para a alfabetização, pois é a sílaba a principal unidade sobre a qual opera o alfabetizando. É importante que, ao pronunciar a palavra, o entrevistador não acentue excessivamente essas unidades</p>

(cont.)

Conhecimentos	Competências	Descritores	Detalhamento
Características da tecnologia da escrita (cont.)	C1. Dominar conhecimentos que concorrem para a apropriação da tecnologia de escrita (cont.)	D6. Identificar sons, sílabas e outras unidades sonoras	A consciência fonológica é considerada uma importante dimensão da alfabetização. As tarefas devem envolver a identificação de sons, sílabas e outras unidades idênticas ou semelhantes (as outras unidades devem ser maiores que a sílaba, mas menores que a palavra). Rimas e sílabas tendem a ser mais facilmente reconhecidas que sons isolados, assim como as vogais. Também tendo em vista uma gradação de dificuldade, deve-se levar em conta a posição da unidade sonora (mais facilmente identificadas, primeiramente, no início de palavra, depois no fim e, por último, no interior) e a tonicidade (as sílabas tônicas são mais perceptíveis que as átonas). Pequenos textos, como parlendas, cantigas, trava-línguas e poemas podem ser utilizados em razão do uso intensivo da rima e da aliteração.
		D7. Distinguir palavras escritas de simulações de escrita	Deseja-se avaliar se o alfabetizando estabeleceu critérios adequados a respeito do que é uma palavra escrita. Um exemplo de questão pode consistir na solicitação de que o alfabetizando indique, tendo em vista seqüências diferentes de letras, onde está escrito "alguma coisa": a) B b) AAAAA c) CVL d) CAVALO

Conhecimentos	Competências	Descritores	Detalhamento
Características da tecnologia da escrita (cont.)	C1. Dominar conhecimentos que concorrem para a apropriação da tecnologia de escrita (cont.)	D8. Distingue diferentes tipos de letra	Trata-se de uma competência bastante avançada, em razão da ausência de uma definição clara, entre alfabetizadores, a respeito do tipo de letra a ser utilizado na alfabetização. Certos alfabetizadores utilizam a cursiva e a imprensa simultaneamente (maiuscula e minúscula); outros utilizam apenas a de imprensa maiúscula (só mais tarde introduzindo a imprensa minúscula e a cursiva). A esse respeito, ver considerações na introdução sobre o tipo de letra a ser utilizado nos textos e palavras a serem lidos autonomamente pelos alfabetizandos.
Escrita	C2. Escrever palavras	D9. Escrever o nome próprio completo	A escrita do nome próprio (completo ou incompleto; com erros ou sem erros) já serviu de critério de alfabetização e representa, ainda hoje, um importante aspecto do processo de aprendizado da língua escrita (seja sob um ponto de vista simbólico, seja sob um ponto de vista pedagógico, já que o nome tende a servir como “forma fixa”, uma palavra que o alfabetizando sabe de cor e pode ser utilizada como instrumento para a análise e a escrita de palavras novas).

(cont.)

Conhecimentos	Competências	Descritores	Detalhamento
Escrita (cont.)	C2. Escrever palavras (cont.)	D10. Escrever palavras ditadas	<p>Pretende-se apreender a capacidade de codificação de palavras isoladas; aquelas mais frequentes no cotidiano dos alfabetizados (como as ligadas à esfera doméstica e às primeiras aprendizagens escolares) são mais fáceis; também são mais facilmente redigidas palavras com estrutura silábica mais simples e que exigem menos esforço físico e da memória de curto termo. O que se espera, nas questões, é apreender se o alfabetizando domina o princípio alfabético (isto é: que o sistema de escrita representa sons ou fonemas) e o grau de domínio das relações entre fonemas e grafemas no português. Desse modo, podem ser obtidos diferentes níveis: o alfabetizando não domina o princípio alfabético; domina o princípio alfabético sem fazer erros; domina fazendo erros. Para não dificultar a análise das respostas aos itens relativos ao descritor, é importante que se evitem palavras que possam gerar escritas incompreensíveis para pessoas pouco familiarizadas com a escrita inicial de alfabetizando, como, por exemplo, “piolho”, que pode ser redigida como, também, por exemplo, “pionho” ou “bioio”. Isto é importante porque, na codificação de respostas, a possibilidade de ler o que foi escrito pelo alfabetizando, mesmo que com muitos erros, será empregada como critério para apreender se este domina ou não o princípio alfabético.</p>

(cont.)

Conhecimentos	Competências	Descritores	Detalhamento
Decifração	C3. Decifrar	D11. Ler palavras em voz alta	<p>Pretende-se, aqui, verificar a competência de decodificar uma palavra. Aspectos ligados à compreensão e a outros processos de decifração (reconhecimento global e instantâneo; levantamento de hipóteses, por exemplo) são explorados por outros descritores. É mais difícil a leitura de palavras com padrão silábico diferente do padrão CV (consoante + vogal) e apresentadas fora de um contexto (ou em contexto pouco familiar – o que mostra a necessidade de controlar os contextos explorados).</p> <p>Durante a leitura, pode-se avaliar se a decodificação é feita com ou sem fluência (a ausência de fluência é marcada por lentidão, por hesitações e pausas e pelo acesso ao sentido da palavra apenas depois da decifração).</p>

(cont.)

Conhecimentos	Competências	Descritores	Detalhamento
Decifração (cont.)	C3. Decifrar (cont.)	D12. Ler em voz alta uma sentença	<p>Pretende-se, aqui, verificar a competência de decodificar uma sentença. Aspectos ligados à compreensão e a outros processos de decifração (reconhecimento global e instantâneo; levantamento de hipóteses, por exemplo) são explorados por outros descritores. É mais difícil a leitura de sentenças sem contexto (ver descritor acima) e com padrão sintático diferente da ordem canônica sujeito + verbo + complementos (por exemplo: posposição de sujeito, deslocamento à esquerda de adjuntos adverbiais, termos intercalados, voz passiva). Durante a leitura, pode-se avaliar se a decodificação é feita com ou sem fluência (a ausência de fluência é marcada por lentidão, por hesitações e pausas e pelo acesso ao sentido da palavra apenas depois da decifração de todas as palavras que compõem a sentença).</p>
		D13. Ler em voz alta um bilhete simples	<p>Avalia-se aqui a fluência em leitura. A declaração acerca da capacidade de ler ou escrever um bilhete simples é utilizada para definir, no Censo, a condição de alfabetizado. A leitura fluente é caracterizada pela automatização do processo de decodificação, o que conduz à inexistência ou à presença reduzida de hesitações, pausas, lentidão excessiva. Para que o bilhete seja simples, é necessário que contenha apenas uma informação central (como, por exemplo: 21/10/2004 Maria, o seu almoço está pronto no fogão. Carlos) e que o padrão sintático seja o canônico.</p>

Conhecimentos	Competências	Descritores	Detalhamento
Compreensão	C4. Recuperar informações	D14. Compreender palavras lidas silenciosamente	<p>O objetivo desse descritor é apreender a capacidade de compreender, por meio da leitura silenciosa e mediante diferentes tipos de processamento da leitura (decifração, reconhecimento global e instantâneo, formulação de hipóteses, apoio em saliências textuais e em contextos, por exemplo). A palavra deve ser apresentada num contexto, já que se pretende avaliar – seja qual for o processamento da leitura – a capacidade de compreensão, da qual fazem parte os elementos que compõem o contexto. Por exemplo: mostre, no desenho, onde se localiza a padaria; o alfabetizando deve marcar, no desenho, a casa com a placa de padaria.</p>
		D15. Retirar uma informação explícita de uma sentença.	<p>O alfabetizando deve extrair uma informação de uma sentença mais longa (aqueles que não desenvolveram automatismos de leitura terão mais dificuldade, em razão de sobrecarregarem sua memória de curto termo) e mais complexa (com padrões sintáticos não canônicos).</p>

(cont.)

Conhecimentos	Competências	Descritores	Detalhamento
Compreensão (cont.)	C4. Recuperar informações (cont.)	D16. Retirar informação explícita de um bilhete	<p>O alfabetizando deve retirar uma informação num bilhete que contém mais de uma informação (22/10/2004 Maria, Tudo bem? Seu almoço está no fogão. Volto amanhã à tarde. Não se esqueça de pagar a conta de luz. Abraço, Carlos).</p> <p>Para aumentar o nível de dificuldade, pode-se solicitar que o alfabetizando retire mais de uma informação.</p>

Referências

- CUNNINGHAM, Patricia M. e ALLINGTON, Richard L. *Classrooms That Work: they can all read and write*. 3.ed. Boston, MA: Pearson Education, 2003.
- SOARES, Magda Becker. *Letramento e alfabetização: as muitas facetas*. 26ª Reunião Anual da ANPEd. GT Alfabetização, Leitura e Escrita. Poços de Caldas, 7 de outubro de 2003.
- SOARES, Magda Becker. “Letramento: como definir, como avaliar, como medir”. In: _____. *Letramento: um tema em três gêneros*. Belo Horizonte: Autêntica, 1998. p.61-125. (Original em inglês: *Literacy Assessment and its implications for Statistical Measurement*. Paris: Unesco, 1992).
- RIBEIRO, Vera Masagão. *Termo de referência para avaliação do Programa Brasil Alfabetizado*. Brasília: Secad/MEC, 2003.
- BATISTA, Antônio Augusto Gomes e RIBEIRO, Vera Masagão. “Cultura escrita no Brasil: modos e condições de inserção”. *Educação e Realidade*. Porto Alegre, 2004 (no prelo).
- BRASIL. *Orientações Gerais do Programa Brasil Alfabetizado*. Brasília: MEC, 2004.
- FNDE. Resolução n. 9 (Manual de orientações para assistência financeira a programas e projetos e projetos educacionais). www.mec.gov.br. (acessado em 16/10/2004).

Matriz de Referência: proposta para o *Programa Brasil Alfabetizado* – Matemática

Maria da Conceição Ferreira
Reis Fonseca¹
Maria da Penha Lopes²
Maria Laura Magalhães Gomes³
Vera Masagão⁴

A proposição de uma matriz de referência para a avaliação de Matemática no *Programa Brasil Alfabetizado* requer algumas considerações sobre o desenvolvimento de competências e conhecimentos matemáticos no contexto desse Programa.

A preocupação com a incorporação do domínio de algumas habilidades matemáticas à concepção de alfabetismo indica, primeiramente, que o Programa assume uma perspectiva mais ampla da *alfabetização*. Leva-se em conta que, para ser alfabetizado hoje, no Brasil, o sujeito precisa ser capaz de atender demandas de leitura e escrita cada vez mais diversificadas e sofisticadas, que solicitam, não raro, a mobilização de conhecimentos e habilidades matemáticos.

Reconhece-se, assim, o papel da educação matemática na promoção do acesso e do desenvolvimento de estratégias e possibilidades de leitura do mundo, para as quais conceitos, relações, critérios, procedimentos e resultados matemáticos podem contribuir significativamente. Assume-se, portanto, uma abordagem que contempla

¹ Pesquisadora do Núcleo de Educação de Jovens e Adultos – NEJA/FaE/UFMG

² Professora aposentada/UFMG

³ Pesquisadora do Instituto de Ciências Exatas – Icx/UFMG

⁴ Coordenadora de programas da Ação Educacional

conhecimentos matemáticos de uso social, trabalhados – e, conseqüentemente, avaliados – em situações contextualizadas na vida dos alfabetizandos.

Contudo, essa preocupação com a realização de um diagnóstico da competência matemática dos alunos do *Programa Brasil Alfabetizado* e com a implementação de ações para desenvolvê-la no contexto de sua realização evidencia ainda uma outra dimensão essencial desse Programa: a da inclusão educacional.

De fato, a meta primordial da eliminação do analfabetismo de adultos em nosso País estabelece-se como ação inicial e estratégica de um processo de (re-)inserção de jovens e adultos no sistema escolar, na garantia de um direito do cidadão e de um espaço privilegiado para o acesso, a discussão, a partilha, a crítica e a construção de modos diversificados e mais democráticos de relação com o mundo da leitura e da escrita. É também na perspectiva da continuidade dos estudos dos jovens e adultos brasileiros, a partir da alfabetização, que se insere a participação da educação matemática no *Programa Brasil Alfabetizado*.

Foi, portanto, considerando as contribuições da educação matemática para as práticas de leitura e escrita que a vida social, nela também incluída a perspectiva da vivência escolar, impõe ou possibilita aos jovens e adultos brasileiros, que se concebeu este Documento.

Apresentam-se a seguir os pressupostos teóricos que subsidiam a concepção da avaliação de conhecimentos e competências matemáticas no âmbito do *Programa Brasil Alfabetizado* e os princípios considerados para a elaboração da matriz de referência para essa avaliação. O documento propõe, então, a matriz com os conhecimentos, as competências, os descritores e detalhamentos, incluindo exemplos de itens de teste relativos a alguns dos descritores, que visam orientar a construção de instrumentos que permitam realizar a avaliação pretendida.

Pressupostos teóricos e suas implicações para a elaboração da matriz de avaliação

A decisão relativa aos conhecimentos matemáticos que seriam contemplados numa matriz de referência fundamenta-se não apenas na relevância de conhecimentos dessa natureza para as práticas de leitura e escrita, mas também na consideração de que as ações no âmbito desse Programa desenvolvem-se num período de tempo relativamente pequeno – seis a oito meses – e que devem, portanto, ser tomadas como pontos de partida para uma inclusão mais duradoura dos sujeitos na escola.

Se a preocupação com o desenvolvimento de conhecimentos matemáticos numa iniciativa de alfabetização de pessoas jovens e adultas justifica-se pela contribuição indispensável que esses conhecimentos oferecem às práticas de leitura na sociedade brasileira atual, o *Programa Brasil Alfabetizado* e, portanto, sua avaliação quanto às habilidades matemáticas dos alfabetizados devem contemplar primordialmente noções e procedimentos matemáticos elementares, de amplo uso social, e nos contextos em que efetivamente se apresentam para os alunos do Programa. O domínio de tais noções e procedimentos, devido aos aportes essenciais que representam para a compreensão de diversos tipos de texto muito frequentes na vida dos sujeitos, é condição imprescindível à sua inclusão social e escolar.

A concepção da avaliação que aqui se propõe considerou ainda os resultados de pesquisas como a do Indicador Nacional de Alfabetismo Funcional Matemático (INAF) – Instituto Paulo Montenegro/Ação Educacional – que, nos anos de 2002 e 2004, contemplou as habilidades matemáticas da população jovem e adulta brasileira. Os resultados do INAF-2002 e do INAF-2004 revelam que 80% dos jovens e adultos brasileiros que não concluíram a 4ª série do ensino fundamental não conseguem ultrapassar o nível 1 de alfabetismo matemático funcional, que se caracteriza pelo sucesso apenas em tarefas de leitura de números de uso frequente em contextos específicos.

Sendo assim, é de se esperar que boa parte do público atendido pelo *Programa Brasil Alfabetizado*, ainda que possa deter algumas estratégias para a resolução de seus problemas diários, vivencie dificuldades impostas por um domínio precário de conhecimentos e/ou procedimentos elementares, próprios de uma cultura matemática socialmente valorizada, que marcam os modos de organização, produção e divulgação do conhecimento em nossa sociedade, e de cujo domínio dependem suas possibilidades de acesso, apreciação, crítica e utilização de informações, critérios ou estratégias em diversas instâncias da vida social; enfim, suas possibilidades de leitura do mundo.

Cabe, pois, aos processos de alfabetização contemplar tais conhecimentos matemáticos em sua proposta pedagógica, considerando os modos próprios pelos quais os alfabetizados lidam com as situações que os envolvem, mas também não se furtando à responsabilidade da apresentação de outras possibilidades de compreensão e/ou solução de problemas.

Tal apresentação deve assumir e explicitar a natureza cultural dos conhecimentos matemáticos, empenhando-se numa negociação respeitosa e democrática de novos significados e dispondo-se a uma discussão honesta e cuidadosa das

vantagens e desvantagens dos diversos procedimentos, inclusive os escolares, em diferentes oportunidades.

Tendo, pois, em vista que se pretende analisar e promover o domínio de conhecimentos matemáticos que são importantes num processo inicial de aquisição e desenvolvimento da leitura e da escrita, mas que ainda representam um obstáculo para grande parte dos alfabetizados jovens e adultos, optou-se por focalizar, na construção de uma matriz para avaliação do Programa, competências relativas a dois blocos de conhecimentos: Números e Operações.

Em virtude da importância fundamental do conhecimento dos números para diversas práticas de leitura, e da constatação de que em seus conceitos e representações residem muitas das necessidades e ainda das dificuldades de jovens e adultos alfabetizados, é nesse bloco que se localizam quatro das cinco competências propostas na matriz de referência. Tais competências configuram-se na contagem (como procedimento e como recurso na constituição do próprio conceito de número), e na leitura, na escrita e na comparação de números (representados no sistema decimal de numeração).

Com efeito, observações, relatos e análises de diversas experiências e pesquisas com alfabetizados jovens e adultos têm revelado dificuldades de algumas pessoas não apenas com a representação numérica, mas com o próprio conceito de número, o que justifica a inclusão de um primeiro descritor pelo qual se busca aferir a habilidade da realização da contagem elementar, de associação um-a-um, e de um outro descritor que se refere às estratégias para contagens elevadas, que mobilizam a realização de agrupamentos, idéia essencial para a compreensão do nosso sistema de numeração.

Ainda levando em consideração a responsabilidade do Programa com a garantia de domínio de algumas práticas importantes na vida social, propõe-se, nos descritores associados à competência da contagem, que ela se realize com objetos e também com quantias de dinheiro, com a utilização de cédulas e moedas.

O mesmo argumento justifica, nos descritores referentes às competências de leitura, escrita e comparação de números, a presença, junto aos números naturais, dos números decimais que expressam valor monetário, mesmo que tradicionalmente os números decimais sejam considerados um conteúdo sofisticado nos níveis de ensino muito elementares.

Quanto ao segundo bloco de conhecimentos escolhido para a composição da matriz de referência, o das operações, ainda considerando que o *Programa*

Brasil Alfabetizado representa somente uma primeira etapa, de duração relativamente pequena, na vida escolar dos sujeitos, optou-se por restringir a avaliação a uma competência – a de resolução de problemas envolvendo apenas a adição ou a subtração.

Também aqui, mesmo reconhecendo que, do ponto de vista estritamente matemático, os números decimais envolvem idéias mais sofisticadas do que as dos números naturais, propõe-se que as situações envolvendo as operações de adição ou subtração refram-se a quantidades discretas (representadas pelos números naturais), mas também tratem de adicionar ou subtrair pequenas quantias em dinheiro. Trata-se, nesse sentido, mais uma vez, de assumir a relevância das práticas e das demandas da vida social dos alunos nas decisões sobre os conhecimentos matemáticos a contemplar no desenvolvimento e na avaliação de um programa de Educação de Jovens e Adultos.

Pretende-se ainda, como se verá, avaliar essa competência de resolução de problemas com a utilização de recursos variados de cálculo, separadamente ou em conjunto, de acordo com a vontade do sujeito durante a aplicação do teste.

O que se espera, portanto, é conhecer a capacidade de resolução dos problemas propostos, com o uso de cálculo mental, lápis e papel e calculadora eletrônica. Essa idéia baseia-se, mais uma vez, na consideração de que os educandos são adultos cujas práticas sociais incluem, freqüentemente, a possibilidade e a necessidade de utilização de todos esses métodos e instrumentos.

Finalmente, cabem algumas palavras em relação às condições de realização do teste. Ao se elaborarem os itens do teste, deve-se considerar que as situações de contagem e de leitura, escrita e comparação de números, bem como aquelas que demandam a resolução de problemas por meio de adições ou subtrações, serão propostas oralmente aos sujeitos por um entrevistador, que recorrerá à manipulação de suportes conhecidos da população em geral, tais como calendário, cédulas e moedas, folhetos de propaganda e jornal, ou outros materiais que se mostrem necessários (e viáveis) para a simulação das situações propostas.

Para a execução de algumas tarefas do teste, o entrevistador oferecerá ao respondente lápis, papel e calculadora, e permitirá o recurso a estratégias pessoais de cálculo, como contar nos dedos, por exemplo.

A resposta produzida pelo entrevistado será também comunicada oralmente ou por meio de recursos gestuais (como apontar) ao entrevistador. Apenas alguns itens, particularmente os relacionados à competência de escrever números, deman-

darão uma produção escrita do respondente. Na resolução de problemas, o aluno poderá registrar seus cálculos, mas isso não lhe será cobrado na avaliação.

Levando em conta que o centro desta proposta reside na idéia de avaliar as habilidades matemáticas dos sujeitos que contribuem para o desenvolvimento da leitura, cabe insistir em que os itens do teste sejam elaborados de uma forma contextualizada, focalizando a contribuição dos conhecimentos matemáticos para as práticas de leitura da população brasileira.

Matriz de referência

Conhecimentos	Competências	Descritores	Detalhamento
Números	C1. Realizar Contagens	D1. Realizar contagens de pequenas quantidades dando o resultado oralmente	Procura-se avaliar a capacidade do aluno para realizar contagem de pequenas coleções de objetos (até aproximadamente 20 unidades, geralmente efetuadas de um em um) Ex: Há um “caixa especial” no supermercado para compras de até 15 volumes. O cliente que fez esta compra (mostrar ilustração), poderá usar este “caixa especial”?
		D2. Realizar contagens de quantidades maiores (por agrupamento ou outras estratégias) produzindo resposta oral	Procura-se avaliar a capacidade do aluno para propor e realizar uma estratégia de contagem para grandes coleções (por agrupamento, por exemplo), bem como se a extensão de seu domínio de contagem atinge a quantidade proposta (aproximadamente 50)
		D3. Realizar contagem de quantias em dinheiro com cédulas e moedas	Realizar contagem de quantias em dinheiro com cédulas de 20, 10, 5, 2 e 1 reais e moedas de 50, 25, 10, 5 e 1 centavos Nível 1: Somente cédulas de 1 real Nível 2: Somente cédulas de igual valor (de 2, de 5 ou de 10 reais) Nível 3: Somente cédulas, mas com valores variados Nível 4: Cédulas e moedas

Conhecimentos	Competências	Descritores	Detalhamento
Números (cont.)	C2. Ler números	<p>D4. Ler números naturais de um algarismo</p>	<p>Procura-se avaliar o nível de reconhecimento dos algarismos, ainda que o aluno não domine a estrutura do “Sistema de numeração decimal” Ex: O entrevistador exibe a foto de um jogador de futebol de costas, com uniforme da seleção brasileira, e o número estampado na camiseta tem apenas um algarismo. “Este é um craque da nossa seleção. Qual é o número que está estampado na sua camisa?”</p>
		<p>D5. Ler números naturais de 2, 3 ou 4 algarismos</p>	<p>Procura-se avaliar a capacidade de ler números que exijam certo domínio das regras do sistema de numeração decimal. Ex: Esta é a placa de uma casa na Rua Direita. Qual é o número da casa (mostrar placa ou cartão com o número da casa)? Nível 1: O número tem 2 algarismos Nível 2: O número tem 3 algarismos sem o zero intermediário Nível 3: O número tem 4 algarismos sem o zero intermediário Nível 4: O número tem 3 ou 4 algarismos com o zero intermediário</p>

(cont.)

Conhecimentos	Competências	Descritores	Detalhamento
Números (cont.)	C2. Ler números (cont.)	D6. Ler números decimais que expressam valor monetário	Procura-se avaliar a capacidade de leitura de preços e outras informações sobre valores monetários inferiores a 100 reais. Ex: Apresenta-se um folheto de propaganda com vários produtos e seu preço. O entrevistador aponta um produto e seu preço e pergunta: “qual é o preço deste produto?”
	C3. Escrever números	D7. Escrever números de 1 algarismo	Procura-se avaliar a identificação do símbolo correspondente ao nome do algarismo e sua habilidade para grafá-lo. Ex: “Você pode anotar o número de um telefone que eu vou ditar? Dois, três, cinco, nove, quatro, zero, oito”
		D8. Escrever números de 2 ou 3 algarismos	Procura-se avaliar o domínio do sistema de numeração decimal evidenciado pela capacidade de registrar um número de 2 ou 3 algarismos ditado pelo entrevistador Nível 1: O número tem 2 algarismos Nível 2: O número tem 3 algarismos sem o zero intermediário Nível 3: O número tem 3 algarismos com o zero intermediário

(cont.)

Conhecimentos	Competências	Descritores	Detalhamento
Números (cont.)	C4. Comparar números	D9. Comparar números naturais	<p>Procura-se avaliar a capacidade de comparação de números registrados no sistema de numeração decimal, ou expressos oralmente</p> <p>Nível 1: Comparar números de 2 algarismos expressos oralmente. Ex: Quem é mais velho: uma pessoa de quarenta e oito anos ou uma pessoa de cinquenta e três anos?</p> <p>Nível 2: Comparar dois números de 2 algarismos registrados no sistema de numeração decimal</p> <p>Nível 3: Comparar dois números de 3 algarismos registrados no sistema de numeração decimal</p> <p>Nível 4: Comparar um número de 2 com outro de 3 algarismos</p> <p>Ex: Qual desses números é maior: 89 ou 213 (mostrar placas)?</p>
		D10. Comparar números decimais que expressam valor monetário	<p>Procura-se avaliar a habilidade de comparar preços de produtos registrados por escrito</p> <p>Ex: O entrevistador mostra um folheto com vários pacotes de feijão, de mesmo tamanho, mas marcas diferentes, e seus respectivos preços</p> <p>“Nesta loja, qual desses produtos está mais barato?”</p>

(cont.)

Conhecimentos	Competências	Descritores	Detalhamento
Operações	C5. Resolver problemas envolvendo adição ou subtração	<p>D11. Resolver problemas envolvendo adição de números naturais ou de quantias em dinheiro por qualquer método, para a produção de uma resposta aproximada</p> <p>D12. Resolver problemas envolvendo adição de números naturais ou de quantias em dinheiro por qualquer método, para a produção de uma resposta exata</p> <p>D13. Resolver problemas envolvendo uma subtração de números naturais ou quantias em dinheiro por qualquer método, produzindo uma resposta oral</p>	<p>Procura-se avaliar a capacidade do aluno para efetuar uma adição por aproximação, sem utilizar registros escritos e nem a calculadora Ex: O entrevistador mostra um folheto de propaganda com os respectivos preços e pergunta: “Com 10 reais eu posso fazer esta compra: um pacote deste arroz, mais este pacote de feijão e mais este de açúcar?”</p> <p>Procura-se avaliar a capacidade de resolver problemas envolvendo apenas adição (de duas ou mais parcelas) utilizando quaisquer recursos de cálculo, inclusive a calculadora Nível 1: problemas que recaem em uma adição com 2 parcelas Nível 2: problemas que recaem em uma adição com 3, 4 ou 5 parcelas</p> <p>Procura-se avaliar a capacidade de resolver problemas envolvendo uma subtração de dois números, utilizando quaisquer recursos de cálculo, inclusive a calculadora</p>

Síntese dos Relatórios de Aplicação dos Instrumentos de Avaliação Cognitiva: conhecimentos matemáticos, língua escrita e perfil socioeconômico

Maria Madalena Torres¹

Quanto à época de aplicação dos instrumentos:

- Período de aplicação dos instrumentos cognitivos nos sete estados: 09 a 17/ 12/2004
- Período de aplicação dos instrumentos cognitivos no DF: 13 a 22/12/2004
- Foram aplicados 264 instrumentos cognitivos e 264 questionários do perfil socioeconômico

Quanto aos locais visitados e respectivos entrevistadores:

- AlfaSol/Universidade Brás Cubas – Mogi das Cruzes (SP): Francijairo Ananias da Silva
- Centro de Educação Paulo Freire de Ceilândia/Sobradinho/São Sebastião/Paranoá (DF): Fernanda Frade, Janaína Sabino e Maria Madalena Torres
- Congresso Nacional Afro-brasileiro/Cnab – São Paulo (SP): Marlete F. de Souza

¹ Técnica da Coordenação-Geral de Pedagogia de Alfabetização e Educação de Jovens e Adultos (DEJA/SECAD/MEC)

- Federação das Mulheres do Brasil – Natal (RN): Ivonete dos Santos Barbosa
- Secretaria de Educação de Aracaju (SE): Werner Ward de Oliveira*
- Secretaria de Educação e Cultura de Natal (RN): Stela Martins Teles
- Serviço Social da Indústria/SESI (DF): Jane Fontenele, Janaína Sabino de Oliveira, Maria Onézia, e Maria Madalena Torres
- Serviço Social da Indústria/SESI - Campina Grande (PB): Adriana Dias de Freitas e Clauzene Lima da Silva

Quanto à receptividade dos coordenadores e alfabetizadores

- A receptividade dos coordenadores, alfabetizadores e alfabetizando aos entrevistadores foi boa, com exceção à do entrevistador Werner Ward que teve algumas dificuldade em sua estada nas turmas em Aracaju, devido a uma situação criada por visitantes do PBA/Secad/MEC no período do monitoramento, que entraram em confronto com a equipe da Seduc local.

Quanto às instalações físicas

- Em alguns lugares , como Ceilândia e Sobradinho, onde as turmas não estavam mais em funcionamento, as visitas foram realizadas nas residências dos alfabetizando, não sendo possível observar as instalações das salas.
- As instalações das classes de alfabetização atendidas por projetos executados pelas Seducs são adequadas, mas as turmas que funcionam nas igrejas e salões comunitários apresentam dificuldades nas instalações elétricas. Nessas últimas, o quadro de giz, geralmente, é pequeno e as cadeiras, muitas vezes, não são apropriadas para o conforto dos alfabetizando.
- As turmas do SESI que funcionam na Empresa Qualix têm boa estrutura de sala de aula, apesar de uma das turmas funcionar no restaurante da empresa. Contudo, não há prejuízo para a qualidade do trabalho.

* Instrumentos aplicados nas turmas de 2004 (do meio para o final do processo de alfabetização).

Quanto à aplicação dos instrumentos

Na maioria dos grupos, o tempo médio de 50 minutos foi suficiente para a aplicação dos dois cadernos: de Língua Escrita e Matemática; no caso do instrumento perfil socioeconômico, quando foi aplicado na sala de aula, a duração foi mais ou menos de 15 minutos; porém, nas residências dos alfabetizandos, esse tempo variou de 30 a 40 minutos.

Considerações sobre os instrumentos cognitivos

No instrumento de **Língua Escrita/Leitura**, as páginas de bilhetes e cartazes devem permanecer separadas das questões. As demais questões devem permanecer com as figuras, mas com a fonte das palavras aumentada no próprio instrumento.

No cartaz “Ana Maria da Silva é sua melhor candidata à vereadora, nº 99444”, a imagem de Ana Maria com traje de médico e um estetoscópio pendurado ao pescoço dá dúbia interpretação no momento da leitura e interpretação da mensagem. Quando perguntamos: Ana Maria é candidata a quê? Os alfabetizandos divagam para responder, fazendo outras perguntas sobre médico, além da perda de tempo. Em relação à mesma figura e no caso específico do DF, não existe essa categoria política nesse distrito.

Quando os alfabetizandos sentem dificuldades na leitura de palavras e pequenas frases, ao se deparar com um dos vários bilhetes que aparecem nos testes, eles entram em desespero e ficam nervosos por não lhes conseguir ler.

No geral, existem questões com alternativas previsíveis para as respostas dos alfabetizandos e que não foram incluídas no instrumento.

O uso de fichas separadas é um processo muito trabalhoso; portanto, sugerimos que palavras isoladas em fichas, a exemplo de “portão” e “trabalhador”, sejam incluídas nas próprias questões. Entretanto, devem permanecer os cartazes e os bilhetes isoladamente, para auxílio da leitura. Contudo, para que o alfabetizando não seja induzido pela disposição das respostas, basta cobri-las com um papel adicional.

Sobre o instrumento cognitivo de Matemática, percebe-se que podemos otimizar o tempo de sua aplicação retirando muitas fichas desnecessárias, permanecendo com as cédulas dos “dinheirinhos”, aumentando a fonte das letras e organizando melhor o espaçamento na página. Enfim, o procedimento deve ser o mesmo do de Língua Escrita: cobrir as respostas para que os alfabetizandos não sejam induzidos por elas.

Algumas figuras não foram compreendidas. As cocadas foram confundidas com um pé de alface, por causa da disposição das cores preta e branca; os pães, como não estavam num tabuleiro, não foram identificados e tivemos que dizer o que eram; além da pintura renascentista da família, que foi alvo de risos para os entrevistados.

Os exercícios de contagem de pães e palitos praticamente se repetem: se aparecem os pães para serem contados abstratamente, não há necessidade de contar palitos de fósforo, pois se são capazes de contar os pães de forma abstrata, para que contar, concretamente, os palitos?

Na figura em que aparece o Ronaldinho, camisa 9, a maioria dos homens olha a imagem e, antes de fazermos alguma pergunta, já dizem com muito orgulho, “Ronaldinho, o camisa 9!”. As mulheres, no entanto, não têm o mesmo comportamento. Talvez fosse bom não elaborar questões óbvias, influenciadas pela mídia.

No cartaz que apresenta a seleção brasileira de futebol, quando pedimos que identifiquem a quantidade de pessoas, eles dizem logo: “Não seriam 11, ou, no máximo, 22? Esse que sobra é o técnico?”

Todos os relatórios apresentam observações por parte dos entrevistadores com relação aos problemas de visão (necessidade e carência de óculos), enfrentados por grande parte dos entrevistados.

Considerações sobre o perfil socioeconômico

No perfil socioeconômico, o problema da cor é impressionante. Uns são brancos e não querem ser, intitulado-se “morenos”; outros são negros e não querem assumir, contentando-se com a opção “pardo”, ou exigindo que se acrescente a cor “morena”. Por isso, no campo 7, sugere-se excluir “raça” e incluir “cor”.

Por duas vezes aparece a pergunta se a pessoa tem carteira de trabalho: na página 01, item 08 e depois na página 03, item 06.

No bloco E, item 2, sugere-se incluir “alfabetização” no histórico familiar.

No bloco F, item 4, sugere-se acrescentar ao campo “emprego temporário” os itens: contrato temporário, CLT, aposentado e pensionista. O mesmo poderia ocorrer no campo renda familiar, acrescentando os itens: renda de aposentado, pensionista ou trabalho autônomo.

No bloco G, item 1, que se refere ao acesso a programas de renda, sugere-se incluir perguntas como: solicitou o benefício na prefeitura? Não foi aprovado? Não recebeu resposta?

Considerações sobre a relação entrevistador/entrevistado

Reflexões desencadeadas durante o processo de aplicação dos instrumentos:

- Em relação ao pagamento das bolsas, há a compreensão de que a gestão não é fácil. Houve muitos atrasos de pagamentos e dificuldades para a continuidade dos alunos no 1º Segmento;
- Inclusão do tema racial nas discussões das turmas no DF, identificado a partir da aplicação do perfil socioeconômico;
- Criação de benefícios, como o exame de vista e a entrega de óculos, que venham a amenizar os problemas visuais enfrentados pelos alfabetizando;
- A união de programas intergovernamentais como o *Bolsa-Escola*, o *Fome Zero* e o *Fazendo Escola* - 1º Segmento da EJA;
- O impacto da saída de sua realidade para se deparar com o inusitado, numa melhor compreensão da realidade da alfabetização no Brasil;
- Participação em um programa de formação continuada que capacite melhor os alfabetizadores na técnica da alfa-numerização;
- Algumas falhas no instrumento cognitivo e no perfil socioeconômico deram espaço para que alguns entrevistadores mediassem a dificuldade entre o que pedia a questão e o entendimento do entrevistado.

Do ponto de vista dos entrevistados, muitos receberam com desconfiança os instrumentos cognitivos, poucos com satisfação, alguns com emoção.

No geral, é possível trabalhar com esses instrumentos cognitivos e com o perfil socioeconômico, desde que sejam adaptados às sugestões apresentadas pelos entrevistadores – já que, no primeiro momento, foram experimentais.

PS: Este relatório foi baseado em informações extraídas dos relatórios de cada entrevistador e da gravação em vídeo da reunião de relatos de experiências, realizada em 29 de dezembro de 2004.

Qualidade psicométrica dos itens da avaliação educacional dos alunos egressos do *Programa Brasil Alfabetizado*

Anamara Ferreira Ribeiro
Josemberg M. de Andrade¹

A avaliação educacional é uma ferramenta útil para melhorar o sistema educacional, fornecendo informações que permitem aos educadores determinar quais práticas promovem resultados desejados e quais não (Sanders & Horn, 1995). Ela deve prover um contínuo monitoramento do sistema educacional, com a finalidade de detectar os efeitos positivos ou negativos de políticas adotadas (Soares, Cesar & Mambrini, 2001).

Em se tratando do desempenho dos alunos, esse é verificado por meio de testes que aferem competências e habilidades. A construção de testes de avaliações educacionais requer um conhecimento sistemático das habilidades específicas a serem alcançadas. Cada item ou questão de um teste deve estar contribuindo com uma parcela significativa para o resultado final da avaliação (Rodrigues, 2002).

Dessa forma, a qualidade do teste depende diretamente da qualidade dos itens que o compõem. Para se obter um teste de alta qualidade técnica, e que permita fazer inferências válidas sobre o desempenho dos alunos, é necessário que os itens tenham sido elaborados com rigor metodológico (Inep, 2001). Como assinalam Laros, Pasquali e Rodrigues (2000), a qualidade do item é a base para a adequação

¹ Consultores da Coordenação Geral de Estudos e Avaliação (DAIE/SECAD/MEC)

de todas as análises subseqüentes, tais como a análise do desempenho escolar e de fatores que o afetam.

A análise psicométrica deve ser empreendida com a finalidade de identificar não apenas os itens que desempenham bem o papel de avaliar a competência requerida como também aqueles itens que não o fazem (Rodrigues, 2002). A análise psicométrica pode ser realizada pela Teoria Clássica dos Testes, ou seja, a TCT (Pasquali, 2003).

A TCT procura explicar o resultado final, ou seja, a soma das respostas dadas pelos indivíduos a um conjunto de itens. Especificamente, o interesse da TCT não recai sobre o traço latente, e sim sobre o comportamento, ou melhor, o escore num teste, o que equivale a um conjunto de comportamentos (Pasquali, 2003). Na verificação da adequação dos itens, são comumente estimados os índices de discriminação e dificuldade dos itens.

O índice de discriminação é utilizado para verificar o quanto cada item pode diferenciar estudantes que possuem diferentes níveis de habilidade. Em outras palavras, a discriminação dos itens refere-se ao grau em que um item diferencia corretamente os examinandos no comportamento que o teste pretende medir (Anastasi e Urbina, 2000).

O cálculo do índice de discriminação pode ser efetuado a partir de grupos critério ou da correlação do item com o escore total dos itens. Considerando essa segunda maneira, existem vários tipos de correlação para a obtenção desse índice. Duas das mais utilizadas são a correlação ponto-bisserial e a bisserial (Pasquali, 2003). Erthal (2003) assinala que, independente do método utilizado, o critério de 0,25 parece ser adequado para identificar itens capazes de diferenciar sujeitos com níveis variados de habilidades.

Na TCT, a dificuldade do item é definida como a porcentagem de sujeitos que respondem corretamente ao item. Um valor para o índice de dificuldade da TCT perto de 1,0 indica que o item é fácil e um valor perto de 0 (zero) indica que se trata de um item difícil. Vale salientar que quanto mais a dificuldade de um item se aproxima de 0,0 ou 1,00, menos informações diferenciais ele oferece sobre o examinado. Por outro lado, quanto mais o nível de dificuldade se aproxima de 0,50, mais diferenciações o item pode fazer (ver, por exemplo, Anastasi e Urbina, 2000; Erthal, 2003; Hambleton e cols., 1991; Pasquali, 2003).

Diante da necessidade de verificar a qualidade psicométrica dos itens de Língua Portuguesa e Matemática, aplicados em uma avaliação educacional de adultos

da Secretaria de Educação Continuada, Alfabetização e Diversidade (Secad), este relatório tem como objetivo apresentar os resultados das análises psicométricas realizadas. A seguir, é apresentado o método utilizado para o alcance dos objetivos deste estudo.

Método

Nesta seção é apresentado o método utilizado para o alcance do objetivo do presente relatório.

Instrumento

Foram analisados 84 itens de um teste cognitivo aplicado em uma avaliação educacional de adultos. Desses itens, 45 foram de Língua Portuguesa e 39 de Matemática.

Procedimentos

Foram realizadas as seguintes análises: (1) análises exploratórias, por meio de análises descritivas; e (2) verificação da dificuldade e discriminação dos itens, por meio da Teoria Clássica dos Testes (TCT).

Os consultores em avaliação educacional receberam o banco de dados por meio da Secretaria de Educação Continuada, Alfabetização e Diversidade (Secad) / Diretoria de Avaliação e Informações Educacionais (DAIE).

Após o recebimento das bases, foi realizada a limpeza no banco, análise exploratória inicial dos dados e verificada sua consistência.

Retirou-se, da base de dados, os alunos para os quais não havia resposta da prova, isto é, os ausentes. A seguir, foram solicitadas as frequências de resposta às alternativas de cada item, a fim de verificar os possíveis erros de digitação. Após a análise exploratória, foram verificados os índices de discriminação e dificuldade dos itens.

Análise de dados

Para a realização da análise exploratória dos dados, foi utilizado o *software SPSS (Statistical Package For Social Sciences)*, na versão 13.0. Para a estimação dos parâmetros dos itens, foi utilizado o *software Bilog for Windows*, na versão 3.11.

Resultados

Nesta seção são apresentados os resultados do presente relatório. Como já assinalado anteriormente, foram realizadas as seguintes análises: análise exploratória dos dados, estimação dos parâmetros dos itens e verificação da fidedignidade das provas.

Análise exploratória

A análise exploratória de dados consiste em um conjunto de procedimentos estatísticos realizado com o intuito de revelar o que está contido em um banco de dados. Ainda, através desse tipo de análise é possível verificar a adequação dos dados aos pressupostos requeridos pelas análises estatísticas (Tabachnick e Fidell, 1996).

Assim, a partir de uma análise exploratória, foi verificada a existência de problemas de digitação dos dados e, em seguida, esses erros foram corrigidos. Optou-se por excluir do banco de dados todos os questionários em branco, o que resultou em uma amostra total de 267 avaliados. Os itens não respondidos por esses avaliados foram desconsiderados das estimações dos parâmetros dos itens.

Estimação dos parâmetros dos itens

Discriminação dos itens

No presente estudo a verificação da discriminação dos itens foi efetuada por meio do cálculo da correlação bisserial (r_{bis}). Esse coeficiente é calculado com base na Teoria Clássica dos Testes (TCT) para todos os itens da prova. As provas de Língua Portuguesa e Matemática foram consideradas separadamente, ou seja, os 45 itens da prova de Língua Portuguesa e os 39 itens de Matemática. Para fins de interpretação, pode-se dizer que uma questão com um alto índice r_{bis} é capaz de separar os estudantes que possuem um alto ní-

vel da habilidade aferida daqueles que possuem um nível baixo e daqueles que não desenvolveram ainda a habilidade aferida.

Na tabela 1 encontram-se os coeficientes de correlação bisserial para cada um dos 39 itens avaliados na prova de Matemática, enquanto na tabela 2, são apresentados os coeficientes dos 45 itens da prova de Língua Portuguesa.

Tabela 1 – Índices de discriminação dos itens da prova de Matemática

Prova de Matemática					
Itens do nível 1 Discriminação		Itens do nível 2 Discriminação		Itens do nível 3 Discriminação	
D01_N1	0,65	D01_N2	0,81	D01_N3	0,42
D02_N1	0,66	D02_N2	0,11	D02_N3	0,40
D03_N1	0,09	D03_N2	0,23	D03_N3	0,99
D04_N1	0,86	D04_N2	---	D04_N3	-0,30
D05_N1	0,89	D05_N2	0,41	D05_N3	0,73
D06_N1	0,94	D06_N2	0,58	D06_N3	1,33
D07_N1	0,25	D07_N2	1,21	D07_N3	1,04
D08_N1	1,31	D08_N2	0,55	D08_N3	0,53
D09_N1	1,20	D09_N2	0,47	D09_N3	0,20
D10_N1	1,67	D10_N2	1,32	D10_N3	0,82
D11_N1	0,21	D11_N2	0,40	D11_N3	0,45
D12_N1	0,63	D12_N2	0,56	D12_N3	0,61
D13_N1	0,56	D13_N2	0,56	D13_N3	0,40
Média	0,76	Média	0,60	Média	0,59
Desvio-padrão	0,45	Desvio-padrão	0,35	Desvio-padrão	0,41
Valor mínimo	0,09	Valor mínimo	0,11	Valor mínimo	-0,30
Valor máximo	1,67	Valor máximo	1,32	Valor máximo	1,33

Como pode ser observado em relação ao nível de dificuldade 1 (N1) da prova de Matemática, a média da correlação bisserial desses itens foi de 0,76, com os índices variando de 0,09 a 1,67. Verifica-se que esses índices são, na maioria, discriminativos. O item D03_N1 é o que apresenta menor poder de discriminação, ou seja, menor capacidade de diferenciar avaliados com diferentes níveis de habilidades. Como já ressaltado anteriormente, são desejáveis índices de discriminação iguais ou superiores a 0,25.

No que se refere ao nível de dificuldade 2 (N2), os itens apresentaram uma média de discriminação de 0,60 com um d.p. de 0,35; o que indica se tratar de itens discriminativos, em sua maioria. O item mais discriminativo é o D10_N2, e o com menor poder de discriminação é o item D02_N2. Ressalta-se que o índice de discriminação do item D04_N2 não pôde ser calculado, já que foi respondido corretamente por todos os avaliados.

O nível de dificuldade 3 (N3) foi o que apresentou uma menor média de discriminação dos itens (0,59), com os índices variando de -0,30 a 1,33. No entanto, a média apresentada indica que os itens conseguem diferenciar avaliados com diferentes níveis de habilidades. O item D04_N3 apresentou o menor poder de discriminação, indicando não ser um item adequado para diferenciar sujeitos com níveis diferentes de habilidades. Ressalta-se, ainda, que não são esperados índices de correlação bisserial negativos.

Em relação ao conjunto total de itens da prova de Matemática, ou seja, juntando os itens dos três níveis de dificuldade, a média de discriminação obtida foi de 0,65, com um d.p. de 0,40. Ainda, os índices de discriminação variaram de -0,30 a 1,67. O menor índice de discriminação foi encontrado no nível três, e o maior no nível um. De maneira geral, com a exceção de poucos itens, pode-se afirmar que a prova de Matemática é capaz de diferenciar sujeitos com níveis de habilidades diferentes.

Na tabela 2 são apresentados os índices de discriminação dos itens da prova de Língua Portuguesa.

Tabela 2 – Índices de discriminação dos itens da prova de Língua Portuguesa

Prova de Língua Portuguesa					
Itens do nível 1 Discriminação		Itens do nível 2 Discriminação		Itens do nível 3 Discriminação	
D02_N1	1,04	D02_N2	---	D02_N3	0,33
D03_N1	0,77	D03_N2	0,78	D03_N3	0,08
D04_N1	1,03	D04_N2	0,39	D04_N3	0,87
D05_N1	0,53	D05_N2	0,39	D05_N3	0,74
D06_N1	0,70	D06_N2	0,58	D06_N3	0,23
D07_N1	0,76	D07_N2	0,84	D07_N3	0,86
D08_N1	0,84	D08_N2	0,66	D08_N3	0,99
D09_N1	0,52	D09_N2	0,58	D09_N3	0,68
D10_N1	0,99	D10_N2	1,03	D10_N3	0,71
D11_N1	1,16	D11_N2	0,89	D11_N3	1,00
D12_N1	0,96	D12_N2	1,07	D12_N3	0,96
D13_N1	1,05	D13_N2	0,95	D13_N3	1,04
D14_N1	0,65	D14_N2	0,95	D14_N3	1,06
D15_N1	0,95	D15_N2	1,15	D15_N3	1,05
D16_N1	1,04	D16_N2	0,99	D16_N3	0,80
Média	0,87	Média	0,80	Média	0,76
Desvio-padrão	0,20	Desvio-padrão	0,25	Desvio-padrão	0,31
Valor mínimo	0,52	Valor mínimo	0,39	Valor mínimo	0,08
Valor máximo	1,16	Valor máximo	1,15	Valor máximo	1,06

Como pode ser observado em relação ao nível de dificuldade 1 (N1) da prova de Língua Portuguesa, a média da correlação bisserial desses itens foi de 0,87, com os índices variando de 0,52 a 1,16. Verifica-se que esses índices são satisfatórios e que os itens são discriminativos.

No que se refere ao nível de dificuldade 2 (N2), os itens apresentaram uma média de discriminação de 0,80 com um d.p. de 0,25; o que indica se tratar de itens discriminativos, ou seja, capazes de diferenciar adequadamente, avaliados com níveis diferentes de habilidades. Ressalta-se que o índice de discriminação do item D02_N2 não pôde ser calculado, já que foi respondido corretamente por todos os avaliados.

O nível de dificuldade 3 (N3) foi o que apresentou uma menor média de discriminação dos itens (0,76), com os índices variando de 0,08 a 1,06. No entanto, a média apresentada indica que os itens conseguem diferenciar avaliados com diferentes níveis de habilidades.

O item D03_N02 apresentou o menor poder de discriminação, indicando não ser um item adequado para diferenciar sujeitos com níveis diferentes de habilidades.

Em relação ao conjunto total de itens da prova de Língua Portuguesa, a média de discriminação obtida foi de 0,81, com um d.p. de 0,26. O menor índice de discriminação foi encontrado no nível três e o maior no nível um. Sendo assim, pode-se afirmar que a prova de Língua Portuguesa é capaz de diferenciar sujeitos com níveis de habilidades diferentes.

Dificuldade dos Itens

A tabela 3 apresenta os índices de dificuldade para cada um dos itens da prova de Matemática.

Tabela 3 – Índices de dificuldade dos itens da prova de Matemática

Prova de Matemática					
Itens do nível 1 Dificuldade		Itens do nível 2 Dificuldade		Itens do nível 3 Dificuldade	
D01_N1	0,98	D01_N2	0,88	D01_N3	0,93
D02_N1	0,88	D02_N2	0,69	D02_N3	0,87
D03_N1	0,97	D03_N2	0,93	D03_N3	0,82
D04_N1	0,95	D04_N2	1,00	D04_N3	0,98
D05_N1	0,92	D05_N2	0,76	D05_N3	0,79
D06_N1	0,87	D06_N2	0,88	D06_N3	0,92
D07_N1	0,98	D07_N2	0,95	D07_N3	0,94
D08_N1	0,95	D08_N2	0,78	D08_N3	0,53
D09_N1	0,95	D09_N2	0,93	D09_N3	0,87
D10_N1	0,97	D10_N2	0,98	D10_N3	0,95
D11_N1	0,84	D11_N2	0,90	D11_N3	0,81
D12_N1	0,79	D12_N2	0,84	D12_N3	0,72
D13_N1	0,82	D13_N2	0,84	D13_N3	0,66
Média	0,91	Média	0,87	Média	0,83
Desvio-padrão	0,07	Desvio-padrão	0,09	Desvio-padrão	0,13
Valor mínimo	0,79	Valor mínimo	0,69	Valor mínimo	0,53
Valor máximo	0,98	Valor máximo	1,00	Valor máximo	0,98

Como já assinalado anteriormente, na Teoria Clássica dos Testes (TCT) a informação sobre a dificuldade é fornecida pela proporção de acertos na alterna-

tiva correta. Ressalta-se que as análises foram realizadas considerando os três níveis de dificuldade dos itens, propostos pelo(s) elaborado(es) dos itens.

Como pode ser observado em relação ao nível de dificuldade 1 (N1) da prova de matemática, a proporção média de acerto desses itens foi de 0,91, o que indica que foram acertados por 91% dos avaliados. A partir desses resultados, e considerando os critérios apresentados por Pasquali (2003), esses itens foram considerados muito fáceis por parte da amostra de avaliados.

Em relação aos itens do nível de dificuldade 2 (N2), a dificuldade média dos itens foi de 0,87, com valores mínimo e máximo iguais a 0,69 e 1,00, respectivamente. Com base na média de dificuldade encontrada e considerando os critérios apresentados por Pasquali (2003), esses itens podem ser considerados muito fáceis. Como visto, o item D04_N2 foi respondido corretamente por todos os avaliados.

Por fim, a média dos itens do nível de dificuldade 3 (N3) foi de 0,83, o que indica que esses itens foram acertados por 83% dos avaliados. Similarmente aos itens anteriormente avaliados, os itens do nível de dificuldade 3 da prova de Matemática também foram considerados muito fáceis para a amostra de avaliados.

Em relação ao conjunto total de itens avaliados na prova de Matemática (39 itens), a média de dificuldade obtida foi de 0,87, com um d.p. de 0,10. O item mais difícil foi encontrado no nível três, e o mais fácil no nível dois. Sendo assim, pode-se afirmar que a prova de Matemática foi uma prova fácil para os avaliados.

A seguir, são apresentados os índices de dificuldade para cada um dos itens da prova de Língua Portuguesa.

Tabela 4 – Índices de dificuldade dos itens da prova de Língua Portuguesa

Prova de Língua Portuguesa					
Itens do nível 1 Dificuldade		Itens do nível 2 Dificuldade		Itens do nível 3 Dificuldade	
D02_N1	0,96	D02_N2	1,00	D02_N3	0,94
D03_N1	0,89	D03_N2	0,96	D03_N3	0,84
D04_N1	0,94	D04_N2	0,70	D04_N3	0,93
D05_N1	0,80	D05_N2	0,62	D05_N3	0,72
D06_N1	0,73	D06_N2	0,73	D06_N3	0,43
D07_N1	0,96	D07_N2	0,91	D07_N3	0,69
D08_N1	0,79	D08_N2	0,82	D08_N3	0,65
D09_N1	0,97	D09_N2	0,94	D09_N3	0,49
D10_N1	0,87	D10_N2	0,74	D10_N3	0,74
D11_N1	0,78	D11_N2	0,76	D11_N3	0,77
D12_N1	0,73	D12_N2	0,66	D12_N3	0,63
D13_N1	0,71	D13_N2	0,67	D13_N3	0,56
D14_N1	0,87	D14_N2	0,67	D14_N3	0,69
D15_N1	0,74	D15_N2	0,78	D15_N3	0,69
D16_N1	0,77	D16_N2	0,61	D16_N3	0,59
Média	0,83	Média	0,77	Média	0,69
Desvio-padrão	0,09	Desvio-padrão	0,13	Desvio-padrão	0,14
Valor mínimo	0,71	Valor mínimo	0,61	Valor mínimo	0,43
Valor máximo	0,97	Valor máximo	1,00	Valor máximo	0,94

Como pode ser observado em relação ao nível de dificuldade 1 (N1) dos itens da prova de Língua Portuguesa, a proporção média de acerto desses itens foi de 0,83, o que indica que foram acertados por 83% dos avaliados. A partir desses resultados, pode-se dizer que esses itens foram considerados muito fáceis por parte da amostra de avaliados.

Em relação aos itens do nível de dificuldade 2 (N2), a dificuldade média dos itens foi de 0,77, com valores mínimo e máximo iguais a 0,61 e 1,00, respectivamente. Com base na média de dificuldade encontrada e considerando os critérios apresentados por Pasquali (2003), esses itens foram considerados fáceis por parte da amostra de avaliados.

Por fim, a média dos itens do nível de dificuldade 3 foi de 0,69, o que indica que esses itens foram acertados por 69% dos avaliados. Nessa categoria, foi encontrado o item mais difícil de toda a prova, a saber: D06_N3. Recomenda-se

que se verifique a adequação pedagógica desse item, bem como se investigue, junto aos alunos, se as habilidades e competências referentes a esse item foram realmente desenvolvidas pelos alunos. No geral, os itens desse nível foram considerados fáceis por parte dos alunos.

Em relação ao conjunto total de itens da prova de Língua Portuguesa, a média de dificuldade obtida foi de 0,77, com um d.p. de 0,13 e valores entre 0,43 e 1,00. O item mais difícil foi encontrado no nível três, e o mais fácil no nível dois. Sendo assim, pode-se afirmar que a prova de Português é uma prova fácil para os avaliados.

Considerações Finais

Os resultados das análises exploratórias e psicométricas encontrados para a prova de Língua Portuguesa e de Matemática indicam que os itens construídos para essa avaliação são muito fáceis para os avaliados. Quando o interesse está em verificar o poder de um teste para discriminar diferentes níveis de habilidades nos sujeitos, como é o caso de provas de avaliações educacionais, requer-se uma distribuição equilibrada dos itens em termos de dificuldade. Itens muito fáceis, que todos os avaliados acertam, e itens muito difíceis, os quais ninguém acerta, são itens que não diferenciam os indivíduos e por isso não trazem nenhuma informação (Pasquali, 2003). A literatura indica que quanto mais o nível de dificuldade se aproxima de 0,50, mais diferenciações o item pode fazer (ver, por exemplo, Anastasi e Urbina, 2000; Erthal, 2003; Hambleton e cols., 1991; Pasquali, 2003).

No que se refere à discriminação dos itens, as médias encontradas foram 0,65 e 0,81 em Matemática e Língua Portuguesa, respectivamente, indicando se tratar de itens, no geral, bastante discriminativos. No entanto, em ambas as provas foram encontrados itens que, praticamente, não diferenciam os examinandos quanto ao seu nível de conhecimento. Smith e McCarthy (1995) afirmam que, para o refinamento de um instrumento, ou seja, para a melhoria de sua representação de um construto, é importante, além de outros objetivos, a inclusão de itens que discriminem os participantes no nível de intensidade desejado para o atributo.

Os resultados mostram, assim, que a prova de Língua Portuguesa encontra-se mais adequadamente construída, com itens mais discriminativos. Dessa forma, com base no que foi encontrado, sugere-se que seja dada uma maior atenção à prova de Matemática, no sentido de melhorar esses itens menos discriminativos. Sugere-se também, que sejam incluídos itens mais difíceis em ambas as provas, com o objetivo de promover uma maior diferenciação entre os avaliados.

Referências

- Anastasi, A. e Urbina, S. (2000). *Testagem psicológica* (7ª ed.). Porto Alegre: Artmed.
- Erthal, T. C. (2003). *Manual de psicometria* (7ª ed). Rio de Janeiro: JorgeZahar Editor.
- Hambleton, R. K., Swaminathan, H. e Rogers, H. J. (1991). *Fundamentals of item response theory*. California: Sage Publications.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep (2001). *Saeb 2001: Novas perspectivas*. Brasília: Ministério da Educação/Inep.
- Laros, J. A., Pasquali, L., e Rodrigues, M. M. M. (2000). “Análise da Unidimensionalidade das provas do Saeb– Relatório Técnico”. Brasília: Centro de Pesquisa em Avaliação Educacional – Universidade de Brasília.
- Pasquali, L. (2003). *Psicometria: teoria dos testes na psicologia e na Educação*. Petrópolis: Editora Vozes.
- Rodrigues, M. M. M. (2002). *Instrumentos de avaliação educacional: uma visão pedagógica e psicométrica integradas: estudos das provas do Saeb, Matemática 8ª série, 1997 e 1999*. Dissertação de mestrado. Brasília: Universidade de Brasília.
- Sanders, W. L. & Horn, S. P. (1995). “Educational assessment reassessed: the usefulness of standardized and alternative measures of student achievement as indicators for the assessment of educational outcomes”. *Education Policy Analysis Archives*, 3, 1-16. Retirado em 25/04/2002, <http://oram.ed.asu.edu/epaa>.
- Smith, G. T. & McCarthy, D. M. (1995). Metodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment*, 7, 300-308.
- Soares, J. F., César, C. C. & Mambrini, J. (2001). “Determinantes de desempenho dos alunos do Ensino Básico brasileiro: evidências do Saeb de 1997”. Em C. Franco (Org.) *Avaliação, ciclos e promoção na Educação* (pp. 121-153). Porto Alegre: Artes médicas.
- Tabachnick, B. G. & Fidell, L. S. (1996). *Using multivariate statistics* (3th ed.). New York: Harper Collins College Publishers.

Desafios da
amostragem

Estudo sobre o desenho da amostra de alunos e turmas do *Programa Brasil Alfabetizado*

Mauricio Teixeira Leite
de Vasconcellos¹

Este capítulo apresenta a seleção de uma amostra probabilística de turmas e alunos da edição 2004 do *Programa Brasil Alfabetizado* (PBA), com base no cadastro de entidades, turmas, alfabetizadores e alfabetizandos, mantido pelo Ministério da Educação e conhecido como Sistema do Brasil Alfabetizado (SBA), bem como da documentação dos procedimentos adotados em sua seleção.

Considera-se que a amostra selecionada para o SBA 2004 não podia ser do conhecimento prévio da Secretaria de Educação Continuada, Alfabetização e Diversidade (Secad), seja por razões éticas associadas ao processo de avaliação externa, seja para resguardar a Instituição de qualquer possibilidade de crítica relacionada à possível atuação no sentido de melhorar os resultados da avaliação. Os arquivos magnéticos contendo a amostra selecionada foram encaminhados diretamente à entidade responsável pela avaliação externa do PBA.

Assim, este trabalho limita-se à descrição do desenho da amostra selecionada e à documentação do arquivo com as unidades selecionadas para a amostra, e não fornece qualquer indicação que possibilite identificar as turmas.

¹ Professor da Escola Nacional de Ciências Estatísticas, Doutor em Saúde Pública – ENSP/FIOCRUZ

Objetivos da amostra do SBA 2004

A amostra de turmas e alunos da edição 2004 do PBA, denominada simplesmente por amostra do SBA 2004, tem o objetivo eminentemente exploratório, fundamental para conhecer e testar todo o procedimento de obtenção dos dados necessários à avaliação do PBA.

Neste sentido, a amostra de 2004 serviu para verificar a possibilidade de uso do SBA como cadastro de seleção de amostras de turmas para avaliação do PBA, visto que o cadastro concebido para fins administrativos nem sempre se presta àquela finalidade. Além da conhecida necessidade de reorganizar cadastros administrativos para geração de cadastros de seleção, foi necessário testar a qualidade de suas informações, sua abrangência e seus mecanismos de atualização. Desta maneira, foram identificadas lacunas de informação, que demandavam preenchimento para viabilizar o uso do SBA como cadastro de seleção.

Essa amostra serviu, também, para testar os instrumentos de coleta (questionários e manuais de instrução) e verificar sua adequação ao perfil dos informantes, bem como para experimentar os testes de avaliação do grau de aprendizado dos alfabetizandos.

Desenho da amostra do SBA 2004

População alvo e população de pesquisa

A população alvo corresponde ao conjunto de alfabetizandos do Edital 2004 do PBA. Esta população está organizada por entidades parceiras subdivididas em turmas, às quais, por sua vez, estão associados os alfabetizandos e os alfabetizadores.

No entanto, como o teste dos instrumentos de coleta teria que ser feito em dezembro de 2005, a população de pesquisa foi limitada às turmas com término neste período. A Tabela 1 indica o tamanho das populações alvo e de pesquisa em cada macrorregião do País.

Tabela 1 – Números de turmas do SBA 2004 e com fim em dezembro de 2005, segundo as macrorregiões

Macrorregiões	Número de turmas do SBA 2004		Turmas com fim em dezembro 2005	
	Absoluto	%	Absoluto	%
Total	69.688	100,00	2.109	100,00
Norte	6.724	9,65	380	18,02
Nordeste	46.602	66,87	1045	49,55
Sudeste	10.124	14,53	553	26,22
Sul	3.524	5,06	97	4,60
Centro-Oeste	2.714	3,89	34	1,61

Observa-se que a distribuição regional do número de turmas com fim em dezembro de 2005 é substancialmente diferente do total de turmas do SBA 2004.

Tamanho da amostra do SBA 2004

Por razões orçamentárias, o tamanho da amostra foi fixado em cerca de 400 alfabetizando em todo o País, a fim de respeitar os limites mínimos de observações necessárias ao algoritmo de Teoria de Resposta ao Item (TRI).

Para reduzir o efeito de conglomeração da amostra, no entanto, foi decidido aplicar os testes cognitivos a quatro alunos de cada turma selecionada, o que resultou em um tamanho de amostra de 100 turmas.

Com o objetivo de representar a diversidade do País, decidiu-se estratificar a amostra de turmas por macrorregião e alocar o tamanho total da amostra proporcionalmente ao número de turmas do SBA 2004 em cada macrorregião, garantindo um tamanho mínimo de 10 turmas por estrato (Tabela 2).

Tabela 2 – Número de turmas do SBA 2004 e alocação do tamanho da amostra por macrorregião.

Macrorregiões	Nº de turmas do SBA 2004 (Nh)	Tamanho da amostra de turmas	
		Calculado (nh)	Ajustado (n*)h
Total	69.688	100,0	112
Norte	6.724	9,6	10
Nordeste	46.602	66,9	67
Sudeste	10.124	14,5	15
Sul	3.524	5,1	10
Centro-Oeste	2.714	3,9	10

Desta forma, pode-se ver na Tabela 2 que o tamanho calculado da amostra no estrato h, representado por n_h , é dado por $n_h = N_h \times 100 / N$, onde $N = \sum N_h$. O tamanho ajustado da amostra, para o estrato h, é dado por $n_h = \text{mínimo}(10; n_h)$.

Seleção de turmas e alunos

No primeiro estágio de seleção, foi selecionada uma amostra aleatória simples de turmas, usando o algoritmo de Hájek (1960)¹. Este algoritmo consiste em associar um número aleatório, uniformemente distribuído no intervalo (0 ; 1), a cada unidade da população de pesquisa (no caso, a cada turma com fim em dezembro de 2005), classificar os registros por estrato e número aleatório, e selecionar as n_h turmas na ordem dos registros, ou seja, selecionar as n_h turmas que receberam os menores números aleatórios do estrato h.

No segundo estágio, quatro alunos deviam ser selecionados, com equi-probabilidade entre os presentes no dia da entrevista, de cada turma selecionada no primeiro estágio. Esta seleção deve ser feita em campo, usando a planilha fornecida. Para tanto, basta digitar o número de alunos presentes e verificar nas colunas correspondentes os números de ordem dos quatro alunos selecionados.

¹ Hájek, J., 1960. Limiting distribution in simple random sampling from a finite population. Publications of the Mathematics Institute of the Hungarian Academy Science, 5:361-374.

Descrição do arquivo da amostra do SBA 2004

A amostra de turmas selecionadas foi fornecida por meio de uma planilha em formato MS-Excel, contendo as colunas indicadas no quadro abaixo, juntamente com as instruções e a documentação.

Colunas	Descrição	Documento/instruções
A a J	Dez números aleatórios para seleção dos alunos (colunas escolhidas)	
K	Código IPEA da turma	(variável que identifica a turma e que deverá constar de todos os questionários preenchidos)
L	Resultado da entrevista	(digite o resultado da entrevista)
M	Número de alunos presentes	(digite o número de alunos presentes para que apreçam os números de alunos a serem entrevistados)
Número dos alunos selecionados na ordem da lista de freqüência		
N	1º aluno	(entreviste os quatro alunos cujos números de ordem na lista de freqüência coincidam com os indicados nestas colunas, começando na coluna 1º aluno e buscando nas seguintes até encontrar quatro número diferentes e associados a alunos presentes)
O	2º aluno	
P	3º aluno	
Q	4º aluno	
R	5º aluno	
S	6º aluno	
T	7º aluno	
U	8º aluno	
V	9º aluno	
W	10º aluno	
Dados de identificação e endereço da turma		
X	Código MEC da turma	U- urbana R- rural
Y	Total previsto de alunos na turma	
Z	Endereço	
AA	Código IBGE do município	
AB	Sigla da UFAC	
	CEP	
AD	Zona	
AE	DDD	
AF	Telefone	
AG	FAX	

Tipo de locação da turma		
AH	Tipo de turma	
AI	Locação da turma	5- Casas do alfabetizador 10 – Escola 15 – Igreja 20 - Salão Paroquial 25- Sociedade de Amigos do Bairro 99 - Outro
Datas de início e término das aulas		
AJ	Início	
AK	Término previsto	
AL	Término efetivo	
	Horário de aula	
AM	Início	
AN	Fim	
Dias previstos com aulas		
AO	Segunda-feira	
AP	Terça-feira	
AQ	Quarta-feira	
AR	Quinta-feira	
AS	Sexta-feira	
AT	Sábado	
AU	Domingo	
Dados do alfabetizador		
AV	Código MEC do alfabetizador	
AW	Nome do alfabetizador	
AX	Sexo	F- feminino M - masculino
AY	Número da identidade	
AZ	Código de escolaridade	02- Antigo Primário 03- Antigo Ginásio 04-Antigo Clássico, Científico, Etc 05-Ensino Fundamental Ou 1º Grau 06 – Ensino Médio Ou 2º Grau 07- Superior- Graduação 08 – Mestrados ou Doutorado
BA	Código da profissão	5 – Alfabetizador 10-Aposentado 15- Desempregado 20 – Estudante 25 – Professor Alfabetizador 30- Professor Educação Infantil 35- Professor Ensino Fundamental 40- Professor Ensino Médio 45 – Professor Universitário 50 – Trabalhador Rural 55- Trabalhador Urbano 99- Outra

Dados para localização do alfabetizador		
BB	Endereço	
BC	Município	
BD	UF	
BE	CEP	
BF	DDD	
BG	Telefone	
BH	Celular	
BI	e-mail	
BE	CEP	
Dados da entidade		
BJ	CNPJ	
BK	Razão Social	
BL	Tipo	1- Prefeitura 2- 2-Secretaria Estadual de Educação 3- Ong 4- Universidade 5- Outras Secretarias Estaduais 6- Secretaria Municipal de Educação 7- Secretaria De Educação do Df 8- Universidade Federal 9- Empresa Privada 10- Empresa Pública 11- IES Privada 12- IES Municipal 13- IES Estadual
BM	Endereço	
BN	Bairro	
BO	Complemento	
BP	Código IBGE do município	
BQ	Sigla da UF	
BR	CEP	
BS	DDD	
BT	Telefone	
BU	Fax	
BV	E-mail	
BW	Dados de seleção Número aleatório	
BX	Código da macrorregião	1- Norte 2- Nordeste 3- Sudeste 4- Sul 5- Centro-Oeste
BY	Número de ordem	

Estudo sobre o desenho da amostra de alunos e turmas do *Programa Brasil Alfabetizado*

Mauricio Teixeira Leite
de Vasconcellos¹

Introdução

Este capítulo trata da seleção de uma amostra probabilística de turmas e alunos da edição 2005 do *Programa Brasil Alfabetizado* (PBA), selecionada com base no cadastro de entidades, turmas, alfabetizadores e alfabetizandos, mantido pelo Ministério da Educação e conhecido como Sistema do Brasil Alfabetizado (SBA), bem como da documentação dos procedimentos adotados em sua seleção.

Considera-se que a amostra selecionada para o SBA 2004 não podia ser do conhecimento prévio da Secretaria de Educação Continuada, Alfabetização e Diversidade (Secad), seja por razões éticas associadas ao processo de avaliação externa, seja para resguardar a Instituição de qualquer possibilidade de crítica relacionada à possível atuação no sentido de melhorar os resultados da avaliação. Os arquivos magnéticos contendo a amostra selecionada foram encaminhados diretamente à entidade responsável pela avaliação externa do PBA.

Assim, este trabalho limita-se à descrição do desenho da amostra selecionada e à documentação do arquivo com as unidades selecionadas para a amostra, e não fornece qualquer indicação que possibilite identificar as turmas.

¹ Professor da Escola Nacional de Ciências Estatísticas, Doutor em Saúde Pública – ENSP/FIOCRUZ

Objetivos da amostra do SBA 2005

A amostra de turmas e alunos da edição 2005 do PBA, denominada simplesmente por amostra do SBA 2005, tem o objetivo geral de conhecer o perfil socioeconômico dos alunos dos cursos de Alfabetização de Jovens e Adultos (AJA) financiados pelo PBA 2005, saber como foram selecionados e avaliar seu grau de aprendizado.

É certo, no entanto, que a avaliação de um programa passa por sua implementação e que esta pode ser diferenciada por segmentos da população. Assim, seria desejável que o objetivo geral pudesse ser cumprido para diferentes segmentos da população, tais como: macrorregiões do País e tipo de entidade parceira, dentre outros.

De fato, o desejo (ou necessidade) de que os objetivos sejam cumpridos para segmentos da população implica que os resultados possam ser estimados com precisão controlada para estes segmentos, o que, habitualmente, é feito com a estratificação da população. A solução assegura que os estratos populacionais sejam domínios de estimação, para os quais os resultados podem ser calculados com exatidão previamente controlada no desenho da amostra.

População alvo e população de pesquisa

A população alvo corresponde ao conjunto de alfabetizandos do Edital 2005 do PBA. Esta população está organizada por entidades parceiras, que se subdividem em turmas, às quais, por sua vez, estão associados os alfabetizandos e os alfabetizadores.

A população de pesquisa corresponde ao conjunto de alfabetizandos do Edital 2005 do PBA presentes no cadastro de seleção da amostra, ou seja, os alfabetizandos de turmas válidas registradas no SBA 2005 na data de referência do cadastro de seleção (04 de abril de 2006), que correspondem a 98,2% do total de alfabetizandos constantes no SBA 2005.

Estratificação a alocação do tamanho da amostra do SBA 2005

Por razões orçamentárias, o tamanho da amostra foi fixado em 800 turmas em todo o País. Este tamanho de amostra foi repartido (ou alocado) entre os es-

tratos definidos para a amostra para viabilizar a seleção das turmas em cada estrato da amostra.

Como documentado no estudo para estratificação e desenho da amostra, a ser usada para avaliação do impacto dos projetos aprovados no âmbito da Resolução de 2005 do *Programa Brasil Alfabetizado* (Science, 2006), a estratificação inicial deveria ser feita por tipo de entidade parceira, usando o algoritmo proposto por Dan Hedlin (Hedlin, 2000), visto que a distribuição do número de turmas por entidade apresenta acentuada assimetria positiva.

O algoritmo de Hedlin foi usado para determinar os pontos de corte dos estratos de forma a minimizar a variância, dado um tamanho total de amostra e um número de estratos, de forma a obter a alocação do tamanho de amostra que conduzisse à maior precisão (ou à menor variância) dentre todas as amostras estratificadas de tamanho total igual a 800.

No entanto, como as entidades estaduais e as Instituições de Ensino Superior apresentavam tamanhos populacionais pequenos (22 e 5, respectivamente), as entidades deste tipo foram alocadas ao estrato certo e o método foi aplicado de forma independente para as entidades municipais e para cada Organização Não-Governamental (ONG).

Para tanto, foi necessário alocar o tamanho total de turmas para cada tipo de entidade e definir o número de turmas a pesquisar em cada entidade de cada estrato. Assim, a solução encontrada foi aplicar o método de Hedlin para diferentes combinações de número de estratos², de tamanhos da amostra de turmas em cada entidade de cada estrato e, portanto, de tamanho total da amostra de entidades nos estratos de entidades municipais e ONGs, como já foi descrito no estudo indicado acima.

Posteriormente, para acomodar uma Instituição de Ensino Superior que tinha apenas duas turmas, o tamanho final da amostra foi aumentado para 802 turmas, conforme indicado na Tabela 1.

A Tabela 1, apresenta o resultado do estudo de estratificação (Science, 2006) e indica que:

- (1) As entidades estaduais e as Instituições de Ensino Superior foram todas incluídas na amostra, sendo a partir de então tratadas como estrato de seleção;

² Segundo Cochran (1977), evidências empíricas sugerem que o número de estratos deve variar entre 5 e 10. Assim, o número de estratos variou de 4 a 9, considerando que haveria o estrato certo.

- (2) Apesar de incluídas na amostra, estas entidades foram estratificadas por classes do número de turmas para ter diferentes tamanhos da amostra de turmas por entidade;
- (3) No terceiro estrato de classes do número de turmas, relativo ao estrato de ONGs, apesar do método de Hedlin ter indicado a seleção de cinco entidades, optou-se por reduzir o ponto de corte do estrato certo e incluir todas as sete entidades no estrato certo, que foi dividido em dois estratos para acomodar diferentes tamanhos da amostra de turmas por entidade (15 turmas para cada entidade com 500 a 2999 turmas, e 20 turmas para cada uma das quatro entidades com 3000 ou mais turmas); e
- (4) O número de turmas a selecionar em cada entidade cresce de acordo com o estrato de número de turmas da entidade.

Observa-se, também, na Tabela 1, que foi feita uma alocação do número de turmas a selecionar por município em cada entidade, indicando que na segunda etapa deveriam ser selecionados (ou incluídos) municípios onde as entidades da amostra tivessem turmas. É certo que nas entidades municipais, a sua seleção já implica a seleção do município de sua atuação. No entanto, para os outros tipos de entidade, a seleção das turmas diretamente do cadastro espalharia a amostra e aumentaria sobremaneira o seu custo.

Desta forma, controla-se o espalhamento da amostra e o custo de coleta, além de assegurar um grau desejável de espalhamento da amostra que garantisse a representação de todo o território nacional.

Tabela 1 – Dados populacionais e tamanho da amostra de entidades e turmas por estágio de seleção, segundo o tipo de entidade e a classe do número de turmas.

Tipo da entidade e classe do número de turmas	População			Primeira etapa da amostra			Segunda etapa da amostra	
	Nº de entidades	Nº de turmas	Nº de alunos	Nº de entidades a selecionar	Nº de turmas por entidade	Total de turmas a selecionar	Nº de turmas da entidade a selecionar por município	Número máximo de municípios na amostra
Total geral	640	101.307	2.006.601	86		802	182	
Estaduais	22	45.205	870.317	22		301		
Até 150 turmas	2	233	5.118	2	6	12	6	2
De 150 a 999 turmas	6	3.085	56.436	6	9	54	3	18
De 1.000 a 2.999 turmas	9	15.652	288.539	9	15	135	5	27
3.000 turmas ou mais	5	26.235	520.224	5	20	100	5	20
Municipais	567	12.578	250.460	33		172		
Até 29 turmas	457	4.572	87.083	13	3	39	3	13
De 30 a 99 turmas	89	4.521	90.311	7	4	28	4	7
De 100 a 149 turmas	13	1.564	33.041	5	5	25	5	5
150 turmas ou mais	8	1.921	40.025	8	10	80	10	8
ONG¹	46	41.072	841.567	26		299		
Até 99 turmas	17	708	15.162	7	6	42	3	14
De 100 a 499 turmas	18	4.204	83.022	8	9	72	3	24
De 500 a 2.999 turmas	7	8.363	165.283	7	15	105	5	21
3.000 turmas ou mais	4	27.797	578.100	4	20	80	5	16
IES²	5	2.452	44.257	5		30		
Até 10 turmas	1	2	7	1	2	2	2	1
De 11 a 50 turmas	2	69	1.317	2	4	8	4	2
51 turmas ou mais	2	2.381	42.933	2	10	20	5	4

¹ ONG é Organização Não-Governamental.

² IES é instituição de ensino superior.

Seleção da amostra do SBA 2005

A estratificação por tipo de entidade e classes de número de turmas, gerou estratos de seleção com quantitativos pequenos demais para usar a entidade como unidade primária de seleção, que foram incluídas na amostra com certeza.

Assim, a descrição do desenho da amostra utilizado será segmentada em quatro etapas: (1) seleção ou inclusão² da entidade; (2) seleção ou inclusão³ de municípios onde cada entidade da amostra atua; (3) seleção de turmas; e (4) seleção de alunos, alfabetizador e coordenador (ou pessoa responsável).

Nas duas primeiras etapas, a seleção, quando houve, foi feita com probabilidade proporcional ao tamanho, expresso em número de turmas. Na terceira etapa, adotou-se um esquema de amostragem inversa, descrito mais adiante, enquanto a seleção de alunos foi feita com equiprobabilidade.

Primeira etapa: tratamento das entidades

Das 640 entidades parceiras da edição 2005 do *Programa Brasil Alfabetizado*, 46 entidades⁴ foram incluídas com certeza na amostra e correspondem a estratos de seleção. As unidades primárias de amostragem nesses estratos de seleção correspondem, em princípio, aos municípios⁵.

Das 594 entidades restantes, havia uma previsão de seleção de 40 entidades. Para controlar o espalhamento geográfico da amostra, no entanto, decidiu-se fracionar as entidades por município de atuação, selecionando combinações de entidade e município dentro de um esquema de probabilidade proporcional ao número de turmas da entidade no município. Assim, as unidades primárias de amostragem são as combinações entre entidade e município, sempre que o tamanho desta combinação não implicar sua inclusão na amostra com certeza (caso em que as unidades primárias de amostragem são as turmas).

É fato, também, que para as entidades municipais o fracionamento já estava dado. Assim, as 25 entidades municipais a serem selecionadas correspondiam a 25 municípios. Neste caso, a unidade primária de amostragem correspondeu à entidade e o município deixou de ser unidade de seleção.

³ Como em todo procedimento de seleção por probabilidade proporcional ao tamanho, unidades muito grandes acabam sendo incluídas com certeza na amostra, o que evita a destruição do esquema probabilístico da amostra com a atribuição de probabilidades de inclusão maiores do que um para estas unidades.

No caso dos estratos de ONGs com menos de 500 turmas, as 15 entidades a selecionar, após fracionamento por município, corresponderam a 38 combinações de entidade e município, que foram as unidades primárias de amostragem.

Desta descrição, observa-se que o estrato de seleção pode ter sido: (1) o tipo de entidade e a classe de número de turmas; (2) a entidade; ou (3) a combinação de entidade e município. Conseqüentemente, a unidade primária de amostragem pode ter sido a entidade; a combinação entidade e município; ou a turma. Diante desta multiplicidade de situações, no processo de expansão da amostra serão incluídas variáveis que indicarão o estrato e unidade primária de amostragem, além do peso amostral.

Na Tabela 2 são indicados os estratos de seleção, as unidades primárias de amostragem que foram utilizados em princípio. Considera-se que em alguns estratos de seleção havia entidades (ou combinações de entidade e município) que foram incluídos com certeza na amostra. Pelos motivos já indicados, os números de estratos e de unidade primárias de amostragem, apresentados na Tabela 2, podem ser maiores do que os números esperados pela descrição das unidades de seleção, visto que os números apresentados já consideram situações especiais observadas durante a seleção da amostra.

Segunda etapa: tratamento dos municípios

Nos casos em que a entidade corresponde ao estrato de seleção, os municípios correspondem às unidades primárias de amostragem e foram selecionados com probabilidade proporcional ao número de turmas.

Como indicado anteriormente, alguns municípios com número expressivo de turmas foram incluídos com certeza na amostra e a unidade primária de amostragem, nesse caso, passou a ser a turma. A Tabela 2 indica, em notas de pé de tabela, os casos onde isto ocorreu.

⁴ Entidades estaduais, entidades municipais com 150 turmas ou mais, ONGs com 500 turmas ou mais, e IES, com indicado na Tabela 1.

⁵ Há casos onde o município, seja por ser único (Secretaria de Educação do Distrito Federal), seja por ter um número de turmas que geraria uma probabilidade de inclusão maior que 1, é incluído com certeza na amostra e a unidade primária de amostragem passa a ser a turma.

Tabela 2 – Descrição e número de estratos e unidades primárias de amostragem, segundo o tipo de entidade e a classe do número de turmas.

Tipo da entidade e classe do número de turmas	Estrato de seleção		Unidade primária de amostragem		
		Descrição	Número	Descrição	Números
Total geral			62		303
Estaduais					
Até 150 turmas ¹	Cada entidade		2	Turma	24
De 150 a 999 turmas	Cada entidade		8	Município	18
De 1.000 a 2.999 turmas	Cada entidade		9	Município	27
3.000 turmas ou mais ²	Cada entidade		6	Município	24
Municipais					
Até 29 turmas	Entidades com até 29 turmas		1	Entidade	13
De 30 a 99 turmas	Entidades com 30 a 99 turmas		1	Entidade	7
De 100 a 149 turmas	Entidades com 100 a 149 turmas		1	Entidade	5
150 turmas ou mais	Cada entidade		8	Turma	80
ONG¹					
Até 99 turmas ³	Entidades com até 99 turmas		6	Entidade/município	24
De 100 a 499 turmas ⁴	Entidades com 100 a 499 turmas		4	Entidade/município	30
De 500 a 2.999 turmas	Cada entidade		7	Município	21
3.000 turmas ou mais	Cada entidade		4	Município	16
IES²					
Até 10 turmas ⁵	Cada entidade		1	Turma	2
De 11 a 50 turmas ⁵	Cada entidade		2	Turma	8
51 turmas ou mais	Cada entidade		2	Município	4

¹ As duas entidades atuavam em um só município.

² Um município incluído com certeza na amostra.

³ Cinco municípios incluídos com certeza na amostra.

⁴ Três municípios incluídos com certeza na amostra.

⁵ As entidades atuavam em um só município.

Terceira etapa: seleção das turmas

A Tabela 2 indica os casos em que as turmas formam unidades primárias de amostragem. Nos demais casos, as turmas formam unidades secundárias de amostragem.

Aparentemente, o problema de seleção da amostra de turmas estaria resolvido com a estratificação e alocação do tamanho da amostra, descrita na seção 4, e com a seleção das entidades e dos municípios. Para tanto, bastaria proceder à seleção das turmas das entidades/ municípios da amostra.

No entanto, como um dos principais objetivos da pesquisa é determinar o valor adicionado pela alfabetização, é necessário selecionar turmas de modo a entrevistar alunos no início e no fim do curso. E como as datas de início e término das turmas não são dados confiáveis do SBA, por não serem atualizadas pelas entidades parceiras, um problema adicional surgiu na etapa de seleção das turmas: como identificar a subpopulação de turmas com início próximo da data de coleta.

Três alternativas são sugeridas na literatura para lidar com subpopulações: (1) aumentar o tamanho da amostra para compensar a seleção de unidades que não pertencem à subpopulação de interesse; (2) realizar um procedimento de varredura (screening), para identificar as unidades da subpopulação de interesse; e (3) adotar um procedimento de amostragem inversa.

Em sua revisão sobre os métodos de amostragem de subpopulações, Graham Kalton e Dallas W. Anderson (Kalton & Anderson, 1986) mostraram ser muito difícil estabelecer a proporção de aumento do tamanho da amostra para compensar os casos de seleção de unidades não-pertencentes à subpopulação de interesse, além de compensar as não-repostas. Sobretudo porque as proporções de unidades não-pertencentes à subpopulação e de não-resposta variam por conglomerado amostrado (no caso por entidade ou município).

Diante disto, os autores sugerem introduzir uma etapa adicional de varredura exaustiva das unidades populacionais (no caso turmas) para todos os conglomerados amostrados. No caso, isto implicaria entrar em contato com todas as entidades selecionadas para atualizar as datas de início e término de suas turmas cadastradas, o que teria um impacto não-desprezível – e originalmente não-previsto – sobre o custo de coleta.

A última alternativa, amostragem inversa, é um método originalmente proposto por Haldane (1945), que consiste em verificar quantas unidades precisam

ser observadas (n) para que sejam obtidos k sucessos (entrevistas realizadas em unidades de interesse). Ou seja, é um método no qual o número de unidades com entrevista realizada é fixado e as unidades são visitadas (ou são contatadas), respeitando uma ordem aleatória previamente estabelecida, até que o número prefixado de entrevistas realizadas seja atingido. No caso, as turmas das entidades, nos municípios incluídos na amostra, seriam ordenadas de modo aleatório e contatadas respeitando esta ordenação até que fosse atingido o número prefixado de turmas a serem entrevistadas.

Desta descrição, depreende-se que a amostragem inversa corresponde a uma varredura inteligente, feita por amostragem e limitada pela obtenção do número prefixado de entrevistas, sendo, portanto, a alternativa de menor impacto sobre o custo de coleta.

Assim, todas as turmas das entidades e nos municípios da amostra foram coordenadas aleatoriamente, por meio do algoritmo de Hájek (1960), estabelecendo a ordem de visita (ou contato) que deveria ser respeitada até atingir o tamanho de turmas entrevistadas prefixado para o caso. No anexo 1, é apresentado um exemplo fictício da planilha de seleção de turmas em um município exemplo. Os dados da entidade e de algumas de suas turmas são reais, mas a entidade e o município não pertencem à amostra.

Quarta etapa: tratamento das unidades a entrevistar

Uma vez selecionadas as turmas, são entrevistados três tipos de unidade informante: aluno, alfabetizador e coordenador (ou pessoa responsável pelo curso). Para tanto, são usados quatro questionários, sendo dois aplicáveis aos alunos, além do teste cognitivo a ser aplicado entre alunos para medir seu conhecimento ao início e ao final do curso.

É certo que a seleção das turmas implica a seleção de seu alfabetizador e de seu coordenador. Mas no caso dos alunos, foi decidido aplicar um questionário socioeconômico completo a uma amostra de alunos e um questionário reduzido aos demais alunos. A seleção dos alunos que deverão responder ao questionário completo foi feita pela seleção equiprovável dos números dos alunos na lista de presença do alfabetizador.

Descrição do arquivo da amostra do SBA 2005

A amostra de turmas selecionadas foi fornecida por meio de cinco arquivos em formato MS-Excel, um para cada macrorregião do País, e com uma planilha para cada unidade da federação da macrorregião.

Em cada planilha relativa a uma unidade da federação existem três tipos de informação: (1) dados do município; (2) dados da entidade; e (3) dados das turmas e de seu alfabetizador. A descrição das colunas das planilhas é feita em função do tipo de informação.

Dados do município

Colunas	Descrição	Documentação/instruções
A	Contém o texto MUNICÍPIO:	
B	Nome do município e, entre parênteses, o Código do IBGE	

Dados da entidade

Linhas	Colunas	Descrição	Documentação/instruções
1	A	Texto do cabeçalho: CNPJ Nº Texto do cabeçalho: nome da entidade	O número de tiras é impresso após o texto indicado à esquerda, visto que este número varia por entidade e município.
1	B	Número de turmas a entrevistar da entidade do município	
2	A	CNPJ da entidade	
	B	Razão social da entidade	
3	A	Texto: "Dados da entidade relativos à sua sede na UF": seguido pela sigla da UF	
A linha 4 contém o cabeçalho e a linha 5 os dados indicados no cabeçalho			
4	A	Município	Trata-se do código IBGE do município
	B	Endereço da entidade	
	C	Complemento do endereço	
	D e E	Colunas em branco	
	F	Bairro	Nome do Bairro
	G	Nº do CEP	
	H	DDD	Código de DDD da entidade
	K	Endereço de e-mail	Endereço do correio eletrônico da entidade

Dados das turmas

Antes de apresentar os dados das turmas da entidade no município, são impressas quatro linhas com as seguintes instruções:

- (1) Contate as turmas abaixo, respeitando a ordem de contato, até que consiga entrevista realizada no número de turmas indicado acima (coluna B da linha 1 dos dados da entidade).
- (2) Registre, para cada turma contatada, o resultado do contato com seu código respectivo (sem esta informação não será possível expandir a amostra).
- (3) Esgotando a relação de turmas abaixo sem atingir o número de turmas a entrevistar, contate a Science para ser feita uma substituição de município e/ou entidade.
- (4) Registre a identificação do questionário, substituindo XX pelo seqüencial da turma entrevistada, iniciando em 01 para cada entidade em cada município.

Em seguida, são fornecidos os dados das turmas e de seus alfabetizadores, além de colunas para registrar o resultado da entrevista e a identificação do questionário. A primeira linha contém o cabeçalho, descrito abaixo, e as demais fornecem os dados correspondentes.

Colunas	Descrição	Documentação/instruções
A	Código da turma	Código MEC da turma
C	Resultado do contato	Coluna em branco para ser indicado o resultado do contato em texto livre, conforme terceira instrução indicada acima.
D	Código	Coluna em branco para ser indicado o resultado do contato, de acordo com uma lista de códigos a ser preparada pelo João Pedro
E	Ordem	Coluna que indica a ordem em que as turmas devem ser contatadas, conforme a primeira instrução indicada acima
F	Identificação questionário	Coluna com parte pré-impressa , com o XX devendo ser substituído pelo número seqüencial da turma entrevista, conforme quarta instrução acima

Complemento do endereço da turma		
G	Nº do CEP	
H	DDD	Código DDD do telefone da turma
I	Telefone	Número do telefone da turma
J	Nº do FAX	Número do FAX da turma
Alfabetizador, dados de início de fim das aulas da turma		
K	Nome do alfabetizador	
L	Início previsto	Data de início previsto das aulas da turma
M	Fim previsto	Data de fim previsto das aulas da turma
N	Início efetivo	Data de início efetivo das aulas da turma
O	Fim efetivo	Data de fim efetivo das aulas da turma
P a V	Dias da semana previstos com aula	Valores: 2ª , 3ª , 4ª , 5ª ,6ª , Sáb e Dom
Característica da turma		
W	Área	Valores: U- urbana, R- rural
X	Locação	Valores: 5- casa do alfabetizador; 10- Escola; 15- Igreja; 20- Salão Paroquial; 25 - Sociedade de amigos do bairro; 99 - outro
Y	Situação	Só ocorre uma : turma ativa
Z	Número de alunos	Número de alunos da turma
Dados do alfabetizador		
AA	Sexo	Valores: F-feminino, M-masculino
AB	Nº do CEP	
AC	Idade (31/3/06)	Idade do alfabetizador, 30 de março de 2006
AD	Escolaridade	Valores: 02 - Antigo primário 03 - Antigo ginásio 04 - Antigo Clássico. Científico, etc 05 - Ensino Fundamental ou 1º Grau 06 - Ensino Médio ou 2º Grau 07 - Superior - Graduação 08 - Mestrado ou doutorado

AE	Profissão	Valores: 5 – Alfabetizador popular 10- Aposentado 15- Desempregado 20 – Estudante 25- Professor alfabetizador 30 – Professor Educação Infantil 35 – Professor Ensino Fundamental 40 – Professor Ensino Médio 45- Professor universitário 50 – Trabalhador Rural 55 – Trabalhador Urbana 99 - Outra
AF	Nacionalidade: Os dados de nacionalidade do alfabetizador são:	

10 - Brasil 21 - Argentina 22 - Bolívia 23 - Chile 24 - Paraguai 25 - Uruguai 26 - Colômbia 27 - Equador 28 - Antigua e Dep Barbuda 29 - Antilhas Holandesas 30 - Alemanha 31 - Bélgica 32 - Grã Bretanha 33 - Aruba 34 - Canadá 35 - Espanha 36 - Estados Unidos da América 37 - França 38 - Suíça 39 - Itália 40 - Comunidade das Bahamas 41 - Japão 42 - China 43 - Coréia 44 - Barbados 45 - Portugal 46 - Belize 47 - Ilhas Turks e Caicos 48 - Outros Latino Americanos 49 - Outros Asiáticos 50 - Outros 51 - Costa Rica 52 - Cuba 53 - Curaçao 54 - Comunidade Dominicana 55 - República Dominicana 56 - República de El Salvador 57 - Estados Associados das Antilhas	58 - Ilhas Falklands (Malvinas) 59 - Granada 60 - Ilhas Guadalupe Canas 61 - Guatemala 62 - Republica do Haiti 63 - Honduras Britânicas 64 - Honduras 65 - Ilhas Serranas 66 - Jamaica 67 - Ilhas Malvinas 68 - Martinica 69 - Ilha Milhos 70 - Monte Serrat 71 - Nicarágua 72 - Panamá 73 - Panamá Zona do Canal 74 - Porto Rico 75 - Quitasueno 76 - Roncador 77 - Santa Lucia 78 - São Cristóvão 79 - São Vicente 80 - Ilhas Turca 81 - Ilhas Virgens Britânicas 82 - Ilhas Virgens Ameri 84 - Groelândia 83 - Bermudas 85 - México 86 - St Pierre et Miquelon 87 - Guiana Francesa 88 - República Guiana 89 - Peru 90 - Suriname 91 - Trinidad e Tobago 92 - Venezuela 93 - Albania 94 - Andorra 95 - Áustria 96 - Bulgária
---	--

97 - Chipre	152 - Karachaevochekess
98 - Dinamarca	153 - Khakass
99 - Eire	154 - Komi
100 - Escócia	155 - Letônia
101 - Ilhas Feroes	156 - Lituânia
102 - Finlândia	157 - Mari
103 - Gibraltar	158 - Moldávia
104 - Grécia	159 - Ossetia Setentrional
105 - Holanda	160 - Quirquistão
106 - Hungria	161 - Tadjiquistão
107 - Ilhas Baleares	162 - Tartaria
108 - Ilhas Cosmoledo	163 - Turcomenistão
109 - Ilhas do Canal	164 - Tuvin
110 - Inglaterra	165 - Ucrânia
111 - Irlanda do Norte	166 - Udmurt
112 - Irlanda	167 - União Soviética
113 - Islândia	168 - Usbequistão
114 - Iugoslávia	169 - Yakut
115 - Liechtenstein	170 - Abissínia
116 - Luxemburgo	171 - Açores
117 - Ilhas de Man	172 - Afar Francês
118 - Mônaco	173 - República da África do Sul
119 - Noruega	174 - Alto Volta
120 - Republica de Malta	175 - Angola
121 - País de Gales	176 - Argélia
122 - Paises Baixos	177 - Bechuana
123 - Polônia	178 - Benin
124 - Romênia	179 - Botswana
125 - São Marinho	180 - Burundi
126 - Suécia	181 - Camarões
127 - Svalbard e Jan Mayer Islands	182 - Ceuta e Lelilla
128 - Tchecoslováquia	183 - Chade
129 - Estado da Cidade do Vaticano	184 - Ilhas Comores
130 - Croácia	185 - Congo
131 - Sérvia	186 - Costa do Marfim
132 - Eslovênia	187 - Daome
133 - República da Macedônia	188 - Djibuti
134 - Bósnia Herzegovina	189 - República Árabe do Egito
135 - República Tcheca	190 - Etiópia
136 - Eslováquia	191 - República do Gabão
137 - Montenegro	192 - Gâmbia
138 - Azerbaijão	193 - Gana
139 - Bashkista	194 - Gaza
140 - República da Bielorrusia	195 - Guiné
141 - Buryat	196 - Guiné Equatorial
142 - Carélia	197 - Infni
143 - Cazaquistão	198 - Ascensão e Tristão da Cunha Islands
144 - Chechen Ingusth	199 - Ilhas Canárias
145 - Chuvash	200 - Lesoto
146 - Dagestão	201 - Libéria
147 - Estônia	202 - Líbia
148 - Geórgia	203 - Madeira
149 - Gorno Altai	204 - Malawi
150 - Kabardino Balkar	205 - Madagascar
151 - Kalmir	206 - Mali

- 207 - Marrocos
 208 - Maurício
 209 - Maurítânia
 210 - Moçambique
 211 - Nguane
 212 - República do Niger
 213 - Nigéria
 214 - Papua Nova Guiné
 215 - Praças Norte Africanas
 216 - Protetor do Sudoeste Africano
 217 - Quênia
 218 - República Centro Africana
 219 - Reunião
 220 - Rodésia
 221 - Ruanda
 222 - Saara Espanhol
 223 - Santa Helena
 224 - São Tomé e Príncipe
 225 - Seychelles
 226 - Serra Leoa
 227 - Somália República
 228 - Suazilândia
 229 - Sudão
 230 - Tanganica
 231 - Território Britânico do Oceano Índico
 232 - Transkei
 233 - Togo
 234 - Tunísia
 235 - Uganda
 236 - Zaire
 237 - Zâmbia
 238 - Burkina Fasso
 239 - Zimbábue
 240 - Namíbia
 241 - Afeganistão
 242 - Arábia Saudita
 243 - Bahrain
 244 - Birmânia
 245 - Brunei
 246 - Bhutan
 247 - Catar
 248 - Ceilão
 249 - China Taiwan
 250 - Coveite
 251 - Emirados Árabes Unidos
 252 - Filipinas
 253 - Hong Kong
 254 - Iêmen
 255 - Índia
 256 - Indonésia
 257 - Ira
 258 - Iraque
 259 - Israel
 260 - Jordânia
 261 - Kmer Camboja
 262 - Kuwait
 263 - Laos
 264 - Líbano
 265 - Macau
 266 - Malásia
 267 - Ilhas Maldivas
 268 - Mascate
 269 - Mongólia
 270 - Nepal
 271 - Oman
 272 - Palestina
 273 - Paquistão
 274 - Ilhas Riquiu
 275 - Singapura (ou Cingapura)
 276 - Sequin
 277 - Síria
 278 - SRI-Lanka
 279 - Tailândia
 280 - Estado Tréguia
 281 - Turquia
 282 - Vietnã do Norte
 283 - Vietnã do Sul
 284 - Mianmar
 285 - Arquipélago Manahiki
 286 - Ilhas Midway
 287 - Ashmore e Cartier
 288 - Austrália
 289 - Bismark (Arquipélago)
 290 - Ilhas Cook
 291 - República de Fiji
 292 - Guan
 293 - Ilhas Baker
 294 - Ilhas Cantão e Enderburg
 295 - Ilhas Carolina
 296 - Ilhas do Pacífico
 297 - Ilhas Christmas
 298 - Ilhas Gilbert
 299 - Ilhas Howland e Jarvis
 300 - Ilhas Johnston e San
 301 - Ilhas Kingman Reef
 302 - Ilhas Macquaire
 303 - Ilhas Mariana
 304 - Ilhas Marshall
 305 - Ilhas Macdonal e Heard
 306 - Ilhas Nieu
 307 - Ilhas Norfolk
 308 - Ilhas Palau
 309 - Ilhas Salomão
 310 - Ilhas Torkelau
 311 - Ilhas Wake
 312 - Kalimantan
 313 - Ilhas Linha
 314 - Nauru
 315 - Ilhas Nova Caledônia
 316 - Nova Guiné

- 317 - Nova Zelândia
- 318 - Ilhas Novas Hebridas
- 319 - Ilhas Território de Papua
- 320 - Ilhas de Páscoa
- 321 - Ilhas Pitcairin
- 322 - Polinésia Francesa
- 323 - Sabah
- 324 - Samoa Americana (ou Samoa Oriental)
- 325 - Samoa Ocidental
- 326 - Ilhas Santa Cruz
- 327 - Sarawak
- 328 - Território de Cocos
- 329 - Timor (ou Timor Leste)
- 330 - Tongas
- 331 - Tuvalu
- 332 - Ilhas Wallis e Futuna
- 333 - Território Antártico Britânico
- 334 - Antártica Francesa
- 335 - Território Antártico da Austrália
- 336 - Antártico Chileno
- 337 - Antártico Argentino
- 338 - Antártico Norueguês
- 339 - Apátrida
- 340 - Dependência de Ross (Nova Zelândia)
- 341 - Terras Austrais
- 342 - Bangladesh
- 343 - Cabo Verde
- 344 - Guiné Bissau
- 345 - Iêmen do Sul
- 346 - Kara Kalpak
- 347 - Armênia
- 348 - Rússia
- 349 - Senegal
- 350 - Tanzânia

Referências

- Cochran W.G., 1977. *Sampling Techniques*, 3rd edition. New York: John Wiley & Sons.
- Hájek, J., 1960. *Limiting distribution in simple random sampling from a finite population*. Publications of the Mathematics Institute of the Hungarian Academy Science, 5:361-374.
- Hedlin D., 2000. *A procedure for stratification by an extended Ekman rule*. Journal of Official Statistics, 16(1):15-29.
- Kalton G. & Anderson, D.W.,1986. *Sampling rare populations*. Journal of the Royal Statistical Society, Series A, 149, part 1:65-82
- SCIENCE, 2006. *Avaliação do Programa Brasil Alfabetizado: Relatório técnico descritivo do estudo para estratificação e desenho da amostra a ser usada para avaliação do impacto dos projetos aprovados no âmbito da Resolução de 2005 do Programa Brasil Alfabetizado*. Relatório técnico, Produto C, Rio de Janeiro:

Anexos

Anexo 1: Planilha de seleção de turmas

A		B		C	D	E
MUNICÍPIO:	Salesópolis	(Código/BGE: 3545001)				
CNPJ Nº	Nome da entidade	Número de turmas a entrevistar da entidade no município: 3				
06205515000104	ASSOCIAÇÃO PAULISTA DE EDUCAÇÃO POPULAR					
Dados da entidade relativos à sua sede na UF: SP						
Município	Endereço da entidade	Complemento				
3518800	RUA RIO PARDO,215	NULL				
Contate as turmas abaixo, respeitando a ordem de contato, até que consiga entrevista realizada no número de turmas indicado acima.						
Registre, para cada turma contatada, o resultado do contato com seu código respectivo (sem esta informação não será possível expandir a amostra).						
Esgotando a relação de turmas abaixo sem atingir o número de turmas a entrevistar, contate a Science para ser feita uma substituição de município e/ou entidade.						
Registre a identificação do questionário, substituindo XX pelo seqüencial da turma entrevistada, iniciando em 01 para cada entidade em cada município.						
Código turma	Endereço da turma		Resultado do contato	Código	Ordem	
359032	RUA DO CASARÃO, Nº 13				1	
358089	PRAÇA PE. SÃO JOÃO MENENDES, S/N				2	
351658	RUA MELO NUNES, 80				3	
349630	RUA MARECHAL DEODORO, 146, JD. PINHEIROS				4	
375369	GONÇALO MOREIRA, 374				5	
375457	RUA PARTICULAR, 108				6	
349826	R ELOY DOMINGUES DA SILVA, 423				7	
370743	RUA MANOEL VELOSO DA COSTA, 46 ALTOS, JD. VILA CARRÃO				8	
354099	AV TRISTE BERRANTE, 45				9	
351621	RUA MELO NUNES, 80				10	
349170	RUA PADRE VIRGILIO CAMPELO, 279				11	
344953	AVENIDA DOS TEXTEIS, 580				12	
346310	R. WASHINGTON LUIS, 236				13	
354098	RUA TRISTE BERRANTE, 45				14	
346306	RUA SARA KUBISCHEK 185				15	

Anexo 1: Planilha de seleção de turmas

F	G	H	I	J	K	L	M	N	O	P	Q
Bairro	Nº do CEP	DDD	Telefone	Nº do FAX	Endereço de e-mail						
JARDIM JACY	7262000	11	64991518	64430135	a.p.e.p@terra.com.br						
Identificação questionário	Nº do CEP	DDD	Telefone	Nº do FAX	Nome do alfabetizador	Início previsto	Fim previsto	Início efetivo	Fim efetivo	2ª	3ª
137XX	8970000				ELIZIA BARBOZA LARANJO	06/02/2006	06/10/2006	20/02/2006	20/10/2006	S	S
137XX	8970000				RITA ABREU DE SOUZA	06/02/2006	06/10/2006	20/02/2006	20/10/2006	S	S
138XX	4408000				MARIA GRACIETE DA SILVA	23/01/2006	28/09/2006	23/01/2006	28/09/2006	S	S
138XX	8565520	11	46366564		TINELLI DAMASCENA PEREIRA DOS SANTOS	15/01/2005	20/09/2005	15/01/2005	20/09/2005	S	S
138XX	4658009				MARIA ANTONIA DA SILVA RUAS	10/04/2006	10/10/2006	10/04/2006	10/10/2006	S	S
138XX	8253000				MARIA APARECIDA DA SILVA	20/04/2006	20/10/2006	20/04/2006	20/10/2006	S	S
138XX	4823410				ADRIANA ROSCHEL	16/01/2006	16/09/2006	16/01/2006	16/09/2006	N	N
138XX	8255000	11	67535911		BARBARA NASCIMENTO DA SILVA	23/02/2006	23/10/2006	23/02/2006	23/10/2006	S	N
138XX	8345310				SUELI GONÇALVES BASTOS	30/01/2006	30/09/2006	30/01/2006	30/09/2006	S	S
138XX	4338000				MARIA GRACIETE DA SILVA	23/01/2006	28/09/2006	23/01/2006	28/09/2006	S	S
138XX	8131310				ANTONIA JOELMA RODRIGUES	15/01/2006	15/09/2006	15/01/2006	15/09/2006	S	S
138XX	8490600	11	62826433		GILDETE SOARES AYRES	15/01/2006	20/09/2006	15/01/2006	20/09/2006	S	S
138XX	1033010	11	33138701	33138701	THALITA SILVA STEIN SALLES	06/01/2006	06/09/2006	06/01/2006	06/09/2006	S	S
138XX	8345310				SUELEN DE SOUZA MENDONÇA	30/01/2006	30/09/2006	30/01/2006	30/09/2006	N	S
138XX	8490600				GILDETE SOARES AYRES	06/01/2006	06/09/2006	06/01/2006	06/09/2006	S	S

Estudo para
estratificação e desenho
da amostra utilizada
para avaliação do
impacto dos projetos
aprovados no âmbito
da Resolução de 2005
do *Programa Brasil
Alfabetizado*

Mauricio Teixeira Leite
de Vasconcellos¹

Introdução

Este capítulo apresenta um estudo para estratificação e desenho de uma amostra probabilística de turmas e alunos da edição 2005 do *Programa Brasil Alfabetizado* (PBA), selecionada com base no cadastro de entidades, turmas, alfabetizadores e alfabetizandos, mantido pelo Ministério da Educação e conhecido como Sistema do Brasil Alfabetizado (SBA).

Este estudo descreve a geração do cadastro de seleção de turmas, apresenta dados do SBA 2005 e do cadastro de seleção, discute alternativas para o desenho da amostra, e apresenta os estudos relacionados à estratificação das unidades constantes do cadastro de seleção.

Geração e exploração do cadastro de seleção

O SBA 2005 é um sistema com arquivos que são dinamicamente alterados pelos parceiros do PBA. A versão utilizada para geração do cadastro de seleção da

¹ Professor da Escola Nacional de Ciências Estatísticas, Doutor em Saúde Pública – ENSP/FIOCRUZ

amostra refere-se a uma fotografia do SBA do dia 4 de abril de 2006 (data de referência do cadastro de seleção), que continha um total de 640 entidades ativas, com 108.794 turmas e 2.043.794 alunos.

No entanto, na data de referência do cadastro de seleção algumas turmas estavam canceladas ou não tinham alunos a elas associados. Assim, estas turmas foram eliminadas e o cadastro de seleção ficou com 101.307 turmas válidas (93,1% do total de turmas) e 2.006.601 alunos (98,2% do total de alunos). Os percentuais entre parêntesis indicam que a maior parte das turmas eliminadas na montagem do cadastro de seleção eram turmas sem alunos associados.

A distribuição, no cadastro de seleção, do número de turmas por macrorregião e unidade da federação, apresentada na Tabela 1, mostra forte concentração no Nordeste, com cerca de 68% das turmas do SBA 2005. Na observação por tipo de entidade, observa-se que o Nordeste só não tem mais da metade das turmas no caso de Instituições de Ensino Superior (IES), caso em que a região Norte concentra cerca de 57% das turmas deste tipo de entidade.

Tabela 1 – Número de turmas por tipo de entidade, segundo as macrorregiões e unidades da federação.

Macrorregiões e unidades da federação	Número de turmas				
	Total	Tipo de entidade			
		Estaduais	Municipais	ONG ¹	IES ²
Total nacional	101.307	45.205	12.578	41.072	2.452
Norte	9.047	2.058	1.716	3.878	1.395
Rondônia	531	250	105	176	0
Acre	1.449	1.226	86	137	0
Amazonas	2.043	0	39	609	1.395
Roraima	91	0	25	66	0
Pará	3.605	0	1.295	2.310	0
Amapá	379	142	72	165	0
Tocantins	949	440	94	415	0
Nordeste	68.385	33.257	6.669	27.473	986
Maranhão	7.824	1.796	1.129	3.913	986
Piauí	9.262	4.539	678	4.045	0
Ceará	14.605	10.275	86	4.244	0
Rio Grande do Norte	5.873	3.999	438	1.436	0
Paraíba	7.702	4.136	831	2.735	0
Pernambuco	6.631	1.520	1.370	3.741	0
Alagoas	3.771	1.157	260	2.354	0
Sergipe	3.563	2.549	285	729	0
Bahia	9.154	3.286	1.592	4.276	0
Sudeste	12.326	2.277	3.138	6.840	71
Minas Gerais	3.788	1.697	1.086	955	50
Espírito Santo	342	0	77	265	0
Rio de Janeiro	4.232	580	469	3.164	19
São Paulo	3.964	0	1.506	2.456	2
Sul	6.502	4.652	723	1.127	0
Paraná	3.394	2.572	318	504	0
Santa Catarina	1.027	889	121	17	0
Rio Grande do Sul	2.081	1.191	284	606	0
Centro-Oeste	5.047	2.961	332	1.754	0
Mato Grosso do Sul	2.146	1.944	64	138	0
Mato Grosso	824	362	128	334	0
Goiás	1.535	564	140	831	0
Distrito federal	542	91	0	451	0

¹ ONG é Organização Não-Governamental.

² IES é Instituição de Ensino Superior.

Na Tabela 2, a distribuição do número de alunos por tipo de entidade, macrorregião e unidade da federação é fornecida e indica o mesmo padrão espacial da distribuição do número de turmas.

Tabela 2 – Número de alunos por tipo de entidade, segundo as macrorregiões e unidades da federação.

Macrorregiões e unidades da federação	Número de alunos				
	Total	Tipo de entidade			
		Estaduais	Municipais	ONG ¹	IES ²
Total nacional	2.006.601	870.317	250.460	841.567	44.257
Norte	170.841	36.040	34.145	78.957	21.699
Rondônia	9.341	3.886	1.881	3.574	0
Acre	25.449	21.077	1.672	2.700	0
Amazonas	35.752	0	1.076	12.977	21.699
Roraima	1.467	0	296	1.171	0
Pará	72.513	0	25.877	46.636	0
Amapá	8.282	3.097	1.492	3.693	0
Tocantins	18.037	7.980	1.851	8.206	0
Nordeste	1.394.413	666.225	138.643	568.311	21.234
Maranhão	164.654	37.179	25.087	81.154	21.234
Piauí	175.193	83.862	11.632	79.699	0
Ceará	291.986	205.021	1.627	85.338	0
Rio Grande do Norte	120.795	80.516	9.809	30.470	0
Paraíba	163.366	89.452	17.963	55.951	0
Pernambuco	133.832	28.979	28.816	76.037	0
Alagoas	82.262	24.810	5.489	51.963	0
Sergipe	72.952	55.033	5.298	12.621	0
Bahia	189.373	61.373	32.922	95.078	0
Sudeste	236.691	38.074	59.230	138.063	1.324
Minas Gerais	67.788	25.600	20.719	20.452	1.017
Espírito Santo	5.948	0	1.285	4.663	0
Rio de Janeiro	86.456	12.474	9.991	63.691	300
São Paulo	76.499	0	27.235	49.257	7
Sul	110.210	77.294	11.954	20.962	0
Paraná	61.392	46.278	5.776	9.338	0
Santa Catarina	15.097	13.182	1.701	214	0
Rio Grande do Sul	33.721	17.834	4.477	11.410	0
Centro-Oeste	94.446	52.684	6.488	35.274	0
Mato Grosso do Sul	35.930	31.749	1.277	2.904	0
Mato Grosso	17.106	7.671	2.492	6.943	0
Goiás	29.262	11.243	2.719	15.300	0
Distrito federal	12.148	2.021	0	10.127	0

¹ ONG é Organização Não-Governamental.

² IES é Instituição de Ensino Superior.

A distribuição do número de turmas por município² indica que o conjunto de turmas válidas do SBA 2005 abrange 3.607 municípios do País, o que representa um média de 28 turmas por município.

No entanto, a distribuição do número de turmas por município apresenta forte assimetria positiva (ou à direita), com alta concentração do número de turmas em 17 municípios, que são listados abaixo, com indicação do seu número de turmas válidas:

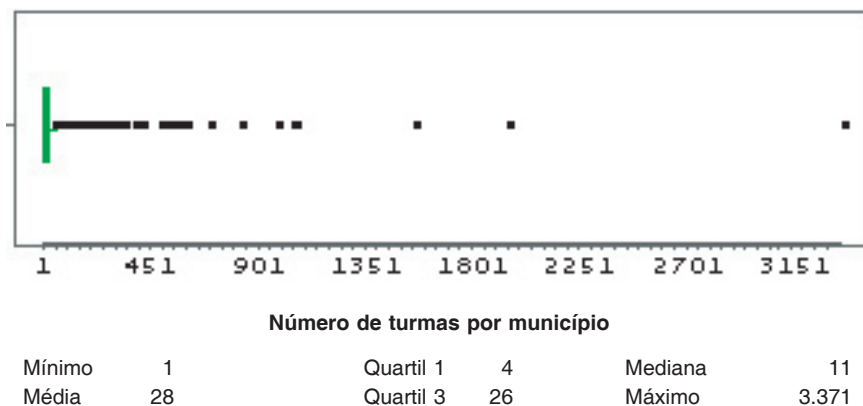
Município	Nº de turmas	Município	Nº de turmas
Fortaleza	3.371	São Luís	1.965
São Paulo	1.572	Natal	1.070
Teresina	1.057	Campina Grande, PB	993
Nova Iguaçu, RJ	838	Rio de Janeiro	708
Tauá, CE	615	Recife	606
Caxias, MA	604	Belém	567
Salvador	556	Brasília	542
Juazeiro do Norte, CE	540	Campo Grande	524
Maceió	500		

Por outro lado, a quantidade de municípios com poucas turmas não é desprezível. Existem 1.184 municípios com até cinco turmas (33,8% dos municípios): 331 com 1 turma; 286 com 2 turmas; 214 com 3 turmas; 192 com 4 turmas; e 161 municípios com 5 turmas.

Desta forma, o gráfico de caixa do número de turmas por município (Figura 1) tem uma forma que não permite ver diferença entre os seus três primeiros quartis.

² A base de dados com a distribuição do número de turmas por município não será apresentada neste estudo, e será fornecida por meio magnético.

Figura 1 – Gráfico de caixa e estatísticas da distribuição do número de turmas por município.



Isto significa que qualquer desenho de amostra que considere o município como unidade de seleção deve usar probabilidades desiguais. Além disto, existem erros de codificação do código de município no SBA 2005, onde entidades municipais mantêm turmas fora de seu município de atuação e, em alguns casos, em outros estados.

A distribuição do número de alunos por turma (Tabela 3) mostra turmas de 1 a 50 alunos. Em termos médios, as turmas têm tamanho variando entre 11,8 e 27,6 alunos, com um valor médio atípico de 3,5 alunos por turma em São Paulo, para IES.

Tabela 3 – Mínimo, média e máximo do número de alunos por turma por tipo de entidade, segundo as unidades da federação.

Unidades da Federação	Estaduais			Municipais			ONG ¹			IES ²		
	Min.	Média	Max.	Min.	Média	Max.	Min.	Média	Max.	Min.	Média	Max.
Total nacional	1	19,3	50	1	19,9	48	1	20,5	50	1	18,0	49
Rondônia	5	15,5	33	7	17,9	27	6	20,3	28			
Acre	5	17,2	25	8	19,4	25	10	19,7	25			
Amazonas				9	27,6	47	1	21,3	25	1	15,6	25
Roraima				6	11,8	20	1	17,7	28			
Pará				1	20,0	46	2	20,2	25			
Amapá	11	21,8	26	11	20,7	29	4	22,4	25			
Tocantins	1	18,1	25	6	19,7	28	5	19,8	25			
Maranhão	1	20,7	25	1	22,2	35	1	20,7	29	1	21,5	49
Piauí	1	18,5	25	1	17,2	32	1	19,7	35			
Ceará	1	20,0	39	8	18,9	25	1	20,1	26			
Rio Grande do Norte	5	20,1	39	2	22,4	27	2	21,2	29			
Paraíba	5	21,6	25	1	21,6	39	1	20,5	26			
Pernambuco	1	19,1	25	1	21,0	31	1	20,3	44			
Alagoas	3	21,4	25	7	21,1	44	1	22,1	25			
Sergipe	1	21,6	50	5	18,6	25	3	17,3	26			
Bahia	1	18,7	26	1	20,7	46	1	22,2	27			
Minas Gerais	1	15,1	25	5	19,1	48	1	21,4	27	1	20,3	25
Espírito Santo				7	16,7	25	1	17,6	25			
Rio de Janeiro	10	21,5	34	7	21,3	36	1	20,1	27	10	15,8	25
São Paulo				1	18,1	48	1	20,1	29	2	3,5	5
Paraná	1	18,0	36	5	18,2	37	1	18,5	25			
Santa Catarina	5	14,8	47	2	14,1	26	9	12,6	24			
Rio Grande do Sul	1	15,0	25	1	15,8	25	5	18,8	25			
Mato Grosso do Sul	2	16,3	27	10	20,0	25	10	21,0	25			
Mato Grosso	3	21,2	27	5	19,5	25	2	20,8	25			
Goiás	1	19,9	43	9	19,4	36	1	18,4	50			
Distrito federal	4	22,2	28				6	22,5	26			

¹ ONG é Organização Não-Governamental.

² IES é Instituição de Ensino Superior.

Como a Tabela 3 pode dar uma idéia distorcida da distribuição do número alunos por turma, foi preparada a Tabela 4, que mostra que são poucas as turmas com poucos ou com muitos alunos. As turmas com mais de 25 alunos representam cerca de 0,8% do total de turmas e o máximo de 2,8% ocorre em entidades municipais. Por outro lado, as turmas com até 5 alunos representam 0,9% das turmas, com um máximo de 1,3% nas entidades estaduais. As turmas

com até 10 alunos correspondem a 7,5% das turmas e, no caso das entidades estaduais, atinge o máximo de 9,6% das turmas.

Tabela 4 – Distribuição do número de turmas por tipo de entidade, segundo classes de número de alunos na turmas.

Classes de número de alunos na turma	Número de turmas				
	Total	Tipo de entidade			
		Estaduais	Municipais	ONG ¹	IES ²
Total	101.307	45.205	12.578	41.072	2.452
De 1 a 5 alunos	862	543	80	215	24
De 6 a 10 alunos	6.048	3.391	751	1.840	66
De 11 a 15 alunos	16.188	8.328	2.288	4.531	1.041
De 16 a 20 alunos	26.813	11.487	2.874	11.983	469
De 21 a 25 alunos	50.691	21.161	6.261	22.419	850
Mais de 25 alunos	705	295	324	84	2

¹ ONG é Organização Não-Governamental.

² IES é Instituição de Ensino Superior.

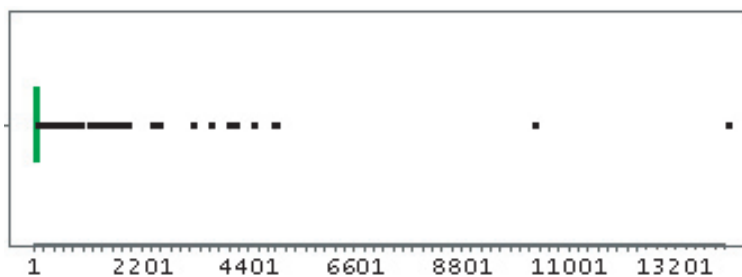
Ao Considerar-se os limites sugeridos na Resolução do PBA e os problemas de montagem e atualização que o SBA enfrenta, parece que esta variável não deve ser levada em conta, no desenho da amostra, como medida de tamanho das turmas, apesar de indicar diferenças de tamanho de turma entre os tipos de entidade.

O tipo de entidade é outra variável interessante para o desenho da amostra, pelo seu papel como variável de estratificação – seja pelas diferenças observadas nos dados, seja pelas mudanças recentes na orientação do PBA.

A distribuição do número de turmas por entidade, da mesma forma que ocorreu por município, também é bastante assimétrica, como mostra o gráfico de caixa na Figura 2. A média (158 turmas), por ser influenciada pelos valores extremos, é bem superior à mediana (12 turmas).

No entanto, o comportamento do número de turmas por entidade é bastante variado por tipo de entidade, como mostram os gráficos de caixa da Figura 3. Apesar de todas as distribuições apresentarem assimetria positiva, as medidas de posição são diferentes, indicando que as entidades municipais e de ensino superior têm menos turmas do que as entidades estaduais e as ONGs.

Figura 2 – Gráfico de caixa e estatísticas da distribuição do número de turmas por entidade.

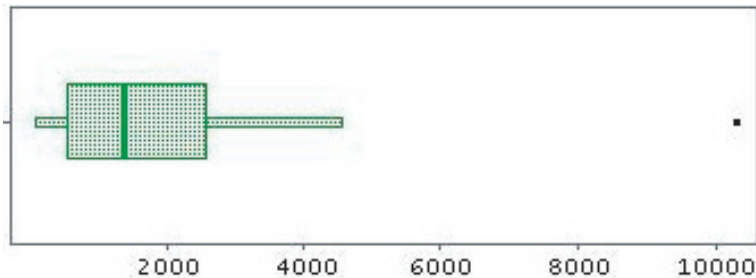


Número de turmas por entidade

Mínimo	1	Quartil 1	5	Mediana	12
Média	158	Quartil 3	35	Máximo	14.242

Figura 3 – Gráfico de caixa e estatísticas das distribuições do número de turmas por entidade, por tipo de entidade.

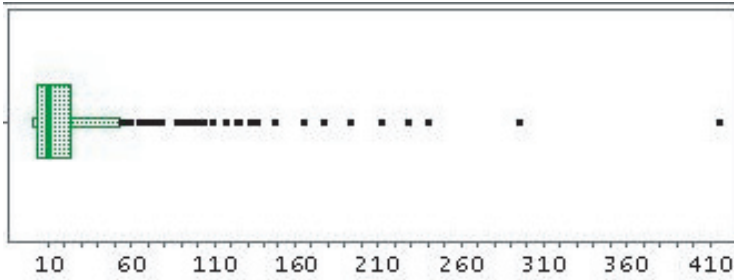
Entidades estaduais



Número de turmas por entidade

Mínimo	91	Quartil 1	564	Mediana	1.373
Média	2.055	Quartil 3	2.572	Máximo	10.275

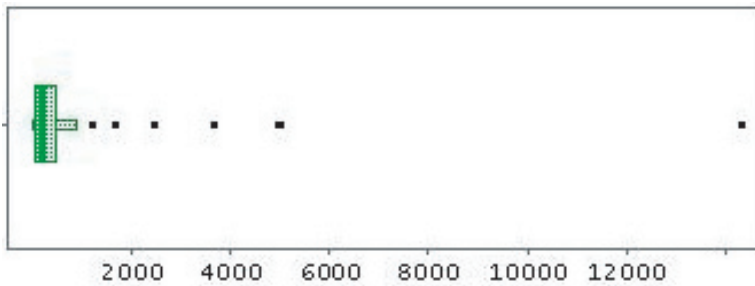
Entidades municipais



Número de turmas por entidade

Mínimo	1	Quartil 1	4	Mediana	11
Média	22	Quartil 3	24	Máximo	415

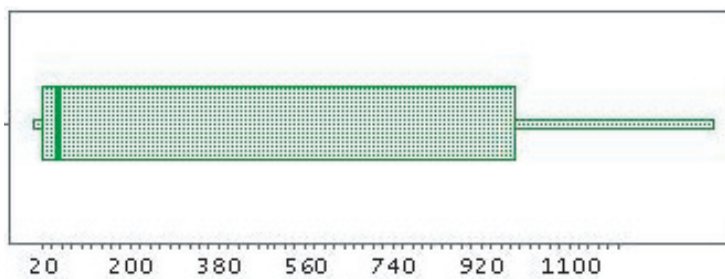
Organizações Não-Governamentais



Número de turmas por entidade

Mínimo	2	Quartil 1	60	Mediana	201
Média	893	Quartil 3	465	Máximo	14.242

Instituições de Ensino Superior



Número de turmas por entidade

Mínimo	2	Quartil 1	19	Mediana	50
Média	490	Quartil 3	986	Máximo	1.395

Além disto, algumas poucas entidades concentram um número expressivo de turmas. De fato, seis entidades (0,9% do total) têm 43.098 turmas, que representam 42,5% do total de turmas do cadastro de seleção, sendo três entidades estaduais (18.950 turmas, ou 41,9 % das turmas de entidades estaduais) e três ONGs (24.148 turmas, ou 58,8 % das turmas de ONGs), como pode ser observado na Tabela 5.

Tabela 5 – Número de entidades e de turmas por tipo de entidade, segundo classes de número de turmas da entidade.

Classes de número de turmas da entidade	Número de entidades e de turmas									
	Total		Tipo de entidade							
			Estaduais		Municipais		ONG ¹		IES ²	
	Entidades	Turmas	Entidades	Turmas	Entidades	Turmas	Entidades	Turmas	Entidades	Turmas
Total	640	101.307	22	45.205	567	12.578	46	41.072	5	2.452
1 a 9	254	1.048	0	0	249	1.027	4	19	1	2
10 a 19	135	1.763	0	0	131	1.711	3	33	1	19
20 a 29	77	1.834	0	0	77	1.834	0	0	0	0
30 a 39	28	956	0	0	27	921	1	35	0	0
40 a 49	23	1.015	0	0	23	1.015	0	0	0	0
50 a 69	28	1.545	0	0	23	1.268	4	227	1	50
70 a 99	22	1.802	1	91	16	1.317	5	394	0	0
100 a 149	18	2.161	1	142	13	1.564	4	455	0	0
150 a 199	4	685	0	0	3	534	1	151	0	0
200 a 299	15	3.581	1	250	4	972	10	2.359	0	0
300 a 499	6	2.456	2	802	1	415	3	1.239	0	0
500 a 999	8	6.075	3	2.033	0	0	4	3.056	1	986
1000 a 1999	10	14.789	7	10.531	0	0	2	2.863	1	1.395
2000 a 2999	3	7.565	2	5.121	0	0	1	2.444	0	0
3000 a 3999	3	10.934	2	7.285	0	0	1	3.649	0	0
4000 a 4999	4	18.581	2	8.675	0	0	2	9.906	0	0
10275	1	10.275	1	10.275	0	0	0	0	0	0
14242	1	14.242	0	0	0	0	1	14.242	0	0

¹ ONG é Organização Não-Governamental.

² IES é Instituição de Ensino Superior.

A Tabela 5 pode causar alguma confusão em sua leitura porque a variável “número de turmas” aparece duas vezes. No entanto, na primeira coluna ela é usada para definir as classes de número de turmas da entidade, enquanto nas demais colunas ela fornece o total de turmas das entidades classificadas em cada linha da Tabela. Assim, na linha de entidades com 1 a 9 turmas, observa-se, nas colunas de total, a existência de 254 entidades com 1 a 9 turmas, que constituíram, em conjunto, 1.048 turmas.

Na discussão entre estratificar a amostra por tamanho de município e por tipo de entidade, um cruzamento entre essas duas variáveis traz aspectos esclarecedores. Na Tabela 6, os municípios foram grupados em pequenos (até 10 mil habitantes), médios (entre 10 e 100 mil habitantes) e grandes (mais de 100 mil habitantes) e o

número de turmas em cada grupo de município foi detalhado segundo os tipos de entidades parceiras.

Tabela 6 – Número de turmas por grupos de população municipal, segundo as combinações de tipos de entidade.

Combinações de tipo de entidade	Número de municípios com turmas válidas							
	Total		População municipal em 01/07/2005					
			Até 10.000		De 10.000 a 100.000		Mais de 100.000	
	Número	%	Número	%	Número	%	Número	%
Total	3.607	100,0	1.479	41,0	1.915	53,1	213	5,9
Só um tipo de entidade	2.276	63,1	1.132	31,4	1.073	29,7	71	2,0
Estadual	1.391	38,6	877	24,3	500	13,9	14	0,4
Municipal	231	6,4	77	2,1	141	3,9	13	0,4
ONG ¹	603	16,7	167	4,6	393	10,9	43	1,2
IES ²	51	1,4	11	0,3	39	1,1	1	0,0
Dois tipos de entidade	1.184	32,8	334	9,3	757	21,0	93	2,6
Estadual e municipal	80	2,2	32	0,9	43	1,2	5	0,1
Estadual e ONG ¹	937	26,0	288	8,0	602	16,7	47	1,3
Estadual e IES ²	8	0,2	2	0,1	6	0,2	0	0,0
Municipal e ONG ¹	124	3,4	6	0,2	80	2,2	38	1,1
ONG ¹ e IES ²	35	1,0	6	0,2	26	0,7	3	0,1
Três tipos de entidade	141	3,9	13	0,4	81	2,2	47	1,3
Estadual, municipal e ONG ¹	122	3,4	9	0,2	69	1,9	44	1,2
Estadual, municipal e IES ²	1	0,0	1	0,0	0	0,0	0	0,0
Estadual, ONG ¹ e IES ²	15	0,4	2	0,1	11	0,3	2	0,1
Municipal, ONG ¹ e IES ²	3	0,1	1	0,0	1	0,0	1	0,0
Todos os tipos de entidade	6	0,2	0	0,0	4	0,1	2	0,1

¹ ONG é Organização Não-Governamental.

² IES é Instituição de Ensino Superior.

Os dados da Tabela 6 indicam que uma estratificação por grupo de população dos municípios dificilmente permitiria representar adequadamente os tipos de entidade parceira, visto que cerca de 63% dos municípios têm turmas constituídas por apenas um tipo de entidade. Além disso, não há uma correlação alta entre o número de turmas (ou de alunos) e a população municipal: enquanto a correlação entre turmas e alunos é da ordem de 0,97, a correlação entre população municipal e número de turmas válidas é menor do que 0,55.

Por outro lado, uma estratificação por tipo de entidade permitiria selecionar municípios de forma a ter uma razoável representação das macrorregiões do País.

A última variável do cadastro de seleção importante para o desenho da amostra é a data efetiva de conclusão do curso de alfabetização, tendo em vista que a seleção de turmas não deve considerar turmas com curso concluído.

A Tabela 7 mostra, no entanto, que esta variável não deve ser considerada no desenho da amostra, visto que, em média, uma a cada quatro turmas não tem esta informação no cadastro e, no caso das IES, mais da metade das turmas não tem data de término efetivo.

Tabela 7 – Número de turmas por tipo de entidade, segundo o ano de término efetivo do curso.

Ano de término efetivo	Total		Tipo de entidade							
			Estaduais		Municipais		ONG ¹		IES ²	
	Turmas	%	Turmas	%	Turmas	%	Turmas	%	Turmas	%
Total	101.307	100,0	45.205	100,0	12.578	100,0	41.072	100,0	2.452	100,0
2000	1	0,0	0	0,0	0	0,0	1	0,0	0	0,0
2005	3.571	3,5	1.780	3,9	347	2,8	1.209	2,9	235	9,6
2006	71.953	71,0	32.183	71,2	10.801	85,9	28.162	68,6	807	32,9
2009	1	0,0	0	0,0	0	0,0	1	0,0	0	0,0
Sem declaração	25.781	25,4	11.242	24,9	1.430	11,4	11.699	28,5	1.410	57,5

¹ ONG é Organização Não-Governamental.

² IES é Instituição de Ensino Superior.

Objetivos, estratificação a alocação da amostra do SBA 2005

Objetivos da amostra do SBA 2005

A amostra de turmas e alunos da edição 2005 do PBA, denominada simplesmente por amostra do SBA 2005, tem como objetivo geral conhecer o perfil socioeconômico dos alunos dos cursos de Alfabetização de Jovens e Adultos (AJA) financiados pelo PBA 2005, saber como foram selecionados e avaliar seu grau de aprendizado.

É certo, no entanto, que a avaliação de um programa passa por sua implementação e que esta pode ser diferenciada por segmentos da população. As-

sim, seria desejável que este objetivo geral pudesse ser cumprido para diferentes segmentos da população, tais como: macrorregiões do País e tipo de entidade parceira, dentre outros.

De fato, o desejo (ou necessidade) de que os objetivos sejam cumpridos para segmentos da população implica que os resultados possam ser estimados com precisão controlada para estes segmentos, o que, habitualmente, é feito com a estratificação da população. Essa solução assegura que os estratos (ou segmentos da população) sejam domínios de estimação para os quais os resultados podem ser aferidos com exatidão previamente controlada no desenho da amostra.

População alvo e população de pesquisa

A população alvo corresponde ao conjunto de alfabetizandos do Edital 2005 do PBA. Esta população está organizada por entidades parceiras, que se subdividem em turmas, às quais, por sua vez, estão associados os alfabetizandos e os alfabetizadores.

A população de pesquisa corresponde ao conjunto de alfabetizandos do Edital 2005 do PBA presentes no cadastro de seleção da amostra, ou seja, os alfabetizandos de turmas válidas registradas no SBA 2005 na data de referência do cadastro de seleção (04 de abril de 2006), que correspondem a 98,2% do total de alfabetizandos constantes no SBA 2005.

Estratificação e alocação do tamanho da amostra do SBA 2005

Por razões orçamentárias, o tamanho da amostra foi fixado em 800 turmas em todo o País. Este tamanho de amostra precisa ser repartido (ou alocado) entre os estratos definidos para a amostra para viabilizar a seleção das turmas em cada estrato da amostra.

Segundo Cochran (1977), evidências empíricas sugerem que o número de estratos deve variar entre 5 e 10. No entanto, a discussão no item 2 indicou que a estratificação inicial deveria ser feita por tipo de entidade parceira e que a distribuição do número de turmas por entidade apresentava acentuada assimetria positiva.

Os tratamentos para populações assimétricas nascem nos anos 1950, com um método baseado na minimização de variância para dois estratos (Dalenius, 1952), posteriormente generalizado para um número de estratos maior do que dois e prefixado por Dalenius & Hodges (1959) e Ekman (1959), que tem por

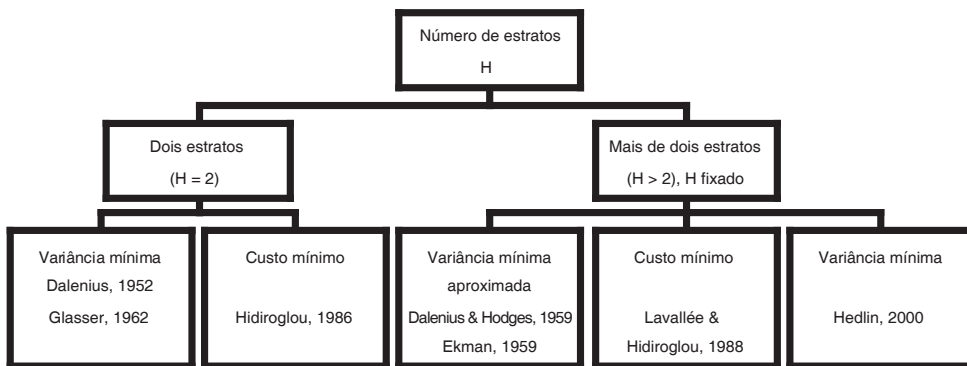
base um estimador aproximado da variância. Nestes casos, o tamanho total da amostra e o número de estratos eram previamente fixados e os pontos de cortes dos estratos eram obtidos de forma a maximizar a precisão da amostra (ou minimizar a variância do estimador).

Outra aproximação para lidar com populações assimétricas foi proposta por Hidiroglou (1986) para dois estratos, generalizada para mais de dois estratos por Lavallée & Hidiroglou (1988). Neste caso, a precisão da amostra e o número de estratos eram previamente fixados e os limites dos estratos eram calculados para produzir o menor tamanho total da amostra.

Em 2000, Dan Hedlin derivou um algoritmo para determinar os pontos de corte dos estratos de forma a minimizar a variância, dado um tamanho total de amostra e um número de estratos.

Todos esses métodos partem da idéia de criar um estrato censitário constituído pelas maiores unidades da população. Desta forma, essas unidades, que carregam a maior parte da variabilidade total, seriam investigadas a 100% e, portanto, não contribuiriam para aumentar a variância do estimador de total, visto não haver erro de amostragem em seu estrato. A Figura 4 esquematiza estes métodos.

Figura 4 – Esquema dos métodos de estratificação para populações positivamente assimétricas.



No caso da amostra do SBA 2005, o tamanho total da amostra foi prefixado em 800 turmas e o emprego do método de Hedlin (2000) permitiu obter os pontos de corte do estrato certo e dos demais estratos de forma a obter a amostra de turmas mais precisa dentre todas as amostras estratificadas de tamanho total igual a 800.

No entanto, como as entidades estaduais e as Instituições de Ensino Superior apresentavam tamanhos populacionais pequenos (22 e 5, respectivamente), as entidades deste tipo foram alocadas ao estrato certo e o método foi aplicado de forma independente para as entidades municipais e para as ONGs.

Para tanto, foi necessário alocar o tamanho total de turmas para cada tipo de entidade e definir o número de turmas a pesquisar em cada entidade de cada estrato. Assim, a solução encontrada foi aplicar o método de Hedlin para diferentes combinações de número de estratos, de tamanhos da amostra de turmas em cada entidade de cada estrato e, portanto, de tamanho total da amostra de entidades nos estratos de entidades municipais e ONGs.

Assim, foi preparado um procedimento iterativo onde o número de estratos variava de 3 a 10 e o tamanho da amostra de turmas variava de 3 a 20 turmas. Definidos estes valores para as entidades estaduais e municipais, resultavam o tamanho total de turmas e de entidades para os parceiros municipais e ONG, que eram submetidos ao método de Hedlin para definição dos limites dos estratos.

A cada aplicação do procedimento descrito, os valores básicos (tipo de entidade, limites dos estratos e tamanho da amostra de turmas em cada estrato) e a variância do estimador de total do número de alunos eram armazenados em um arquivo auxiliar, que foi usado para definir a estratificação proposta para a amostra do SBA 2005.

Posteriormente, para acomodar uma entidade de ensino superior que tinha apenas duas turmas, o tamanho final da amostra foi aumentado para 802 turmas, conforme indicado na Tabela 8.

Na Tabela 8, observa-se que:

- (1) Que as entidades estaduais e as Instituições de Ensino Superior serão todas incluídas na amostra;
- (2) Apesar de incluídas na amostra, essas entidades foram estratificadas por classes do número de turmas, para ter diferentes tamanhos da amostra de turmas por entidade;
- (3) No terceiro estrato de ONGs, apesar do método de Hedlin ter indicado a seleção de cinco entidades, optamos por selecionar todas as sete, deixando o estrato separado do estrato certo para ter diferentes tamanhos da amostra de turmas por entidade; e
- (4) O número de turmas a selecionar em cada entidade cresce de acordo com o estrato de número de turmas da entidade.

Tabela 8 – Dados populacionais e tamanho da amostra de entidades e turmas por estágio de seleção, segundo o tipo de entidade e a classe do número de turmas

Tipo da entidade e classe do número de turmas	População			Amostra do primeiro estágio			Amostra do segundo estágio	
	Nº de entidades	Nº de turmas	Nº de alunos	Nº de entidades a selecionar	Nº de turmas por entidade	Total de turmas a selecionar	Nº de turmas da entidade a selecionar por município	Número máximo de municípios na amostra
Total geral	640	101.307	2.006.601	86		802		182
Estaduais	22	45.205	870.317	22		301		
Até 150 turmas	2	233	5.118	2	6	12	6	2
De 150 a 999 turmas	6	3.085	56.436	6	9	54	3	18
De 1.000 a 2.999 turmas	9	15.652	288.539	9	15	135	5	27
3.000 turmas ou mais	5	26.235	520.224	5	20	100	5	20
Municipais	567	12.578	250.460	33		172		
Até 29 turmas	457	4.572	87.083	13	3	39	3	13
De 30 a 99 turmas	89	4.521	90.311	7	4	28	4	7
De 100 a 149 turmas	13	1.564	33.041	5	5	25	5	5
150 turmas ou mais	8	1.921	40.025	8	10	80	10	8
ONG ¹	46	41.072	841.567	26		299		
Até 99 turmas	17	708	15.162	7	6	42	3	14
De 100 a 499 turmas	18	4.204	83.022	8	9	72	3	24
De 500 a 2.999 turmas	7	8.363	165.283	7	15	105	5	21
3.000 turmas ou mais	4	27.797	578.100	4	20	80	5	16
IES ²	5	2.452	44.257	5		30		
Até 10 turmas	1	2	7	1	2	2	2	1
De 11 a 50 turmas	2	69	1.317	2	4	8	4	2
51 turmas ou mais	2	2.381	42.933	2	10	20	5	4

¹ ONG é organização não-governamental.

² IES é instituição de ensino superior.

Referências

- Hájek, J., 1960. *Limiting distribution in simple random sampling from a finite population*. Publications of the Mathematics Institute of the Hungarian Academy Science, 5:361-374.
- Cochran W.G., 1977. *Sampling Techniques*, 3rd edition. New York: John Wiley & Sons.
- Dalenius, T., 1952. *The problem of optimum stratification in a special type of design*. Skandinavisk Aktuarietidskrift, 35:61-70.
- Dalenius, T. & Hodges, J. L.Jr., 1959. *Minimum variance stratification*. Skandinavisk Aktuarietidskrift, 54:133-148.
- Ekman, G., 1959. *An approximation useful in univariate stratification*. Annals of Mathematical Statistics, 30:219-229.
- Glasser, G.J. *On the complete coverage of large units in a statistical study*. Annals of Mathematical Statistics, 32:28-32.
- Hedlin D., 2000. *A procedure for stratification by an extended Ekman rule*. Journal of Official Statistics, 16(1):15-29.
- Hidiroglou M.A., 1986. *The construction of a self-representing stratum of large units in survey design*. The American Statistician, 40(1):27-31.
- Lavallée P. & Hidiroglou M.A., 1988. *On the stratification of skewed populations*. Survey Methodology, 14(1):33-43.

Estudo sobre o desenho da amostra para análise da demanda por cursos de Alfabetização de Jovens e Adultos

Pedro Luis do Nascimento
Silva¹
Mauricio Teixeira Leite
de Vasconcellos²
Nuno Duarte da Costa
Bittencourt³

Introdução

Em função de sua importância, o componente de demanda para a avaliação do *Programa Brasil Alfabetizado* vem sendo alvo de estudos no âmbito do Projeto de Avaliação do *Programa Brasil Alfabetizado* já há algum tempo. No Anexo 1, consta a Nota Técnica nº 3, de 22 de março de 2005, onde são discutidas opções para a amostra deste componente. Esta nota técnica recomenda a realização de uma pesquisa domiciliar de âmbito nacional para levantar os dados necessários a este componente.

Em princípio, dimensionar a população analfabeta não é problema complicado, visto que estão disponíveis informações sobre pessoas analfabetas no âmbito do Sistema Estatístico Nacional. No entanto, as pesquisas domiciliares, sejam as anuais (Pesquisa Nacional por Amostra de Domicílios), seja o Censo Demográfico decenal, não captam a demanda pelos cursos e muito menos seu sucesso.

¹ Professor da Escola Nacional de Ciências Estatísticas (ENCE/IBGE), Doutor em Estatística (University of Southampton)

² Professor da Escola Nacional de Ciências Estatísticas, Doutor em Saúde Pública – ENSP/FIOCRUZ

³ Professor da Escola Nacional de Ciências Estatísticas (ENCE/IBGE), Mestre em Estatística (IMPA/CNPq)

Nessas pesquisas, não é possível identificar pessoas que fizeram Alfabetização de Jovens e Adultos (AJA) ou Educação de Jovens e Adultos (EJA) e prosseguiram nos estudos. Porém, a proporção dos que fizeram AJA e definem-se como analfabetos pode ser estimada sem maiores problemas, deixando a impressão de que os cursos de AJA não conseguem alfabetizar.

Neste sentido, uma pesquisa que permita determinar a demanda pelos cursos de AJA e EJA, e que colete informações sobre a história educacional dos entrevistados – e em particular sobre os que cursaram AJA ou EJA e prosseguiram nos estudos – poderia resolver as duas questões apontadas.

No entanto, uma pesquisa domiciliar de âmbito nacional tem um custo elevado. Assim, a idéia de realizar este levantamento em conjunto com a Pesquisa de Avaliação do Impacto do Programa Bolsa Família (AIBF), sobre a responsabilidade do Centro de Desenvolvimento e Planejamento Regional (Cedeplar) da Universidade Federal de Minas Gerais (UFMG), com a participação da Science, se fortaleceu e resultou em um acordo entre a Secretaria de Educação Continuada, Alfabetização e Diversidade (Secad/MEC) e a Secretaria de Avaliação e Gestão da Informação (Sagi/MDS), contratante da AIBF.

A Science desenvolveu o plano de amostragem da AIBF, descrito neste relatório, e, em parceria com o Cedeplar, elaborou os questionários e manuais de instrução da AIBF.

Desenho da amostra da AIBF

O objetivo principal da AIBF é desenvolver um conjunto de instrumentos que permitam estimar efeitos ou impactos do Programa Bolsa Família (PBF) sobre as famílias usuárias. Por concepção, a AIBF inclui duas pesquisas: (1) uma pesquisa base (baseline), que será realizada para coletar os dados das famílias usuárias ou não do PBF; e (2) uma pesquisa de seguimento (follow-up) com as mesmas famílias, para determinar o impacto do PBF.

Neste capítulo, são descritos os procedimentos adotados para a seleção da amostra da pesquisa base. O capítulo está organizado em seções que descrevem os cadastros utilizados, a população alvo da AIBF, uma visão geral do plano amostral e, posteriormente, seções que descrevem pormenorizadamente cada uma das etapas do processo de amostragem, desde a seleção de municípios, passando pela seleção de setores, chegando à descrição dos métodos que serão utilizados para seleção de domicílios nos quais serão conduzidas as entrevistas.

Cabe registrar que os procedimentos para a seleção de municípios e setores foram realizados em escritório, e os resultados estão sumarizados neste capítulo. Os procedimentos para seleção de domicílios dependem da operação de coleta, e por este motivo, são aqui descritos sem resultados que ilustrem sua aplicação.

Considerando que a amostra selecionada para a AIBF não pode ser disseminada antes da realização da pesquisa de seguimento, seja por razões éticas associadas ao processo de avaliação externa, seja para resguardar a Sagi (MDS) de qualquer possibilidade de crítica relacionada com a possível atuação no sentido de melhorar os resultados da avaliação. Os arquivos magnéticos que contém a amostra selecionada não constam neste estudo, e estão disponíveis apenas para as equipes encarregadas do processo de avaliação externa.

Cadastros utilizados

Como em toda pesquisa, um elemento fundamental de apoio para preparação da amostra é o sistema de referência ou cadastro. Ao considerar-se a decisão de realizar uma pesquisa de base populacional, foram utilizados para apoiar o trabalho de amostragem da pesquisa AIBF dois cadastros principais.

O primeiro cadastro é formado a partir do Arquivo Agregado por Setores Censitários dos Resultados do Universo do Censo Demográfico 2000 (IBGE, 2003). O arquivo agregado por setores é produzido a partir dos microdados do Censo Demográfico 2000, com medidas resumo que descrevem os setores. Contém um registro para cada um dos 215.811 setores censitários em que foi dividido o País para a realização da operação de coleta dos dados do Censo. Foram excluídos deste arquivo 7.094 setores por razões indicadas a seguir, restando 208.717 setores que participaram do cadastro efetivamente empregado para a seleção da amostra de setores do AIBF.

Além deste cadastro, foi também utilizado um arquivo derivado do Cadastro Único do Governo Federal, contendo informações sobre os benefícios pagos às famílias no âmbito dos programas Bolsa Família; Bolsa Escola; Cartão Alimentação; Bolsa Alimentação; e Vale Gás. Este segundo arquivo foi utilizado apenas para apoiar a estratificação de municípios ou grupos de municípios, e não para a seleção direta de unidades de pesquisa (domicílios ou famílias).

Após a seleção de setores em que a pesquisa será realizada, serão utilizadas informações da Base Operacional Geográfica do IBGE, na forma de mapas e descrições dos setores selecionados, que irão apoiar a realização da operação de lista-

gem e, posteriormente, a coleta de dados nos domicílios selecionados. Espera-se, ainda, utilizar as folhas de coleta do Censo Demográfico 2000 (CD2000) dos setores selecionados, para apoiar o trabalho de listagem, e contribuir dessa maneira para reduzir a margem de erro de cobertura desta operação.

Foi decidido utilizar, para o desenho amostral, a divisão territorial (em municípios) vigente à época do Censo Demográfico 2000. No entanto, na amostra de municípios e setores selecionados, são identificadas as eventuais mudanças de municípios, para utilização das informações atualizadas no momento da coleta de dados. Qualquer decisão distinta desta levaria a um trabalho adicional de alocação das informações dos setores nos municípios criados após a realização do CD2000, tarefa que não acrescentaria qualidade à base usada para seleção da amostra, e permitiria, apenas, recompor, nos dados do CD2000, a malha municipal atualizada.

Foi decidido, também, excluir do cadastro de seleção (e, portanto, da amostra) os setores rurais da macrorregião Norte³. No entanto, foram mantidos no cadastro de seleção dos setores de “Aglomerado rural de extensão urbana” (código de situação igual a 4) da região Norte, visto que os custos de coleta nestes setores são ligeiramente maiores que nos da área urbana. Além disto, é nestes setores que costuma ocorrer boa parte da expansão da área urbana, e é onde são maiores as chances de mudança de situação de rural para urbana desde o CD2000.

Além dos setores rurais da região Norte, foram excluídos todos os setores de tipo especial⁵ com códigos 2, 4, 5, 6 e 7. Isto implicou em manter no cadastro apenas os setores de tipo 0 (“Comum”), 1 (“Especial de aglomerado subnormal”) e 3 (“Especial de alojamento, acampamentos, etc.”). Diante das características da pesquisa, pareceu importante manter os dois tipos “especiais” (códigos 1 e 3) no universo de setores de onde a amostra foi selecionada.

Em conjunto, estas duas classes de exclusões de setores levaram a deixar de fora da população de pesquisa cerca de 2,1% das unidades domiciliares, e cerca de 2,4% da população brasileira. A Tabela 1, no Anexo 2, mostra as contagens e proporções de setores, domicílios e pessoas, total e excluídos, por situação e tipo do setor, considerados os dados do Censo Demográfico 2000.

Uma última alteração cadastral feita antes do uso do cadastro para a seleção da amostra de municípios e setores foi o agrupamento do município de Fernando de Noronha (código 2605459) ao município de Recife (código 2611606). Esta de-

⁴ No Quadro 1, no Anexo 2, é apresentada a relação dos códigos de situação dos setores.

⁵ A relação de códigos de tipos de setor constam do Quadro 2, no Anexo 2.

cisão foi tomada por se tratar de município insular isolado, cujo acesso é feito por avião a partir de Recife, contendo apenas um setor no CD2000. Assim, o município de Fernando de Noronha foi tratado, para fins da seleção da amostra, como se fosse mais um setor do município de Recife.

População alvo e população de pesquisa

A população alvo é composta pelas famílias e pessoas residentes em domicílios particulares (permanentes ou improvisados) no território nacional. Por questões relacionadas com as dificuldades esperadas na realização da operação de coleta, foram feitas exclusões de partes desta população ao definir a população de pesquisa, conforme descrito na seção 2.1.

Assim sendo, a população de pesquisa ficou definida como o conjunto das famílias e pessoas residentes em domicílios particulares (permanentes ou improvisados) no território nacional, excluídos os localizados em setores especiais (tipos 2, 4, 5, 6 e 7) e os localizados em setores rurais da região Norte (códigos de situação iguais a 5, 6, 7 e 8).

Visão Geral do Plano Amostral

O objetivo geral do plano amostral é selecionar uma amostra que permita obter entrevistas em cerca de 15.000 domicílios, que em seu conjunto representem a população de pesquisa. A amostra de domicílios deverá ser distribuída em proporções desiguais entre três categorias:

- a) domicílios com famílias usuárias do Programa Bolsa Família, denominados “casos”;
- b) domicílios com famílias cadastradas no Cadastro Único, mas que ainda não são usuárias do PBF, denominados “controles tipo 1” (estes domicílios podem ser usuários de outros programas de transferência de renda, mas não do PBF); e
- c) domicílios sem famílias cadastradas ou usuárias, denominados “controles tipo 2”.

Como o interesse prioritário do projeto é avaliar o impacto do Programa Bolsa Família, a idéia inicial é que a composição da amostra de 15.000 domicílios

seja distribuída nas seguintes proporções: 4.500 casos (30% do total da amostra), 9.000 controles do tipo 1 (60% do total da amostra), e 1.500 controles do tipo 2 (10% do total da amostra).

Os controles do tipo 1 são domicílios onde as famílias seriam potencialmente elegíveis ao benefício do PBF, mas ainda não recebem este benefício. O maior interesse na observação deste grupo é justamente poder obter dados para uma parcela de famílias num instante anterior ao do recebimento do benefício (linha base ou baseline) com os quais posteriormente serão comparados dados de outras rodadas da pesquisa, realizadas em intervalos de um ano (follow-ups). Com a eventual conversão destes controles em casos (usuários) entre a primeira rodada e as rodadas posteriores, seria possível fazer comparações mais eficazes do tipo antes versus depois, para uma amostra da população de usuários que não foram atendidos nas primeiras etapas do programa.

O segundo grupo de controles presta-se, essencialmente, a permitir avaliar a cobertura do programa (estimar a proporção de domicílios e famílias elegíveis não atendidas) e também para dar medidas da variação de padrões de renda e consumo para domicílios e famílias não atendidos pelo programa, que permitirão examinar efeitos globais de variação da economia.

Já o grupo de casos, formado pela amostra de famílias já atendidas, será acompanhado para examinar como varia ao longo do tempo a situação das famílias já atendidas pelo programa.

Como não existe um cadastro atualizado e disponível de todos os domicílios existentes na área de interesse da pesquisa, foi necessário lançar mão de um plano por amostragem de conglomerados em dois ou três estágios, com estratificação das unidades primárias de seleção. Além da amostragem conglomerada, a pesquisa utilizará amostragem dupla com estratificação, como será descrito adiante.

A primeira fase da amostragem compreenderá a seleção de setores censitários nos quais será feita uma operação de listagem (varredura ou screening) para cadastrar os domicílios encontrados, classificando-os nas três categorias (casos, controles de tipo 1, e controles de tipo 2).

Na segunda fase de amostragem, os domicílios serão selecionados por amostragem estratificada dentro de cada uma destas categorias, em cada município ou grupo de municípios amostrado.

Amostragem na primeira fase

A primeira tarefa da amostragem na fase 1 é a estratificação da população. A coordenação do projeto indicou que a pesquisa deveria fornecer resultados para três estratos definidos em função das macrorregiões: Nordeste (NE), Sudeste e Sul (SE+S), e Norte (exceto os setores rurais) e Centro-Oeste (N+CO). Portanto, esta é precisamente a primeira estratificação feita.

A obtenção de resultados separados para cada um destes estratos com precisão similar sugere também uma alocação igual da amostra (de domicílios) entre as três áreas. Assim, em princípio, a idéia é que a amostra de cada um destes estratos geográficos seja de cerca de 5.000 domicílios, repartidos entre os três grupos de casos e controles nas mesmas proporções propostas para a amostra inteira (30%, 60% e 10%, respectivamente, para casos, controles tipo 1 e controles tipo 2).

Uma segunda tarefa do planejamento amostral consiste em definir, dentro de cada estrato, as Unidades Primárias de Amostragem (UPAs). Os municípios surgem como candidatos naturais a serem utilizados como UPAs. Entretanto, surgem duas dificuldades derivadas da grande variação de tamanho dos municípios.

Em primeiro lugar, alguns municípios são muito grandes, e se utilizados como UPAs em qualquer plano amostral com sorteio baseado em probabilidades proporcionais ao tamanho, seriam incluídos na amostra com certeza. Nessa hipótese, deixariam de operar como unidades conglomeradas de seleção, passando a funcionar como estratos, dentro dos quais são selecionadas unidades primárias de seleção menores, como por exemplo, setores censitários.

Assim, uma das primeiras tarefas foi definir que municípios deveriam ser incluídos na amostra. Foram preparadas duas listas dos maiores municípios segundo a população residente total em 2000 (dados do Censo Demográfico 2000, do IBGE) e em 2005 (dados de projeções de população elaboradas pelo Cedeplar).

Entre os 40 maiores municípios, há apenas dois municípios “discrepantes” nas duas listas. Na lista de 2000, São João do Meriti entraria, mas ficaria de fora na lista de 2005. Em seu lugar entraria Ananindeua, que fica de fora por muito pouco em 2000. Em função disto, decidiu-se ficar com uma lista de 41 municípios “grandes” para inclusão na amostra. Estes 41 municípios foram os 40 municípios com maior população residente total em 2000 e o município de Ananindeua.

Com este raciocínio em mente e considerando outras necessidades do projeto, decidiu-se que os 41 maiores municípios do País, em termos de população, seriam tratados como estratos, e não como unidades primárias de amostragem. Com

isto, mais tarde será possível recompor qualquer das duas relações de 40 maiores municípios: a de 2000 ou a de 2005, conforme o interesse do analista.

Assim, a estratificação geográfica da amostra é feita em dois níveis: primeiro, as três grandes áreas formadas por combinações das macrorregiões (N+CO; NE; e SE+S); depois, dentro de cada uma destas áreas, cada município entre os 41 maiores é considerado um estrato de seleção, e o conjunto dos demais municípios de cada área é também um estrato.

A segunda dificuldade surge com os municípios com população muito pequena. Neste caso, a utilização destes municípios como UPAs representaria elevação dos custos de pesquisa, porque obrigaria a espalhar mais a amostra de municípios, pois em muitos deles não seria possível obter amostras de tamanhos razoáveis dos três grupos de interesse. Assim, antes de passar à seleção de municípios, foi realizada uma operação de agregação de municípios pequenos, de modo que os resultantes “grupos de municípios” tivessem tamanho mínimo que permitisse uma coleta eficiente, caso selecionados para a amostra da primeira etapa.

Assim, em função da estratificação, o plano amostral da pesquisa na fase 1 pode ser especificado ao considerar o seguinte esquema geral, descrito no Quadro 1.

Quadro 1 – Descrição do esquema de amostragem da pesquisa AIBF na Fase 1

Aspecto	Estrato	
	Municípios grandes (os 41 maiores em população)	Municípios “pequenos” (todos os demais)
Plano amostral	Amostragem estratificada de conglomerados em um estágio	Amostragem estratificada de conglomerados em dois estágios.
Estratificação	três grandes áreas, e cada município dentro destas.	Três grandes áreas, e dentro destas, três estratos baseados na proporção de pessoas usuárias de programas de transferência de renda.
UPA	Setor censitário	Municípios ou grupo de municípios
USA ¹		Setor censitário

¹ USA é abreviatura de Unidade Secundária de Amostragem

Amostragem nos municípios grandes

Nos municípios grandes, cada município é um estrato para seleção da amostra de setores (UPAs). Os setores foram selecionados por amostragem sistemática com probabilidades proporcionais ao tamanho (PPT Sistemático) dentro de cada estrato (município). A medida de tamanho empregada foi uma variável derivada que trazia informação sobre o número de domicílios cujos responsáveis foram considerados pobres⁶, com base na renda informada durante o CD2000. Os responsáveis pobres foram definidos como todos os que tinham renda nominal mensal inferior a dois salários mínimos.

Antes do sorteio, os setores foram ordenados de forma descendente segundo a medida “Proporção dos responsáveis pobres”⁷. Este método de seleção confere um efeito de estratificação implícita da amostra de setores com relação à variável “Proporção dos responsáveis pobres”. O número de municípios, de setores a selecionar por município e o número total de setores na amostra de cada grande área são apresentados na Tabela 2, no Anexo 2.

Amostragem nos municípios pequenos

No estrato dos municípios “pequenos”, isto é, de todos os municípios não-classificados como “grandes” que foram incluídos na amostra com certeza, decidiu-se usar como unidade primária de amostragem grupos de municípios contíguos contendo pelo menos 50 setores.

No momento da implementação do procedimento para agrupar municípios, três municípios foram considerados “especiais”. Trata-se dos municípios de Nilópolis (código 3303203), Diadema (código 3513801) e São Caetano do Sul (código 3548807) que não possuem municípios vizinhos no estrato dos municípios pequenos. Portanto, cada um destes municípios foi declarado uma unidade primária de amostragem, e nenhum deles foi submetido ao processo de agrupamento de municípios realizado. Desta forma, 5.462 municípios foram submetidos ao processo de agrupamento de municípios.

⁶ A medida de tamanho de cada setor foi definida como:
Tamanho = $(V0602 + V0603 + V0604 + 1) \cdot 2/3$, onde as variáveis V0602, V0603 e V0604 são todas do arquivo agregado por setores censitários (IBGE, 2003), sendo:
V0602 = Número de domicílios cujo responsável tem renda até $\frac{1}{2}$ Salário Mínimo;
V0603 = Número de domicílios cujo responsável tem renda de $\frac{1}{2}$ até 1 Salário Mínimo;
V0604 = Número de domicílios cujo responsável tem renda de 1 até 2 Salários Mínimos.

⁷ A proporção dos responsáveis pobres é definida como a razão entre a soma das variáveis V0602, V0603, V0604 dividida pela variável V0621, todas do arquivo agregado de setores, sendo V0621 o número total de domicílios no setor.

O agrupamento de municípios para formar unidades primárias de amostragem foi realizado utilizando uma abordagem implementada no sistema Skater (Assunção, Lage e Reis, 2003). Esta abordagem é denominada de *Spatial clustering by tree edge removal* (conglomerção espacial por remoção de arestas num grafo), e consiste essencialmente de um procedimento para agrupar unidades vizinhas de modo a minimizar uma função de distância entre as unidades, com uma opção que permite impor uma restrição de tamanho mínimo dos grupos de unidades vizinhas a serem formados.

A técnica de agrupamento foi aplicada inicialmente ao conjunto de municípios do país como um todo, sem separar os municípios pelas grandes áreas (macrorregião ou grupo de macrorregiões). O resultado levou à formação de 1.417 grupos, dos quais apenas 27 continham municípios em mais de uma grande área. Como eram poucas as exceções, estas foram tratadas com realocação manual dos municípios em outros grupos, sempre assegurando que em cada grupo de municípios restassem apenas municípios de uma mesma grande área. O número total de grupos ao final desta etapa de realocação permaneceu igual a 1.417. Somados aos três municípios tratados de maneira especial conforme acima descrito, o total de UPAs no estrato de municípios pequenos ficou igual a 1.420.

A formação inicial dos grupos de municípios foi efetuada com os seguintes critérios:

- 1) Aplicação do Skater com a opção de formar grupos de municípios contendo ao menos 50 setores; e
- 2) Uso da variável “Proporção de pessoas atendidas em programas de transferência de renda”⁸ como guia para agrupamento dos municípios.

O sistema Skater foi usado de modo distinto do usual, em que se busca agrupar unidades similares segundo a variável ou as variáveis de agrupamento. Na presente aplicação, buscou-se reunir num mesmo grupo municípios com valores bastante distintos da variável de agrupamento. Este novo critério de agrupamento tem por objetivo formar grupos com a maior heterogeneidade interna possível, já que os grupos de municípios funcionarão como unidades primárias de amostragem (conglomerados). Para conseguir este efeito com a aplicação do software Skater, a vari-

⁸ A proporção de pessoas atendidas por algum programa federal de transferência de renda (Bolsa Família, Bolsa Escola, Cartão Alimentação ou Bolsa Alimentação) foi calculada com base no Cadastro Único de beneficiários, tomando como denominador desta proporção em cada município ou grupo de municípios a estimativa da população municipal publicada pelo IBGE referente ao ano de 2002.

ável de agrupamento foi transformada mediante três passos: a) cálculo do posto (rank) de cada município segundo esta variável; b) padronização dos postos de modo a ficarem com média igual a 0 e variância igual a 1; c) cálculo do módulo (valor absoluto) dos postos padronizados.

Com esta transformação da variável de agrupamento, o uso do Skater fez com que municípios com alto posto fossem juntados primeiramente com os de baixo posto, numa espécie de “pareamento com seu oposto” em termos dos postos centrados da variável de agrupamento. Além disso, o uso do programa, com a opção de tamanho mínimo de 50 setores por grupo de municípios, levou à máxima divisão possível da população de municípios em grupos, como era desejado.

Assim, na primeira tentativa, todos os grupos ficaram com número de setores igual ou maior que 50. Esta regra foi violada para uns poucos grupos após a realocação feita para assegurar que em cada grupo ficassem apenas municípios de uma mesma grande área. No total, cinco grupos ficaram abaixo deste limite, sendo que o menor deles ficou com 42 setores.

Além do número de setores por grupo, é interessante analisar como ficaram as distribuições dos grupos quanto ao número de municípios em cada grupo (Tabela 3, no anexo 2). Verifica-se, ao analisar a Tabela 3, que 97,4% dos grupos têm até 10 municípios, e que há poucos grupos com mais municípios, sendo que o maior de todos tem 41 municípios. Isto não deverá provocar problemas ao processo de implementação da amostra, pois é viável fazer, em tais grupos muito grandes, etapas adicionais de amostragem, se necessário.

Após a formação dos grupos de municípios, estes grupos (as UPAs nesta parte do plano amostral) foram estratificados segundo a “Proporção de pessoas atendidas em programas de transferência de renda”. Em cada grande área, foram formados três estratos de UPAs de acordo com esta variável. A amostra de UPAs necessária foi alocada entre os três estratos de maneira desigual, dando maior fração amostral ao estrato onde é maior a proporção de pessoas atendidas em programas de transferência de renda.

O número total de UPAs a selecionar nesta parte do plano amostral foi determinado de maneira a respeitar os limites especificados para o planejamento amostral. Como a amostra dos municípios grandes ficou em 41, no estrato dos municípios pequenos decidiu-se selecionar 65 UPAs. Com isto, a amostra seria “concentrada” em 106 áreas contíguas, correspondentes a municípios grandes ou grupos de municípios pequenos (em termos de sua população). A definição deste número de grupos de

municípios a selecionar no estrato dos municípios pequenos foi feita após considerar a recomendação dada pela coordenação do projeto de seleção de cerca de apenas 100 municípios para a amostra. A motivação para esta recomendação decorreu da estrutura planejada para realização da coleta dos dados e em função do limite orçamentário já fixado para o projeto, o qual considerava a inclusão deste número de municípios na amostra. Na verdade, o esforço de agrupamento de municípios pequenos em UPAs maiores foi feito justamente para permitir satisfazer esta recomendação e ao mesmo tempo conseguir um plano que viabilizasse a distribuição da amostra de maneira uniforme entre as UPAs da amostra, deste estrato de municípios pequenos.

Os resultados da estratificação de UPAs e sua alocação na amostra são apresentados na tabela 4, no Anexo 2.

A seleção dos grupos de municípios (UPAs) em cada estrato foi efetuada utilizando amostragem com probabilidades proporcionais ao tamanho⁹. O método usado para a seleção PPT foi Amostragem Sequencial de Poisson (Ohlsson, 1998).

Uma descrição resumida deste método é apresentada no Anexo 3. Sua escolha nesta aplicação foi motivada pela simplicidade do método e pela facilidade que este método propicia para implementar mecanismos de rotação amostral, caso isto venha se tornar necessário.

Após a seleção das unidades primárias de amostragem (grupos de municípios, neste caso), dentro de cada uma destas UPAs foram selecionados setores por amostragem sistemática, com probabilidades proporcionais ao tamanho (PPT Sistemático). A medida de tamanho empregada foi a mesma utilizada na seleção de setores no estrato dos municípios grandes, isto é, uma variável derivada que trazia informação sobre o número de domicílios cujos responsáveis foram considerados pobres¹⁰ com base na renda informada no Censo Demográfico 2000. Os responsáveis pobres foram definidos como todos os que tinham renda nominal mensal inferior a 2 salários mínimos. Antes do sorteio, os setores foram ordenados dentro de cada UPA de forma descendente, segundo a medida “Proporção dos responsáveis pobres”¹¹. Este método de seleção confere um efeito de estratificação implícita da amostra de setores com relação à variável “Proporção dos responsáveis pobres”.

⁹ A medida de tamanho de cada grupo de municípios foi definida como:
Tamanho = $(V0602 + V0603 + V0604) \frac{1}{2}$, onde as variáveis V0602, V0603 e V0604 são todas do arquivo agregado por setores.

¹⁰ A medida de tamanho de cada setor foi definida como:
Tamanho = $(V0602 + V0603 + V0604 + 1) \frac{2}{3}$, onde as variáveis V0602, V0603 e V0604 são todas do arquivo agregado por setores.

¹¹ A proporção dos responsáveis pobres é definida como a razão entre a soma das variáveis V0602, V0603, V0604 dividida pela variável V0621, todas do arquivo agregado por setores.

O número de municípios, de setores selecionados por município e o número total de setores na amostra de cada grande área são apresentados na Tabela 5, no Anexo 2.

Amostragem na segunda fase

Uma vez selecionados os setores censitários para a amostra, na primeira fase, um processo de listagem, com entrevista aos domicílios encontrados em cada setor selecionado, foi realizado pela equipe de campo, na semana que antecede à semana agendada para coleta de dados no setor. Esta operação gerou um cadastro dos domicílios de cada setor amostrado, onde os domicílios encontrados foram classificados segundo o tipo (caso, controles tipo 1, controles tipo 2). Este tipo de operação de listagem é comumente chamado de *screening* ou varredura para cadastramento.

A seleção dos domicílios foi feita com estratificação nestas três categorias, dentro de cada setor, de modo a permitir controlar os tamanhos de amostra nas diferentes categorias, como desejado. O método de seleção dos domicílios usou, como estratos, as listas de domicílios compiladas reunindo os domicílios de cada tipo em cada município (ou grupo de municípios), e buscou distribuir as amostras de maneira uniforme entre os setores da amostra. A seleção dos domicílios em cada estrato (categoria x setor) foi feita com probabilidades iguais, mas domicílios em diferentes estratos podem ter probabilidades de seleção diferentes.

A seleção estratificada de domicílios foi uma aplicação de amostragem dupla (Cochran, 1977) ou amostragem em duas fases (*two-phase sampling*) descrita em Särndal et al (1992).

Alocação geral da amostra

A Tabela 6, no Anexo 2, apresenta a alocação geral da amostra planejada para a pesquisa base do Projeto AIBF. Os arquivos com a identificação da amostra de municípios e setores foram elaborados no formato do Excel, e entregues ao Cedeplar em formato digital. O protocolo metodológico pactuado para o projeto requer que tais arquivos sejam mantidos em sigilo pelas equipes da Science e Cedeplar, e não sejam disseminados nem entregues ao MDS antes do final do projeto.

Referências

- Assunção R.M., Lage J.P. e Reis E.A., 2003. *Análise de Conglomerados Espaciais via Árvore Geradora Mínima*. Revista Brasileira de Estatística.
- Cochran, W. G., 1977. *Sampling Techniques*, 3rd edition. Nova Iorque: John Wiley & Sons.
- Fundação Instituto Brasileiro de Geografia e Estatística (IBGE), 2003. *Censo Demográfico 2000: Agregado por Setores Censitários dos Resultados do Universo*, 2ª edição. Rio de Janeiro: IBGE.
- Ohlsson, E., 1998. *Sequential Poisson Sampling*. Journal of Official Statistics, 14: 149-162.
- Pessoa, D. G. C. & Nascimento Silva, P. L. do, 1998. *Análise de Dados Amostrais Complexos*. São Paulo: Associação Brasileira de Estatística.
- Särndal, C.E., Swensson, B., Wretman, J., 1992. *Model Assisted Survey Sampling*. Nova Iorque: Springer-Verlag.

**Nota técnica 3: reflexões sobre a amostra
para o componente da demanda
do *Programa Brasil Alfabetizado***

Nota técnica 3: reflexões sobre a amostra para o componente da demanda

Versão 22/03/2005

Mauricio Teixeira Leite de Vasconcellos

Science

Introdução

Esta nota técnica descreve duas opções para a seleção de uma amostra de analfabetos (15 anos ou mais), comparando suas vantagens e desvantagens.

A primeira opção baseia-se no cadastro do Sistema de Informação da Atenção Básica (Siab), enquanto a segunda caracteriza-se por ser uma pesquisa domiciliar com amostra selecionada a partir dos dados do Censo Demográfico 2000 (CD2000).

Nas duas opções, a amostra representará a nação, não permitirá inferências para domínios regionais, e será conglomerada por município. Estas duas medidas visam à redução do custo da pesquisa.

Desenho das amostras

Amostra do SIAB	Pesquisa domiciliar
<p>Primeiro estágio: Município (n=60) Seleção com probabilidade proporcional ao número de analfabetos do município (A_i), condicionada a um corte pré-definido. Municípios onde $A_i > \text{corte}$ serão automaticamente incluídos na amostra (estrato certo), onde $\text{Corte} = \text{número de analfabetos do Siab} / 60$. A diferença entre 60 e o número de municípios do estrato certo será selecionada de forma aleatória. Problemas:</p> <ul style="list-style-type: none">• Existem 269 municípios do CD2000 que não constam do Siab;• Existem diversos municípios onde o Siab tem baixa cobertura (nº de analfabetos bem menor que o do Censo).	<p>Primeiro estágio: Município (n=60) Seleção com probabilidade proporcional ao número de analfabetos do município em 2000 (AC_i), condicionada a um corte pré-definido. Municípios onde $AC_i > \text{corte}$ serão automaticamente incluídos na amostra (estrato certo), onde $\text{Corte} = \text{número de analfabetos do Censo} / 60$. A diferença entre 60 e o número de municípios do estrato certo será selecionada de forma aleatória. Problemas:</p> <ul style="list-style-type: none">• Existem 45 municípios do Siab que não constam do CD 2000 (municípios instalados em 01/01/2001 e em 01/01/2005). <p>As áreas geográficas dos municípios novos serão consideradas nos municípios que lhes deram origem, o que resolve a questão de cobertura mas traz pequenos problemas para a coleta (identificação do setor ou setores selecionados em dois municípios).</p>

Amostra do SIAB	Pesquisa domiciliar
<p>Geração de cadastro nos municípios selecionados: Em cada município selecionado, deve-se consultar a secretaria municipal de saúde para obter o cadastro de pessoas analfabetas ou de domicílios com pessoa analfabeta.</p> <p>Diversos cenários podem ser encontrados:</p> <ul style="list-style-type: none"> • Existem cadastros informatizados com indicação do domicílio e das pessoas analfabetas e os formatos destes cadastros podem variar por município em função de seus recursos de hardware e software; • O cadastro existente está distribuído entre as unidades de atenção básica à saúde (US); • Não existem cadastros informatizados mas há fichas ou pastas sobre os domicílios nas US; • Não existe cadastro informatizado e a cobertura do Siab é decorrente do cadastramento de clientes que procuram a US; e • Nada assegura não haver duplicidade de cadastramento de pessoas entre duas ou mais US. <p>Em qualquer hipótese, deve-se buscar padronização, gerando um arquivo para exercer o papel de cadastro de seleção. Para uma amostra de pessoas, cada registro do cadastro deve ser numerado seqüencialmente e indicar uma pessoa, seu endereço e sua condição de analfabeto ou não. Para uma amostra de domicílios, cada registro do cadastro deve ser numerado seqüencialmente e indicar um domicílio, seu endereço e seu número de analfabetos.</p> <p>Qualquer que seja o tipo de cadastro montado e diante da falta de certeza sobre a inexistência de duplicidades e sobre a cobertura do cadastramento feito no âmbito das US, a amostra não poderá estimar sem viés (desconhecido) o total de analfabetos do País.</p>	<p>Geração de cadastro nos municípios selecionados:</p> <p>Nesta hipótese, o cadastro é o de setores censitários do CD2000.</p> <p>Para cada setor existem mapas (formato pdf), descrição dos limites, dados sobre domicílios e população (do CD2000) e folhas de coleta do CD2000</p> <p>Os setores censitários podem ser as unidades secundárias de amostragem.</p> <p>Se dentre os objetivos estiver a estimação do total de analfabetos do País, é importante dar probabilidade de seleção a todos setores, mesmo os que têm poucos ou nenhum analfabeto. Não havendo este objetivo, a amostra pode ser limitada aos setores com um número mínimo de analfabetos (ou de domicílios com analfabeto residente).</p>
<p>Segundo estágio: Unidade de Saúde</p> <p>Para os municípios selecionados com mais de duas US, seria feita a seleção de US. Nos demais, todas as US entrariam na amostra.</p> <p>De acordo com o porte do município, seriam selecionadas de 2 a 6 US, com probabilidade proporcional ao número de domicílios com pelo menos um analfabeto, ou proporcional ao número de pessoas analfabetas.</p>	<p>Segundo estágio: setor censitário (n=200)</p> <p>Seleção com probabilidade proporcional ao seu número de analfabetos ou ao número de domicílios com pelo menos um analfabeto.</p> <p>Número de setores variando entre 2 e 6 por município, em função da população e do número de setores do município, condicionado à restrição de chegar a um total de 200 setores na amostra.</p>

Amostra do SIAB	Pesquisa domiciliar
<p>Esta decisão sobre seleção de domicílios ou analfabetos deve ser tomada anteriormente ao lançamento da pesquisa e será usada para todos os municípios selecionados.</p> <p>Em princípio, parece ser mais adequado a seleção de domicílios, visto que dados domiciliares serão coletados e analisados em conjunto com dados pessoais.</p> <p>No entanto, o que será dito para domicílios poderá ser aplicado a pessoas, sem perda de generalidade. Assim, em nome da simplicidade, o texto que se segue fará menção apenas à seleção de domicílios.</p> <p>O número de US por município selecionado será definido em função do porte do município selecionado, de forma a ter pelo menos 10 domicílios com analfabetos (ou 10 pessoas analfabetas) selecionados por US. Nos casos em que não há US (o atendimento é feito em município vizinho) ou há até 2 US, todas entrarão na amostra.</p>	<p>A seleção de setores pode ser feita de forma sistemática, após ordenação dos setores em função de renda ou de número de analfabetos, como forma de induzir uma estratificação implícita dos setores.</p> <p>Geração de cadastro nos setores selecionados: Não há necessidade de fazer uma listagem ou arrolamento de todos os domicílios da amostra, basta fazer uma contagem rápida de domicílios para ter o total atualizado de domicílios do setor.</p> <p>Pode-se apoiar esta operação com as folhas de coleta do CD2000 (FC2000). O uso das FC2000 assegurará a total cobertura do setor, o respeito ao percurso de observação do setor, e facilitará o registro sobre a eliminação de domicílios (demolição, tornou-se não-residencial, etc.) e sobre a criação de novos domicílios.</p> <p>As FC2000 são disponibilizadas por meio de imagens. Por razões ligadas ao sigilo estatístico, foram eliminados destas imagens os dados dos domicílios que indicam o nome do seu chefe e sua composição.</p>
<p>Terceiro estágio: domicílios (n=2000)</p> <p>Neste caso, como não será possível estimar o total de analfabetos do País, pode-se proceder à seleção de domicílios que tenham pelo menos um analfabeto diretamente dos cadastros gerados nos municípios selecionados.</p> <p>Esta operação pode ser feita em campo, por meio de uma planilha Excel, elaborada previamente para cada município selecionado. Esta planilha terá um único campo aberto, onde deve ser digitado o número total de domicílios com analfabetos. Uma vez digitado este número, uma coluna indicará os números de ordem dos domicílios a entrevistar e quantidade de domicílios a entrevistar.</p> <p>De posse destes números de ordem, volta-se ao cadastro da US para identificar a localização dos domicílios a entrevistar.</p> <p>As não-respostas deverão ser indicadas por meio do questionário (ou de uma folha de coleta da US) para serem processadas, a fim de permitir a correção de não-resposta no fator de expansão da amostra.</p>	<p>Terceiro estágio: domicílios (n=2000)</p> <p>O número de domicílios entrevistados por setor será fixado em 10 e será usado um procedimento de amostragem inversa, baseado na seleção prévia de 40 domicílios a serem visitados seqüencialmente até serem obtidas as 10 entrevistas realizadas ou até serem visitados todos os 40 domicílios.</p> <p>Também neste caso, será fornecida uma planilha para ser digitado o total de domicílios particulares contados no setor. A planilha indicará a seqüência de números de ordem dos 40 domicílios a visitar.</p> <p>De posse destes números de ordem, volta-se à FC2000 atualizada pela contagem rápida para identificar o endereço dos domicílios a visitar, respeitando a ordem de seleção.</p> <p>Deverá ser feita uma folha de coleta para indicar as ocorrências observadas na visita a cada um dos domicílios (e, portanto, a indicação de quantos foram visitados), inclusive a realização de entrevista, para viabilizar o cálculo do fator de expansão da amostra.</p>

Estas são as idéias centrais dos dois desenhos de amostra possíveis, com a indicação de que a pesquisa domiciliar é um desenho muito mais robusto, que produzirá estimativas mais confiáveis, para uma população amostrada mais bem definida.

Além disso, parece que seu custo deve ser da mesma ordem de grandeza do da pesquisa baseado no Siab, cujo tempo e custo de montagem de cadastro, quando possível, deve ser não-desprezível, podendo ser até maior do que o de uma contagem rápida, operação correspondente à montagem de cadastro nas US.

Anexo II

Quadros e Tabelas

Quadro 1 - Códigos de situação do setor.

Código	Descrição da situação
1	Área urbanizada de cidade ou vila
2	Área não-urbanizada de cidade ou vila
3	Área urbana isolada
4	Aglomerado rural de extensão urbana
5	Aglomerado rural isolado – povoado
6	Aglomerado rural isolado – núcleo
7	Aglomerado rural isolado outros aglomerados
8	Zona rural, exclusive aglomerado rural

Quadro 2 - Códigos de tipo do setor.

Código	Descrição do tipo do setor
0	Comum ou não especial
1	Especial de aglomerado subnormal
2	Especial de quartéis, bases militares, etc.
3	Especial de alojamento, acampamentos, etc.
4	Especial de embarcações, barcos, navios, etc.
5	Especial de aldeia indígena
6	Especial de penitenciárias, colônias penais, presídios, cadeias, etc.
7	Especial de asilos, orfanatos, conventos, hospitais, etc.

Tabela 1 – Número e proporção de setores, domicílios particulares permanentes (DPP) e pessoas residentes em DPP, segundo situação em relação à inclusão no cadastro de seleção da pesquisa do projeto AIBF.

Macror-região	Situação do setor no cadastro de seleção	Setores		Domicílios		População	
		Contag em	Proporção (%)	Contagem	Proporção (%)	Contagem	Proporção (%)
Norte	Mantido	8.409	58,6%	2.062.419	72,1%	9.012.486	69,9%
	Excluído - tipo especial	133	0,9%	6.784	0,2%	23.084	0,2%
	Excluído - situação rural	5.811	40,5%	793.063	27,7%	3.865.134	30,0%
	Total na Região	14.353	6,7%	2.862.266	6,3%	12.900.704	7,6%
Nordeste	Mantido	53.506	99,7%	11.511.594	99,8%	47.701.703	99,9%
	Excluído - tipo especial	181	0,3%	20.425	0,2%	40.008	0,1%
	Total na Região	53.687	24,9%	11.532.019	25,3%	47.741.711	28,1%
Sudeste	Mantido	95.628	99,4%	20.454.633	99,4%	72.285.368	99,8%
	Excluído - tipo especial	613	0,6%	114.159	0,6%	127.043	0,2%
	Total na Região	96.241	44,6%	20.568.792	45,2%	72.412.411	42,6%
Sul	Mantido	36.679	99,7%	7.303.887	99,8%	25.082.119	99,9%
	Excluído - tipo especial	121	0,3%	15.930	0,2%	25.497	0,1%
	Total na Região	36.800	17,1%	7.319.817	16,1%	25.107.616	14,8%
Centro-oeste	Mantido	14.495	98,4%	3.213.153	99,6%	11.604.491	99,7%
	Excluído - tipo especial	235	1,6%	11.469	0,4%	32.237	0,3%
	Total na Região	14.730	6,8%	3.224.622	7,1%	11.636.728	6,9%
Total	Mantido	208.717	96,7%	44.545.686	97,9%	165.686.167	97,6%
	Excluído - tipo especial	1.283	0,6%	168.767	0,4%	247.869	0,1%
	Excluído - situação rural	5.811	2,7%	793.063	1,7%	3.865.134	2,3%
	Total Geral	215.811	100,0%	45.507.516	100,0%	169.799.170	100,0%

Fonte: Arquivo Agregado por Setores Censitários dos Resultados do Universo do Censo Demográfico 2000

Tabela 2 – Resumo da amostra de setores nos municípios grandes.

Macrorregião	Número de Municípios	Setores por Município	Total de Setores na Amostra
Norte + Centro-Oeste	7	18	126
Nordeste	11	15	165
Sudeste + Sul	23	15	345
Total	41	–	636

Tabela 3 – Distribuição de frequência do número de municípios por grupo.

Número de Municípios no Grupo	Número de Grupos
1	253
2	265
3	245
4	213
5	151
6	97
7	84
8	42
9	20
10	13
11	14
12	8
13	3
14	3
15	2
16	2
19	1
20	1
21	1
30	1
41	1
Total	1.420

Tabela 4 – Descrição dos estratos de UPAs, no estrado dos municípios pequenos, e alocação da amostra nestes estratos.

Grande Área	Estrato	Limites		Número de UPAs	
		Inferior	Superior	Na população	Na Amostra
Nordeste	1	0,0812	0,2663	240	5
Nordeste	2	0,2664	0,3132	119	7
Nordeste	3	0,3133	0,4444	121	13
Norte e Centro-Oeste	1	0,0290	0,1296	78	5
Norte e Centro-Oeste	2	0,1297	0,1857	40	6
Norte e Centro-Oeste	3	0,1858	0,8033	39	12
Sudeste e Sul	1	0,0163	0,0938	392	4
Sudeste e Sul	2	0,0939	0,1388	195	5

Tabela 5 – Resumo da amostra de setores nos municípios pequenos

Grande Área	Número de UPAs	Setores por Município	Total de Setores na Amostra
Norte + Centro-Oeste	23	12	76
Nordeste	25	12	300
Sudeste e Sul	17	12	204
Total	65	–	780

Tabela 6 – Dados da população e alocação geral da amostra

Grande área	População					Amostra		Fração Amostral			Parâmetros			
	Tipo de município	Número de pessoas	Número de domicílios particulares permanentes	Número de setores	Número de municípios ou grupos	Número de municípios	Número de municípios ou grupos	Número total de setores	Número de domicílios (Por Mil)	Domicílios (Por Mil)	Setores (Por Mil)	Setores por grupo ou município	Domicílios por setor	Área (km²)
N+CO	Grandes	7.371.138	1.889.755	7.934	7	7	7	126	1.512	0,80	15,88	18	12	30.827
	UPAs	13.270.671	3.304.221	14.767	157	888	23	276	3.312	1,00	18,69	12	12	5.352.009
	Total	20.641.809	5.193.976	22.701	164	895	30	402	4.824	0,93	17,71	-	-	5.382.836
NE	Grandes	11.226.902	2.841.500	10.688	11	12	11	165	1.980	0,70	15,44	15	12	6.522
	UPAs	36.487.506	8.555.726	42.675	480	1.775	25	300	3.600	0,42	7,03	12	12	1.542.717
	Total	47.714.408	11.397.226	53.363	491	1.787	36	465	5.580	0,49	8,71	-	-	1.549.239
SE+S	Grandes	32.937.172	9.577.682	42.456	23	23	23	345	3.450	0,36	8,13	15	10	16.806
	UPAs	64.484.468	17.845.901	89.072	783	2.802	17	204	2.040	0,11	2,29	12	10	1.460.591
	Total	97.421.640	27.423.583	131.528	806	2.825	40	549	5.490	0,20	4,17	-	-	1.477.397
Total	Grandes	51.535.212	14.308.937	61.078	41	42	41	636	6.942	0,49	10,41	-	-	54.155
	UPAs	114.242.645	29.705.848	146.514	1.420	5.465	65	780	8.952	0,30	5,32	-	-	8.355.317
	Total	165.777.857	44.014.785	207.592	1.461	5.507	106	1.416	15.894	0,36	6,82	-	-	8.409.472

Anexo III

Algoritmo para amostragem seqüencial de Poisson

Passo 1: obtenha para cada unidade no cadastro um número pseudo-aleatório, isto é, um número gerado com base na distribuição Uniforme no intervalo (0; 1), de forma que a geração para diferentes unidades no cadastro seja independente. Seja X_i o número aleatório gerado para a unidade i do cadastro, $i=1, 2, \dots, N$.

Passo 2: Em seguida, calcule os valores relativos da medida de tamanho t a ser empregada para selecionar a amostra.

Denotando por T_i o tamanho t da unidade i do cadastro, calcule os valores de $N p_i = T_i / \sum_{k=1}^N T_k$ para todo $i=1, 2, \dots, N$. Observe que os valores p_i satisfazem à relação:

$$\sum_{i=1}^N p_i = 1$$

- a) Calcule os “números aleatórios modificados” dados por $Z_i = X_i / p_i$ para todas as unidades do cadastro, $i=1, 2, \dots, N$.
- b) Ordene as unidades do cadastro segundo os números aleatórios modificados, do menor para o maior valor.
- c) Inclua na amostra as n primeiras unidades na lista ordenada, isto é, aquelas com os n menores valores dos números aleatórios modificados Z_i .

Coleção Educação para Todos

Volume 01: Educação de Jovens e Adultos: uma memória contemporânea, 1996-2004

Volume 02: Educação Anti-racista: caminhos abertos pela Lei Federal nº 10.639/03

Volume 03: Construção Coletiva: contribuições à educação de jovens e adultos

Volume 04: Educação Popular na América Latina: diálogos e perspectivas

Volume 05: Ações Afirmativas e Combate ao Racismo nas Américas

Volume 06: História da Educação do Negro e Outras Histórias

Volume 07: Educação como Exercício de Diversidade

Volume 08: Formação de Professores Indígenas: repensando trajetórias

Volume 09: Dimensões da Inclusão no Ensino Médio: mercado de trabalho, religiosidade e educação quilombola

Volume 10: Olhares Feministas

Volume 11: Trajetória e Políticas para o Ensino das Artes no Brasil: anais da XV CONFAEB

Volume 12, Série Vias dos Saberes n. 1: O Índio Brasileiro: o que você precisa saber sobre os povos indígenas no Brasil de hoje

Volume 13, Série Vias dos Saberes n. 2: A Presença Indígena na Formação do Brasil

Volume 14, Série Vias dos Saberes n. 3: Povos Indígenas e a Lei dos “Branços”: o direito à diferença

- Volume 15, Série Vias dos Saberes n. 4:** Manual de Lingüística: subsídios para a formação de professores indígenas na área de linguagem
- Volume 16:** Juventude e Contemporaneidade
- Volume 17:** Católicos Radicais no Brasil
- Volume 18, Série Avaliação n. 1:** Brasil Alfabetizado: caminhos da avaliação
- Volume 19, Série Avaliação n. 2:** Brasil Alfabetizado: a experiência de campo de 2004
- Volume 20, Série Avaliação n. 3:** Brasil Alfabetizado: marco referencial para avaliação cognitiva
- Volume 21, Série Avaliação n. 4:** Brasil Alfabetizado: como entrevistamos em 2006
- Volume 22, Série Avaliação n. 5:** Brasil Alfabetizado: experiências de avaliação dos parceiros
- Volume 23, Série Avaliação n. 6:** O que fazem as escolas que dizem que fazem Educação Ambiental?
- Volume 24, Série Avaliação n. 7:** Diversidade na Educação: experiências de formação continuada de professores
- Volume 25, Série Avaliação n. 8:** Diversidade na Educação: como indicar as diferenças?
- Volume 26:** Pensar o Ambiente: bases filosóficas para a Educação Ambiental
- Volume 27:** Juventudes: outros olhares sobre a diversidade
- Volume 28:** Educação na Diversidade: experiências e desafios na educação intercultural bilíngüe
- Volume 29:** O Programa Diversidade na Universidade e a Construção de uma Política Educacional Anti-racista
- Volume 30:** Acesso e Permanência da População Negra no Ensino Superior

ISBN 85-98171-63-8



9 788598 171630



ipea

Ministério
da Educação

