

AVALIAÇÃO DA TENDÊNCIA DO DESEMPENHO DE CANDIDATOS REPROVADOS NO EXAME DE CERTIFICAÇÃO, UTILIZANDO MODELOS HIERÁRQUICOS MULTIVARIADOS

*Lilia Carolina Carneiro da Costa, Nilo Batista da Silva Junior
y Paulo Roberto Pinheiro Leal*

1. INTRODUÇÃO

Entre os anos de 1991 a 2000, a gestão governamental deparou-se com o seguinte panorama educacional: a taxa de analfabetismo caiu de 35,2% para 21,6% e houve um acréscimo na matrícula de 62,7% no ensino fundamental e 255,9% no ensino médio (Reis, 2003). Entretanto, se por um lado o ensino na Bahia crescia em termo quantitativo, com cerca de 98% das crianças em idade escolar matriculadas e cursando, por outro lado, observavam-se um baixo aproveitamento escolar, altos índices de repetência, evasão escolar, defasagem idade/série, configurando, assim, um dos piores indicadores de aprendizagem do país e do mundo. Nesse cenário, a administração pública, iniciada em 1999, elegeu a Educação como prioridade absoluta, criando o Projeto Educar para Vencer, cujo objetivo era oferecer aos cidadãos uma educação pública de qualidade (Hashimoto, 2003).

O programa era composto por seis projetos prioritários, articulados entre si, a saber:

- Regularização do Fluxo Escolar de 1^a. à 4^a. série e Regularização do Fluxo Escolar de 5^a. à 8^a. Série.- que visam reduzir o alto índice de defasagem idade/série;
- Gestão Educacional.- que tem por finalidade fortalecer a capacidade local de planejamento e gerenciamento para o desenvolvimento da autonomia e melhoria da qualidade dos serviços educacionais prestados à população;
- Certificação Ocupacional de Profissionais da Educação.- objeto desse artigo, é um sistema que estabelece padrões de mérito e competência para a escolha de seus profissionais;
- Avaliação Externa.- objetiva avaliar o desempenho do sistema educacional através da aplicação de teste aos alunos do ensino fundamental nas disciplinas centrais do currículo;
- Programa de Enriquecimento Instrumental (PEI).- criado pelo psicólogo romeno Reuven Feuerstein, um aplicativo da Teoria da Modificabilidade Cognitiva Estrutural —TMCE— que utiliza como estratégia metodológica a Experiência de Aprendizagem Mediada —EAM— visando ampliar o potencial cognitivo do ser humano.

“As ações desenvolvidas pelos seis projetos colocaram o aluno no centro da reforma, a escola como o eixo da mudança e buscaram assegurar condições adequadas de funcionamento para cada espaço educativo estadual, dentro de um padrão de qualidade discutido e determinado pelos profissionais da educação do próprio Estado.” (Reis, 2003).

A Secretaria de Educação do Estado da Bahia estabeleceu parcerias estratégicas com diferentes entidades com a finalidade de concretizar esse programa. Uma dessas entidades que se engajou em vários projetos do Programa Educar para Vencer foi a Fundação Luís Eduardo Magalhães (FLEM) – Centro de Modernização e Desenvolvimento da Administração Pública. Esta é uma empresa pública de direito privado, sem fins lucrativos, que tem o compromisso de estimular o

desenvolvimento de soluções inovadoras para a administração pública e uma nova governança (Fundação Luís Eduardo Magalhães, 2007).

Em janeiro de 2000, a FLEM desenvolveu o Projeto de Certificação Ocupacional para Profissionais de Educação. Como o Brasil não possuía experiência significativa em certificação ocupacional, a FLEM contratou o AIR - *American Institutes for Research*, instituição americana, com 50 anos de experiência em certificação ocupacional, cujas responsabilidades foram as de transferir tecnologia e formar especialistas.

Esse projeto desenvolveu o Exame de Certificação Ocupacional para Dirigente Escolar, que avalia se o candidato possui as competências necessárias para o exercício do cargo de diretor e vice-diretor de escola. O público alvo, utilizado nesse estudo, foram os servidores públicos da Secretaria de Educação do Estado da Bahia e Secretarias Municipais de Feira de Santana, Ilhéus, Itaparica, Santo Antônio de Jesus e Conde.

Esse exame é composto de três etapas eliminatórias: TCE – Teste de Conhecimentos Específicos; TCELP – Teste de Comunicação e Expressão em Língua Portuguesa e QP – Teste de Questões Práticas. O candidato reprovado em uma etapa pode repeti-las quantas vezes forem oferecidas, em um prazo de dois anos. Frequentemente, os testes são realizados duas vezes ao ano.

O respectivo estudo tem como objetivo avaliar a tendência do desempenho dos candidatos reprovados, ou seja, verificar se melhoram, pioram ou mantêm seu nível de conhecimento/habilidade de uma versão para outra do teste. Também, esse artigo apresenta uma análise da relação entre essa tendência nos dois testes TCE e QP e as características demográficas dos candidatos, como sexo e idade, por meio de modelos hierárquicos multivariados.

Na seção 2, apresentam-se o Exame de Certificação Ocupacional e a origem dos dados. Na seção 3, encontra-se a formulação teórica do modelo Hierárquico Bayesiano Bivariado utilizado nesse estudo. A seção 4 traz os resultados e a discussão das análises. A seção 5 apresenta as conclusões finais.

2. EXAME DE CERTIFICAÇÃO OCUPACIONAL

A certificação ocupacional é um processo de avaliação desenvolvido para atestar que os profissionais possuem as competências necessárias ao desempenho das atividades relacionadas com as suas ocupações. Entende-se por competências um conjunto de conhecimentos, habilidades e atitudes indispensáveis ao exercício de uma atividade específica, em que:

- CONHECIMENTO é a dimensão do saber;
- HABILIDADE é a dimensão do saber fazer – prática ocupacional;
- ATITUDE é a dimensão do querer fazer, envolvendo aspectos motivacionais, sociais e afetivos, referentes ao trabalho.

O processo de certificação ocupacional está relacionado às competências efetivas dos profissionais no trabalho e não somente aos aspectos formais de titulação, formação, etc. (Fundação Luís Eduardo Magalhães, 2000). A constante evolução nos processos de trabalhos, devido a novos estudos científicos na área ou a introdução de novas tecnologias, faz com que os diplomas tradicionais, que algumas vezes são obtidos há décadas, possam não mais representar a realidade atual do trabalho.

Ainda, a certificação ocupacional se diferencia de uma avaliação de desempenho. Enquanto a primeira atesta se o profissional possui as competências para exercer certo cargo, a segunda verifica se o profissional está cumprindo com as responsabilidades e compromissos assumidos com o empregador, como afirma Marras (2002 *apud* Hashimoto, 2003) sobre avaliação de desempenho:

“é um instrumento gerencial que permite ao administrador mensurar os resultados obtidos por um empregado ou por um grupo, em período e área específicos (conhecimentos, metas, habilidades, comportamentos, e outros)”

Outros processos de avaliação, como vestibular e concursos públicos, também são confundidos com a certificação. Nesses processos, deseja-se selecionar os melhores candidatos para uma quantidade específica de vagas. Já na certificação, os candidatos não são ordenados, segundo seu desempenho, e todos aqueles que demonstram ter competências são aprovados. Uma analogia à certificação é o exame para obter a carteira de motorista, em que não há um número limitado de carteiras a serem emitidas e, entre os candidatos aprovados, não se sabe quem é melhor motorista.

Para que a certificação possa atingir o seu objetivo de avaliar se um candidato detém ou não as competências necessárias ao exercício de um cargo, o exame tem que refletir fielmente a ocupação em questão. Assim, o primeiro passo para a elaboração do processo de certificação é o desenvolvimento dos Padrões de Competências – documento que expressa as competências (conhecimentos, habilidades e atitudes) requeridas para o cargo analisado, como afirma Hashimoto (2003):

“O sistema desenvolvido para os profissionais de educação teve como premissa básica, desenvolver padrões nacionais que refletissem a praxis educacional em todas as partes do país, apesar de sua diversidade social, econômica e cultural.

Em se tratando de cargos e funções não operacionais, esta atividade reveste-se de maior complexidade, tendo em vista que os cargos a serem certificados caracterizam-se pela variedade, abstração, incerteza e criatividade.”

Em seguida, são desenvolvidas as “especificações do teste”, matriz que lista os domínios de conhecimentos, habilidades e atitudes a serem medidos, além da natureza e da quantidade de questões dos testes.

Mediante essas especificações, o teste é então construído, como D’Almeida (2003) descreve:

“Esta etapa inclui diversas ações, tais como: redação de itens originais que irão compor os testes de certificação; revisão e modificação, quando necessária, desses quesitos de prova, por um comitê de especialistas no conteúdo avaliado; pré-teste dos itens através da aplicação dos mesmos a uma população semelhante à qual a certificação se destina ou pela análise crítica estruturada de uma banca de especialistas; análise dos dados psicométricos obtidos através do pré-teste, para controle da qualidade dos itens desenvolvidos e escolha daqueles que apresentam melhor desempenho; refinamento dos itens cujo desempenho não tenha sido satisfatório no pré-teste e composição final dos cadernos de teste.”

Com o teste pronto, é definida a linha de corte, isto é, o desempenho mínimo que o candidato precisa ter no teste para ser aprovado. Existem na literatura diversos métodos que podem ser utilizados. Um dos mais difundidos é o método Angoff (Crocker; Algina, 1986). Após o teste aplicado, é realizada a pontuação que varia de acordo com a natureza do teste.

O exame de certificação ocupacional é aplicado em vários momentos distintos, como por exemplo, uma vez por ano. Dessa forma, os candidatos respondem a uma versão diferente do mesmo exame. O uso de diferentes versões pode conduzir a testes com diversos graus de dificuldade, apesar

de terem sido construídos para serem similares em dificuldade e conteúdo. Assim, todas as versões do exame de certificação ocupacional passam por um processo de equalização - um processo estatístico usado para ajustar as notas dos examinados, de modo que sejam permutáveis (Kolen; Brennan, 1995). Esse processo tem por finalidade tornar o exame justo, em que o resultado do candidato dependa apenas da sua competência e não se ele respondeu uma versão mais fácil ou mais difícil.

Todo o processo de certificação ocupacional é desenvolvido por comitês formados por especialistas dos setores público e privado com reconhecida competência. Esse processo passa ainda pelo crivo de uma câmara, composta por profissionais de renomado prestígio no cenário da educação no Brasil. As principais responsabilidades desse grupo são: assegurar e atestar a exatidão técnica das ações, assegurar a legitimidade do processo, contribuir para a projeção do Sistema de Certificação em âmbito nacional, decidir sobre possíveis áreas de expansão e firmar todos os documentos de Certificação (Agência De Certificação Ocupacional, 2006).

A certificação ocupacional traz vantagens tanto para a sociedade que tem o profissional com competência comprovada prestando serviço, quanto para o candidato, que, dentre outras coisas, tem uma maior consciência do seu papel profissional, com o perfil do cargo definido e declarado publicamente. Além disso, o candidato certificado tem reconhecimento público de suas competências e qualificação profissional contínua, uma vez que ele deve ser recertificado a cada dois anos, comprovando, assim, a atualidade de sua competência.

O Projeto de Certificação Ocupacional de Profissionais da Educação teve início com o exame de Dirigentes Escolares, em 2001, no qual se buscaram introduzir critérios objetivos de mérito e competência profissional para a escolha de diretores e vice-diretores, como relata Ferreira (2003):

“... os dirigentes se mantinham insatisfeitos com os processos de seleção para ocupação dos cargos de diretor e vice-diretor, devido ao fato de preponderar às influências políticas partidárias em vez da meritocracia.”

Ainda com a Secretaria de Educação do Estado da Bahia, foi desenvolvida a certificação ocupacional dos Professores Alfabetizadores, Professores de Disciplinas para as séries finais do Ensino Fundamental e Ensino Médio, Coordenadores Pedagógicos e Secretários Escolares.

Em 2002, em parceria com a Secretaria Municipal de Educação de Campo Grande (MS), aconteceu o primeiro exame de certificação ocupacional para Dirigentes Escolares.

Em 2003, a *National Organization for Competency Assurance (Noca)*, comissão internacional que congrega o restrito universo das 500 principais agências de certificação ocupacional de todo o mundo, reconheceu o padrão internacional desse projeto de certificação. Dessa forma, em 2004, o projeto passou a integrar a estrutura da FLEM na condição de uma Agência de Certificação Ocupacional (Acert).

Desde o início do projeto, em 2001, já foram certificados 2.770 Dirigentes Escolares em 7 exames aplicados, 2.176 Professores Alfabetizadores em 5 exames, 9.106 Professores de Disciplinas em 3 exames e 542 Coordenadores Pedagógicos em 2 exames.

Em 2004, a Secretaria da Administração do Estado da Bahia (Saeb) contratou a FLEM para certificar os profissionais da gestão pública, envolvendo inicialmente as carreiras de: Especialista em Políticas Públicas e Gestão Governamental; Técnico em Fiscalização; Especialista em Fiscalização; Jornalista; Analista Técnico; Analista Universitário; Analista em Radiodifusão; Analista em Orientação Social; e Analista em Infra-Estrutura e Transportes. Também está incluída, neste contrato,

a certificação para a função de Presidente e Membro de Comissão de Licitação e Pregoeiro (Agência de Certificação Ocupacional, 2006).

O primeiro cargo que foi certificado e fruto desse artigo é o de Dirigente Escolar. Isto porque, hoje, já se sabe que o dirigente escolar tem um papel fundamental para se construir uma escola eficaz, como relata Reis (2003),

“Inúmeras tentativas de melhorar a eficiência da educação pública baiana revelaram a necessidade de colocar a escola como foco central das políticas educacionais. Nesse sentido, surgiu uma série de estudos que evidenciaram a importância da gestão como um dos fatores determinantes para o sucesso escolar. Os estudos referidos colocaram em pauta temas como liderança do dirigente escolar, democratização da gestão, clima escolar, planejamento e aperfeiçoamento contínuo das equipes envolvidas.”

Ainda, Lück (2000) comenta que o maior objetivo de uma escola é a melhoria do desempenho de aprendizagem de seus alunos e, este, está freqüentemente associado aos seguintes fatores: liderança educacional, flexibilidade e autonomia, clima escolar, apoio da comunidade, processo ensino aprendizagem, avaliação do desempenho acadêmico, supervisão de professores, materiais e textos de apoio pedagógico e espaço adequado.

O Grupo de Avaliação e Medidas Educacionais (2002) também mostra a importância do cargo de dirigente escolar. Este fez um estudo de caso em três escolas da rede pública de ensino do Estado de Minas Gerais. Essas escolas foram selecionadas de acordo com graus variados de desempenho de seus alunos em Matemática e Língua Portuguesa no PROEB/SIMAVE-2000. O objetivo desse estudo foi descobrir se as razões do sucesso ou fracasso dessas escolas podem ser explicadas pelos mesmos fatores apontados pela literatura estrangeira. Dentre outras coisas, o estudo chegou à seguinte conclusão:

“A liderança administrativa é considerada por Sammons et al. (1995) como a mensagem mais clara das pesquisas sobre “Escola Eficaz”. A análise dos dados desta pesquisa indica o mesmo caminho. Os exemplos da Escola Azul, que possuiu uma diretora com espírito de liderança muito grande, envolvida com os assuntos da escola e que procura realizar uma administração participativa, e do Colégio Verde, que apresenta uma estrutura administrativa muito bem organizada, mostram como uma liderança ativa dentro da escola é fundamental para o seu bom funcionamento. Por outro lado, a situação da Escola Amarela, que tem uma diretora pouco envolvida com as questões escolares, reforça a mesma idéia na medida em que essa escola se mostra incapaz de colocar em prática qualquer proposta pedagógica.”¹

Em suma, os resultados da pesquisa mostraram que a liderança administrativa ocupa um papel central dentro das escolas. Uma liderança que apresente como características o envolvimento com as questões escolares e a preocupação com a construção de uma participação dos outros profissionais de ensino na escola, entre outras, é capaz tanto de, pelo menos em parte, reverter as situações adversas colocadas pela realidade socioeconômica, quanto de colocar em prática propostas pedagógicas que podem modificar os resultados acadêmicos dos estudantes.”

O Dirigente Escolar deve possuir competências que lhe permitam promover a educação de todos os alunos, o desenvolvimento profissional dos professores e o funcionamento da escola integrada à comunidade. Dessa forma, após uma análise do cargo de dirigente, o conjunto das

¹ Em Português, a Escola Amarela ficou em um nível inferior na escala de proficiência, seguida pela Escola Azul em uma posição intermediária e pelo Colégio Verde em um patamar mais elevado. Em Matemática, a Escola Amarela permaneceu em um nível inferior, ficando a Escola Azul e o Colégio Verde na posição seguinte da escala de proficiência.

expectativas a respeito das suas funções e do seu papel foram agrupados em 6 categorias (Agência de Certificação Ocupacional, 2006):

Padrão I: Desenvolvimento Integral dos Alunos e Implementação do Currículo.

Padrão II: Desenvolvimento da Equipe.

Padrão III: Administração da Escola.

Padrão IV: Visão e Planejamento Estratégico.

Padrão V: Estabelecimento e Ampliação de Relacionamentos.

Padrão VI: Gestão do Processo Pedagógico.

Os três primeiros padrões são avaliados na primeira etapa do exame de dirigente escolar, o TCE - Teste de Conhecimentos Específicos, e os três últimos na terceira etapa, o QP – Teste de Questões Práticas. Como os padrões expressam toda a diversidade do trabalho do dirigente escolar, esses dois testes foram usados nesse estudo.

A segunda etapa do exame é o TCELP – Teste de Comunicação e Expressão em Língua Portuguesa que avalia conhecimentos e habilidades em Língua Portuguesa.

3. MODELO HIERÁRQUICO BAYESIANO BIVARIADO

Alguns dados são coletados de modo que apresentem uma estrutura hierárquica, como por exemplo, uma localidade possui várias escolas, uma escola possui várias turmas e uma turma possui vários alunos.

Gelman *et al* (1995) afirmam que é inadequado modelar dados hierárquicos de maneira não-hierárquica. Se, por exemplo, tem-se um modelo para cada escola, então se estimam vários parâmetros, produzindo modelos que podem ajustar bem os dados existentes, mas gerar previsões inferiores para novos dados. Ou ainda, quando se trabalha com grandes bancos de dados e se ajusta um modelo único para todas as escolas, com poucos parâmetros não se consegue um modelo com um ajuste acurado. Com os modelos hierárquicos, esses problemas são evitados usando uma estrutura de dependência entre os parâmetros populacionais.

Para melhor entendimento desse modelo, considera-se um exemplo típico em educação: Y_{ij} = proficiência do aluno i da escola j (variável dependente) e X_{ij} = sexo do aluno i da escola j (variável independente). O modelo de regressão clássico em que os parâmetros são considerados fixos é escrito da seguinte maneira:

$$Y_{ij} = \alpha + \beta * X_{ij} + \varepsilon_{ij},$$

onde ε_{ij} 's, erros de medida, são independentes e com distribuição Normal com média 0 e variância σ^2 .

No caso do modelo hierárquico (*cf.* Bryk & Raudenbush, 1992), ou modelo multinível (*cf.* Goldstein, 1995) com o intercepto aleatório, tem-se:

$$Y_{ij} = \alpha_j + \beta * X_{ij} + \varepsilon_{ij},$$

$$\alpha_j = \eta_0 + \eta_1 * W_j + v_j,$$

onde W_j é uma variável independente associada à escola j , como por exemplo, rede municipal,

estadual ou particular e v_j 's, os erros aleatórios do nível 2 associados ao intercepto, são independentes com distribuição normal com média 0 e variância σ_v^2 . Pressupõe-se que o erro ε , nível do aluno, seja independente do erro v , nível da escola. β , η_0 e η_1 são os parâmetros fixos a serem estimados. Também é possível considerar β como um parâmetro aleatório e escrevê-lo como em α_j .

Desejando utilizar mais variáveis dependentes no modelo, pode-se trabalhar com modelos hierárquicos multivariados (Congdon, 2001). Suponha observações contínuas Y_{ij} , para $i = 1, \dots, N$ (indivíduos) e $j = 1$ ou 2 (variáveis dependentes), agrupadas em L categorias. Utilizando ainda o exemplo da escola, Y_{i1} = proficiência do aluno i em português, Y_{i2} = proficiência do aluno i em matemática e os alunos estão distribuídos em L escolas.

Desejando-se estudar a relação existente entre essas observações e as covariáveis X_{i1}, \dots, X_{ic} e assumindo um modelo bivariado, então

$$Y_i \sim \text{Normal}_2(\mu_i, P)$$

onde

$$Y_i = (Y_{i1}, Y_{i2});$$

$$i = 1, \dots, N;$$

P = matriz de dispersão 2×2 para as observações;

$\mu_i = (\mu_{i1}, \mu_{i2})$; μ_{ij} = média do indivíduo i na variável j , i.e. a proficiência média do aluno i em português ($j=1$) ou em matemática ($j=2$), escrita como

$$\mu_{ij} = \alpha_{l(i)j} + \beta_{1j} * X_{i1} + \dots + \beta_{cj} * X_{ic}$$

Nesse modelo, β_{cj} é denotada como efeito fixo ($c = 1, \dots, C$ e $j = 1$ ou 2) e $\alpha_{l(i)j}$ como efeito aleatório, com $l(i)$ sendo a escola que o aluno i pertence. Assim, para $l = 1, \dots, L$ categorias, é assumida para α_l uma distribuição normal bivariada

$$\alpha_l \sim \text{Normal}_2(\eta, Q)$$

$$\text{onde } \alpha_l = (\alpha_{l1}, \alpha_{l2});$$

$$\eta = (\eta_1, \eta_2);$$

Q = matriz de dispersão 2×2 para os efeitos das categorias.

A seguinte interpretação pode ser dada aos parâmetros do modelo: β_{cj} é um parâmetro de efeito fixo, ou seja, assume-se que a relação entre a covariável X_c e a proficiência em português ou matemática, Y , seja a mesma para todas as escolas; α_l é o intercepto aleatório, em que cada escola terá o seu o nível médio de proficiência estimado; η é o nível médio geral de proficiência de todas as escolas; P e Q são os componentes de variância do modelo.

O modelo hierárquico bayesiano requer a especificação de uma distribuição a priori para os parâmetros do modelo. Informações anteriores podem ser incorporadas ao modelo através dessa distribuição, como por exemplo, $\beta \sim N(A, B)$, sendo A e B valores específicos. Pode-se também atribuir uma priori vaga, i.e., uma densidade a priori uniforme. Ou ainda, uma distribuição a priori não-informativa freqüentemente usada que é $\sigma^2 \sim \text{inversa-}\chi^2$ com um número pequeno no grau de liberdade como 2 (Gelman *et al*, 1995). Em inferência clássica, têm-se os estimadores dos parâmetros; já em inferência bayesiana, tem-se a distribuição a posteriori que pode fornecer qualquer estatística desejada a respeito da distribuição dos parâmetros. Um possível estimador é a média a posteriori (Migon y

Gamerman, 1999).

Como o Hewson (2005) cita, existe uma gama de procedimentos para ajuste dos modelos hierárquicos. Primeiramente, Goldstein (2003) mostra vários procedimentos baseados no método iterativo mínimos quadrados generalizado (*Iterative Generalised Least Squares* - IGLS). Goldstein and Rasbash (1992) estabeleceu o procedimento de máxima verossimilhança e, assim, muitos outros procedimentos foram desenvolvidos, utilizando uma aproximação desse. Além de métodos baseados em MCMC (Gamerman y Lopes, 2006).

Igualmente, há softwares com esses procedimentos incorporados, como R *software* (R Development Core Team, 2004); o MLn, desenvolvido e descrito por Rasbash *et al.* (1991) e vários exemplos, utilizando esse pacote, podem ser encontrados em Kreft e de Leeum (1998); e WinBUGS (Spiegelhalter *et al.*, 1998) o qual Congdon (2001) apresenta várias aplicações de modelos hierárquicos, utilizando esse software.

4. RESULTADOS E DISCUSSÃO

O Exame de Certificação Ocupacional para Dirigente Escolar foi aplicado 7 vezes, sendo a primeira vez em 2001. O número de candidatos que responderam pelo menos a uma versão é de 8.257.

4.1. Teste de Conhecimentos Específicos - TCE

A primeira etapa, TCE, é um teste de múltipla-escolha, com cerca de 60 questões, nas quais o desempenho do candidato é avaliado por meio da Teoria da Resposta ao Item, usando o modelo Rasch, com o parâmetro do grau de dificuldade, b , variando em uma escala de média 0 e desvio padrão 1 (Hambleton *et al.*, 1991). O processo de equalização do TCE é feito mediante a colocação de alguns itens comuns nas diferentes versões. Dessa forma, o desempenho dos candidatos pode ser comparado, independente da versão a que eles responderam (Andrade *et al.*, 2000).

Essa primeira análise do estudo foi realizada utilizando o *software* SPSS (SPSS, 1995). A tabela 1 traz o desempenho dos candidatos no TCE em cada exame realizado. O 3º, 4º e 5º exames apresentaram os piores desempenhos e tiveram o maior percentual de candidatos repetentes.

Observou-se ainda que o desempenho médio dos candidatos que estavam respondendo pela primeira vez ao exame diferiu, significativamente ao nível de 5%, do desempenho médio dos que estavam repetindo, sendo essa diferença positiva. Entretanto, o 2º exame foi uma exceção, visto que essa diferença não foi significativa.

TABELA 1. ANÁLISE DO DESEMPENHO DOS CANDIDATOS NO TCE.

Exame	Geral			Respondendo pela 1ª vez (P)		Repetentes (R)		Intervalo de Confiança de 95% para a diferença das médias: P - R	
	Nº	Média	Desvio Padrão	Nº	Média	Nº	Média	LI	LS
1º	1.979	0,975	0,6119	1.979 (100%)	0,975	0 (0%)	--	--	--
2º	4.062	0,738	0,7404	3.661 (90,1%)	0,736	401 (9,9%)	0,762	-0,102	0,050
3º	1.896	0,222	0,7057	595 (31,4%)	0,578	1.301 (68,6%)	0,059	0,455	0,583
4º	1.189	0,310	0,7690	338 (28,4%)	0,873	851 (71,6%)	0,086	0,701	0,873
5º	878	0,273	0,6853	211 (24,0%)	0,800	667 (76,0%)	0,106	0,598	0,789
6º	515	1,095	0,7186	433 (84,1%)	1,228	82 (15,9%)	0,392	0,683	0,990
7º	606	1,346	0,7135	560 (92,4%)	1,408	46 (7,6%)	0,587	0,616	1,025

Os candidatos que obtiveram um desempenho acima de 0,74 foram considerados aprovados. Aqueles que foram reprovados, poderiam repetir o exame em no máximo 4 vezes. Mediante a tabela 2, tem-se que a maioria dos candidatos só respondeu apenas a uma versão (73,3%), grande parte porque já havia sido aprovado (82,0%).

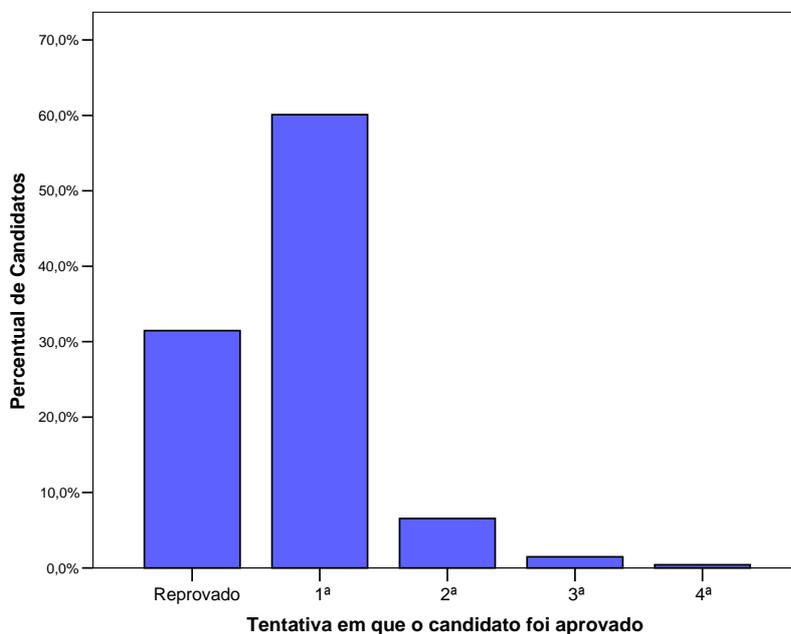
À medida que os candidatos vão respondendo aos exames, cai pela metade a chance de serem aprovados (o índice de aprovação dos que responderam pela primeira vez é de 82%, dos que responderam pela segunda vez é de 44%, pela terceira vez é de 20% e pela quarta vez é de 9%).

TABELA 2. DISTRIBUIÇÃO DE CANDIDATOS PELO NÚMERO DE EXAMES RESPONDIDO E POR SEU RESULTADO.

Respondendo o teste pela		Reprov.	Aprov. nesta tentativa	Total
1ª vez	Nº	1.028	4.673	5.701 (73,3%)
	%	18,0%	82,0%	100,0%
2ª vez	Nº	651	510	1.161 (14,9%)
	%	56,1%	43,9%	100,0%
3ª vez	Nº	448	115	563 (7,2%)
	%	79,6%	20,4%	100,0%
4ª vez	Nº	320	32	352 (4,5%)
	%	90,9%	9,1%	100,0%
Total	Nº	2.447	5.330	7.777 (100,0%)
	%	31,5%	68,5%	100,0%

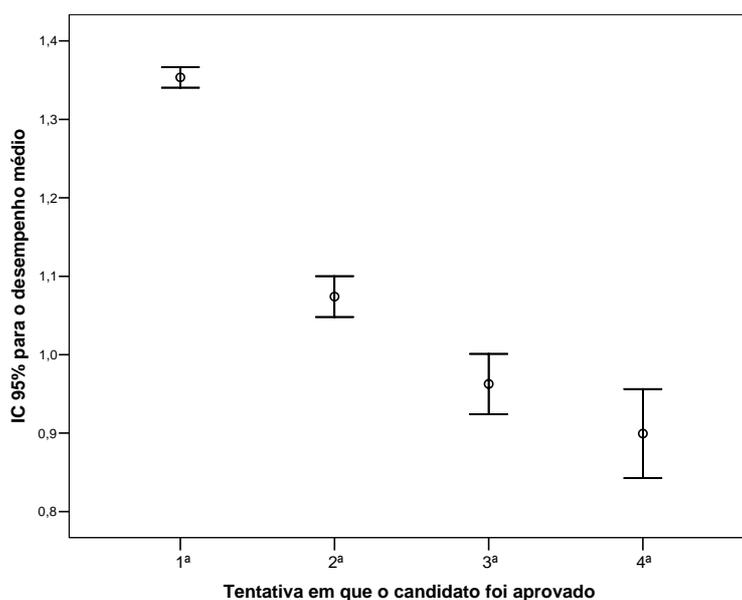
A maioria dos candidatos consegue êxito no TCE logo na primeira tentativa (60,1% – Figura 1).

FIGURA 1. PERCENTUAL DE CANDIDATOS APROVADOS NAS DIFERENTES TENTATIVAS



O procedimento ANOVA com comparações múltiplas foi realizado com o objetivo de verificar se existe diferença entre os desempenhos dos candidatos que foram aprovados nas diversas tentativas. Como resultado (figura 2), tem-se que os candidatos aprovados na primeira tentativa apresentam, em média, um desempenho maior que os candidatos aprovados a partir da segunda.

FIGURA 2. DESEMPENHO MÉDIO E SEU INTERVALO DE CONFIANÇA DE 95% DOS CANDIDATOS APROVADOS NAS DIFERENTES TENTATIVAS



4.2. Teste de Questões Práticas – QP

A terceira etapa, QP, é um teste formado pelos seguintes elementos: um cenário hipotético; alguns documentos, que são preparados com o mesmo padrão daqueles utilizados pelos candidatos em suas escolas; e seis questões discursivas. Cada questão é avaliada em uma escala de 0 a 3 pontos, portanto o desempenho do candidato no teste pode variar de 0 a 18 pontos. O candidato é considerado aprovado quando obtém, no mínimo, uma pontuação média de 2 pontos no conjunto das questões que compõem o teste.

Cada versão do QP é construída mediante uma mesma especificação do teste em que se detalham os conteúdos a serem cobrados. Com as questões prontas, comitês de especialistas adaptam, a cada versão, documentos de expectativas de respostas, denominadas rubricas, definindo quais serão os componentes de resposta, correspondentes a cada nota (Arter e McTighe, 2001). A comparabilidade das notas de diferentes versões é possível, considerando que o procedimento de criação e os parâmetros genéricos de pontuação das questões são os mesmos, além do nível de severidade das avaliações que é mantido a partir dos julgamentos dos especialistas. Estes lêem as respostas dos candidatos e atribuem uma nota de acordo com as rubricas.

Antes do início das pontuações, os especialistas passam por um treinamento e, se forem qualificados, continuam no processo. Durante toda a pontuação, é feito um controle de qualidade para verificar se as notas atribuídas estão de acordo com as rubricas. Além disso, as questões são pontuadas, de forma independente, por dois especialistas, cada um deles atribuindo a sua nota. A nota final da questão é a média entre estas duas notas, excetuando-se os casos em que um dos especialistas atribui uma nota que aprova o candidato e o outro, uma nota que não o aprova. Neste caso, é formado um comitê composto pelos dois primeiros especialistas e um terceiro, a fim de definirem, em consenso, a nota final do candidato (Agência de Certificação Ocupacional, 2006).

Analisando o QP (tabela 3), observou-se um resultado inverso ao do TCE, isto é, os exames que apresentaram os piores desempenhos dos candidatos (1º e 6º) foram aqueles que tiveram o menor percentual de candidatos repetentes.

TABELA 3. ANÁLISE DO DESEMPENHO DOS CANDIDATOS NO QP.

Exame	Geral			Respondendo pela 1ª vez (P)		Repetentes (R)		Intervalo de Confiança de 95% para a diferença das médias: P - R	
	Nº	Média	Desvio Padrão	Nº	Média	Nº	Média	LI	LS
1º	1.312	7,0	2,07	1.312 (100%)	7,0	0 (0%)	--	--	--
2º	2.845	9,4	2,30	2.120 (74,5%)	9,3	725 (25,5%)	9,6	-0,43	-0,04
3º	1.907	9,8	2,47	381 (20,0%)	10,0	1.526 (80,0%)	9,8	-0,10	0,45
4º	1.359	9,0	2,76	267 (19,6%)	9,4	1.092 (80,4%)	9,0	0,08	0,81
5º	820	8,9	2,46	213 (26,0%)	9,1	607 (74,0%)	8,8	-0,08	0,68
6º	375	7,4	2,65	291 (77,6%)	7,5	84 (22,4%)	7,0	-0,21	1,08
7º	551	9,6	2,50	384 (69,7%)	9,4	167 (30,3%)	10,1	-1,08	-0,17

Apenas no 2º e no 7º exames, o desempenho médio dos candidatos que estavam respondendo pela primeira vez diferiu, significativamente, ao nível de 5%, do desempenho médio dos que estavam repetindo. Essa diferença foi negativa, ou seja, os candidatos repetentes apresentaram melhor desempenho. Um resultado também significativo, mas inverso, foi observado no 4º exame, em que essa diferença foi positiva, mostrando que nesse exame os candidatos repetentes apresentaram pior desempenho.

No QP, a grande parte dos candidatos só respondeu a esse teste apenas uma vez (47,9% - tabela 4), com a maioria destes já sendo aprovados nessa versão (63,6%).

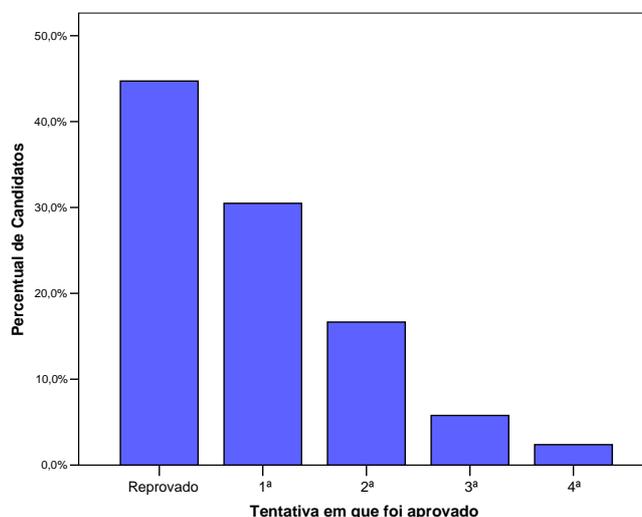
A chance dos candidatos serem aprovados, à medida que vão respondendo aos exames, diminui cerca de 12% de uma versão para outra (o índice de aprovação dos que respondem pela primeira vez é de 64%, dos que respondem pela segunda vez é de 58%, pela terceira vez é de 39% e pela quarta vez é de 29%).

TABELA 4. DISTRIBUIÇÃO DE CANDIDATOS PELO NÚMERO DE EXAMES RESPONDIDOS E POR SEU RESULTADO NO QP.

Respondendo o teste pela		Reprov.	Aprov. nesta tentativa	Total
1ª vez	Nº	868	1.514	2.382 (47,9%)
	%	36,4%	63,6%	100,0%
2ª vez	Nº	612	827	1.439 (29,0%)
	%	42,5%	57,5%	100,0%
3ª vez	Nº	450	287	737 (14,8%)
	%	61,1%	38,9%	100,0%
4ª vez	Nº	292	118	410 (8,3%)
	%	71,2%	28,8%	100,0%
Total	Nº	2.222	2.746	4.968 (100,0%)
	%	44,7%	55,3%	100,0%

A maior parte dos candidatos que responderam a pelo menos uma versão do QP ainda não foi aprovada (44,7%). A versão que mais aprova ainda é a primeira (30,5% – Figura 3).

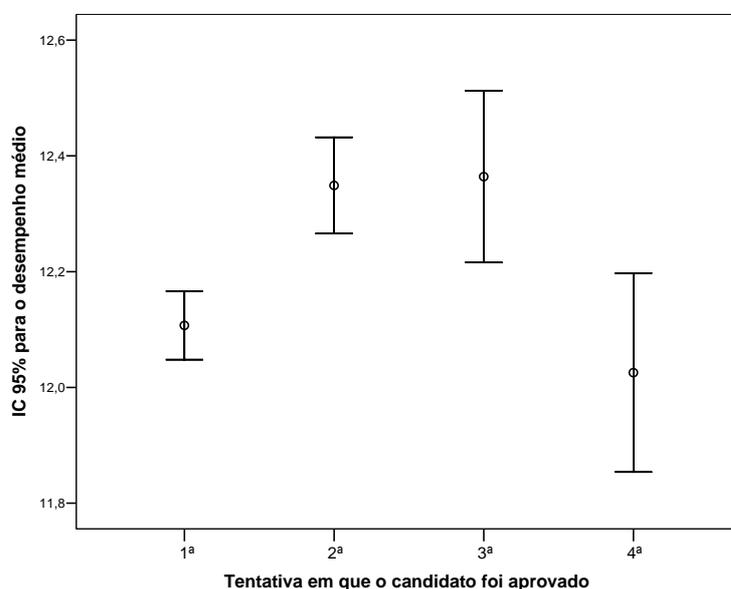
FIGURA 3. PERCENTUAL DE CANDIDATOS APROVADOS NAS DIFERENTES TENTATIVAS DO QP



Mediante procedimiento ANOVA, foi verificado que o desempenho dos candidatos aprovados na 2ª. tentativa não diferiu significativamente do desempenho dos aprovados na 3ª. vez (figura 4). Assim como o desempenho médio dos candidatos aprovados na 1ª. vez não diferiu significativamente dos aprovados na 4ª. vez.

Apesar de o percentual de aprovação ter sido maior na primeira tentativa, esses candidatos evidenciaram um desempenho médio significativamente menor que os candidatos que foram aprovados na 2ª. ou 3ª. tentativas.

FIGURA 4. DESEMPENHO MÉDIO E SEU INTERVALO DE CONFIANÇA DE 95% DOS CANDIDATOS APROVADOS NAS DIFERENTES TENTATIVAS DO QP.



4.3. Avaliação da Tendência do Desempenho no TCE e QP

As figuras 5 e 6 mostram os desempenhos de cinco candidatos que responderam a 4 versões dos dois testes, TCE e QP, respectivamente. Por meio dessas figuras, percebeu-se que alguns candidatos melhoraram e outros pioraram seu desempenho.

FIGURA 5. DESEMPENHO DE ALGUNS CANDIDATOS NO TCE

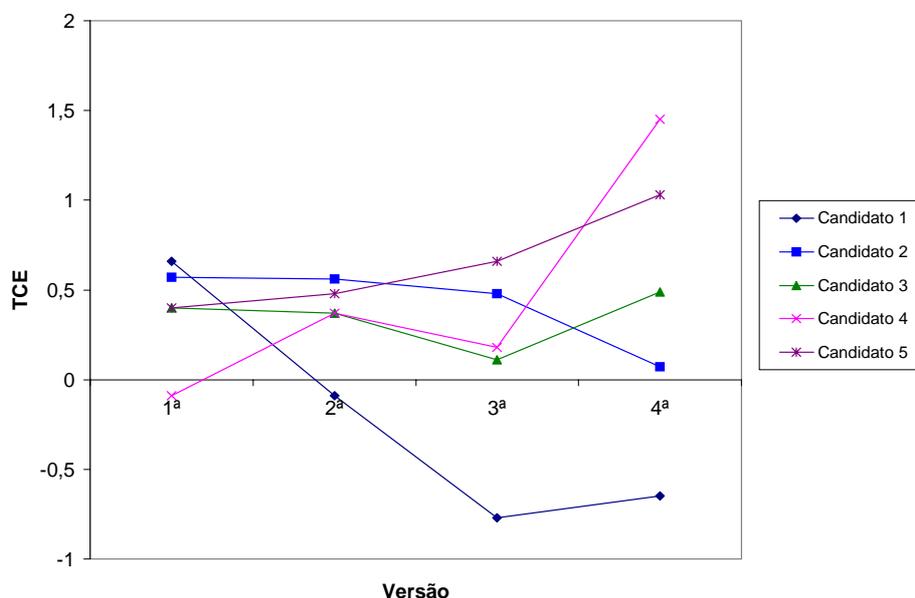
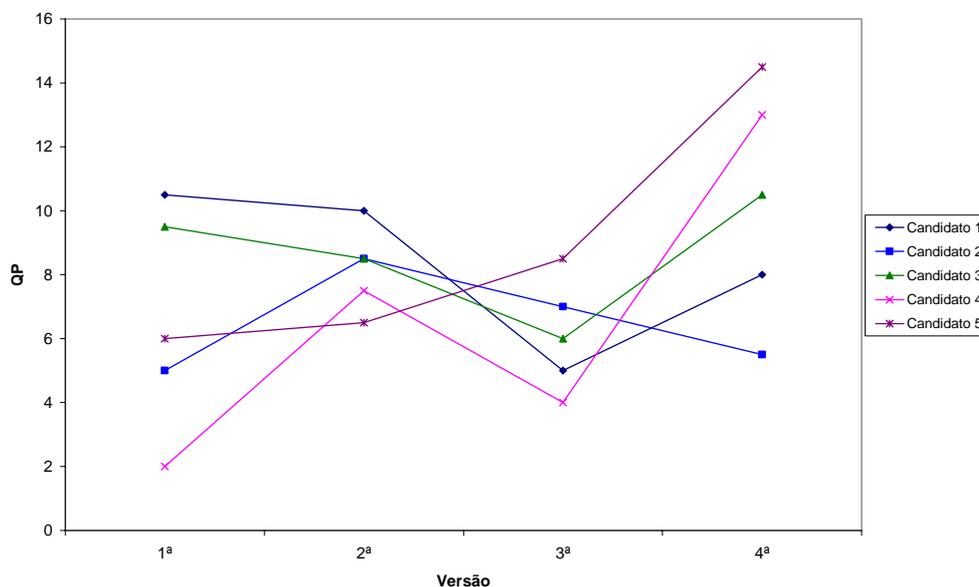


FIGURA 6. DESEMPENHO DE ALGUNS CANDIDATOS NO QP



Para avaliar a tendência do desempenho dos candidatos ao longo das versões, foi utilizado o método de tendência linear (Moretin y Toloí, 2004), em que

$$Z_t = a_0 + a_1t + \epsilon_t, t = 1, \dots, N$$

Z_t é o resultado do candidato na versão t do TCE ou QP e a_1 é a tendência do desempenho dos candidatos ao longo dos exames. Os parâmetros desse modelo foram estimados, segundo o método dos

mínimos cuadrados (Box y Jenkins, 1976).

As figuras 7 e 8 e tabela 5 trazem o histograma e a análise descritiva de a_1 . Observou-se que existe uma tendência positiva no desempenho dos candidatos reprovados, isto é, à medida que os candidatos respondem às versões do teste, vão adquirindo mais conhecimentos e habilidades. No TCE, em média, o desempenho dos candidatos aumenta 0,143 pontos de uma versão para a outra. Já no QP, essa média é de 1,431 pontos.

FIGURA 7. HISTOGRAMA DO a_1 DO TCE.

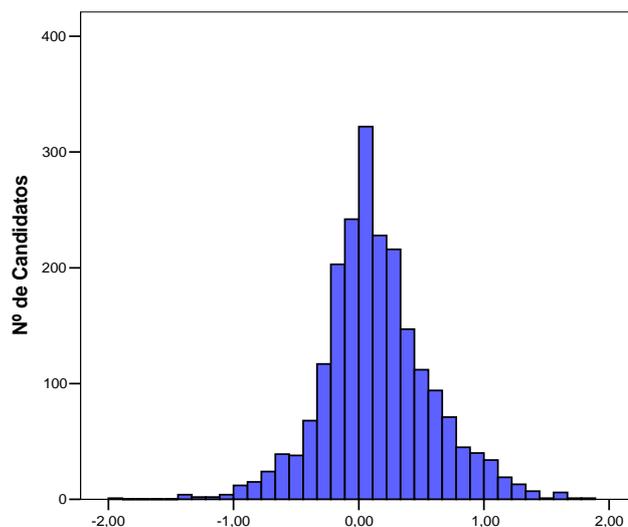
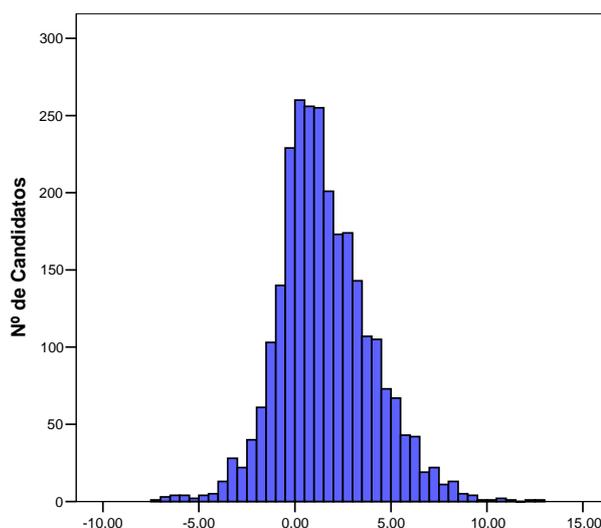


FIGURA 8. HISTOGRAMA DO a_1 DO QP.



Essa seção apresenta resultados de estudos feitos com o objetivo de verificar a relação existente entre a tendência do desempenho dos candidatos com algumas variáveis demográficas.

Os dirigentes foram o público alvo do primeiro exame os quais haviam sido aprovados em um curso dado anteriormente pela SEC. A partir do segundo exame, o público se estendeu para outros profissionais de educação. Além dessa diferença clara dos inscritos no primeiro exame, espera-se uma distinção entre os perfis dos candidatos nos diferentes exames, visto que aqueles que demoraram a se inscrever poderiam se sentir não tendo as competências necessárias para o cargo e estariam se preparando melhor para a certificação. Dessa forma, modelando os parâmetros, segundo uma estrutura hierárquica, combinando essas diversas informações, é possível obter significativo aumento na precisão das estimativas individuais de cada exame.

Utilizando o *software* WinBUGs (Spiegelhalter *et al.*, 1998), foi desenvolvido o seguinte modelo hierárquico bivariado, com o primeiro nível definido como sendo o candidato e, o segundo nível, sendo o exame para o qual o candidato se inscreveu:

$$Y_i \sim \text{Normal}_2(\mu_i, P)$$

onde

$Y_i = (Y_{i1}, Y_{i2})$; Y_{ij} = tendência do desempenho (a_i) do candidato i no teste j ;

$i = 1, \dots, N$ (número de candidatos); $j=1$ para TCE e $j=2$ para QP;

P = matriz de dispersão 2×2 para as observações dos candidatos;

$\mu_i = (\mu_{i1}, \mu_{i2})$; μ_{ij} = média do candidato i no teste j escrita como

$$\mu_{ij} = \alpha_{l(i)j} + \beta_{1j} * \text{SEXO}_i + \beta_{2j} * \text{DIRIG}_i + \beta_{3j} * \text{PROF}_i + \beta_{4j} * \text{IDADE}_i + \beta_{5j} * \text{ESC}_i$$

com $l(i)$ = o exame em que o candidato i se inscreveu ($l = 1, \dots, 6$);

$\text{SEXO}_i = 1$ se o sexo do candidato i for masculino e 0 se for feminino;

CARGO_i = cargo do candidato i . Essa variável foi transformada em:

$\text{DIRIG}_i = 1$ se o candidato i for diretor ou vice-diretor e 0, caso contrário e

$\text{PROF}_i = 1$ se o candidato i for professor e 0, caso contrário;

IDADE_i = idade do candidato i ;

ESC_i = grau de escolaridade. Assume valor 1, se o candidato possuir até o nível médio completo e 0, se possuir pelo menos o nível superior completo.

Assim, foi assumida uma variação entre o desempenho dos candidatos sobre os 6 diferentes exames, mediante o efeito aleatório α_{ij} , como descrito abaixo:

$$\alpha_{ij} \sim \text{Normal}_2(\eta_j, Q)$$

onde

η_j = média geral do teste j ;

Q = matriz de dispersão 2×2 para os efeitos dos exames.

Foi utilizada uma distribuição a priori não-informativa para os parâmetros (GELMAN *et al.*, 1995). Mediante a figura 9, observou-se que os candidatos do segundo exame tiveram o menor crescimento de desempenho entre as versões no TCE. Já a figura 10 apresenta que os candidatos inscritos no 4º e 5º exames tiveram menores α 's no QP, enquanto os inscritos do 6º exame tiveram uma maior tendência média de desempenho.

FIGURA 9. MÉDIA A POSTERIORI E INTERVALO DE CREDIBILIDADE DE 95% PARA A NO TCE

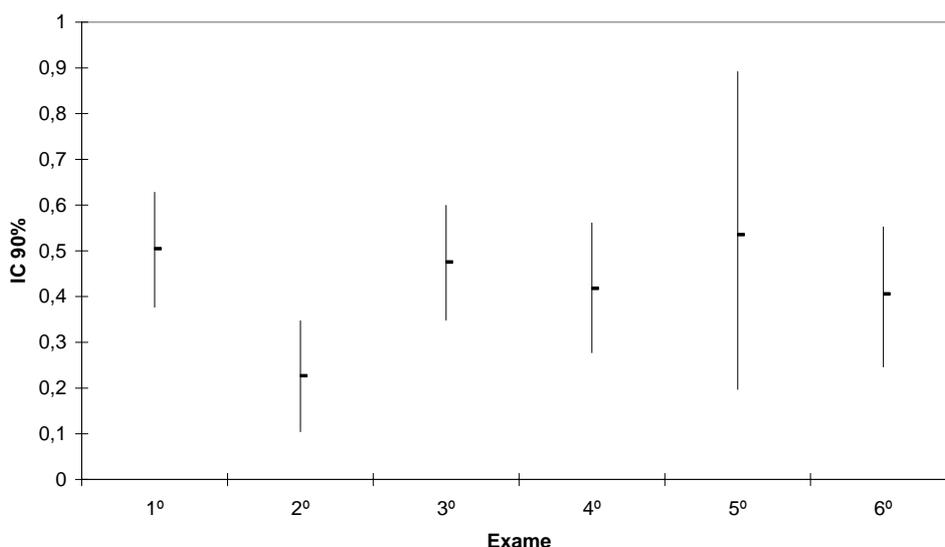
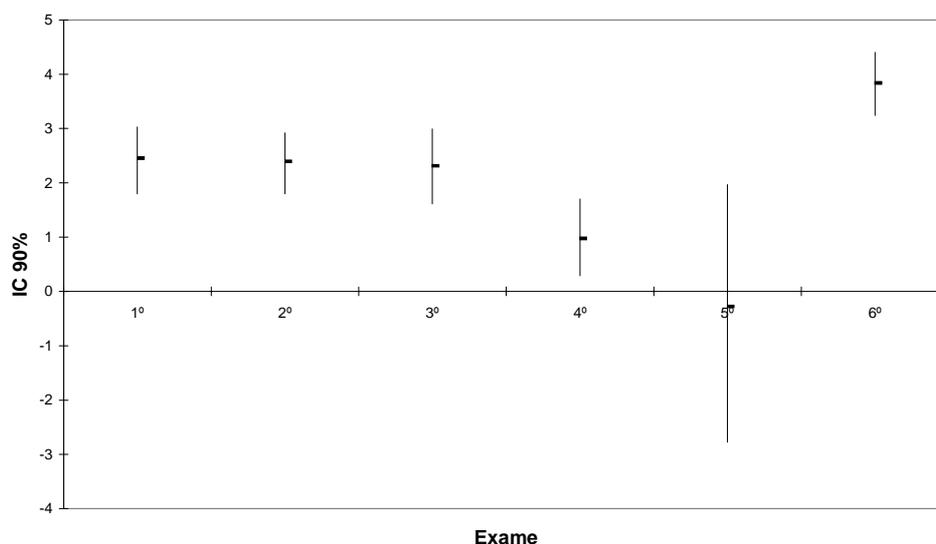


FIGURA 10. MÉDIA A POSTERIORI E INTERVALO DE CREDIBILIDADE DE 95% PARA A NO QP



O resultado para os demais parâmetros pode ser visto na tabela 6. Os candidatos do sexo masculino apresentaram um menor crescimento de conhecimentos e habilidades que os do sexo feminino no QP. Os professores apresentaram uma maior tendência do desempenho nos testes que os candidatos de outros cargos. Já para a variável idade, percebeu-se que os candidatos mais jovens

tendem a ter um melhor crescimento no teste. Os candidatos que têm nível superior ou mais também tendem a ter um melhor crescimento no TCE que os candidatos com no máximo o ensino médio completo.

A correlação entre as tendências dos candidatos nos dois testes ($\rho_1 = P_{12} / \sqrt{P_{11}P_{22}}$) foi estimada em $-0,1222$, isto é, à medida que o candidato tende a melhorar em um teste, ele piora em outro.

A correlação entre os efeitos dos exames no TCE com o QP não foi significativa ($\rho_2 = Q_{12} / \sqrt{Q_{11}Q_{22}}$).

TABELA 6. ESTATÍSTICAS DA DISTRIBUIÇÃO A POSTERIORI DOS PARÂMETROS.

Parâmetro	Média	Desvio Padrão	2,50%	Mediana	97,50%
β_{11}	-0,0256	0,0305	-0,0854	-0,0264	0,0338
β_{12}	-0,5605	0,1287	-0,8023	-0,5607	-0,3020
β_{21}	-0,0228	0,0305	-0,0818	-0,0222	0,0371
β_{22}	-0,0260	0,1273	-0,2692	-0,0259	0,2176
β_{31}	0,3911	0,1368	0,1302	0,3888	0,6535
β_{32}	1,3290	0,3564	0,6320	1,3320	2,0270
β_{41}	-0,0039	0,0010	-0,0059	-0,0039	-0,0017
β_{42}	-0,0215	0,0053	-0,0304	-0,0221	-0,0097
β_{51}	0,0557	0,0199	0,0169	0,0557	0,0946
β_{52}	0,1777	0,1105	-0,0370	0,1778	0,3956
P_{11}	0,1676	0,0054	0,1574	0,1676	0,1782
P_{12}	-0,1162	0,0381	-0,1907	-0,1166	-0,0429
P_{22}	5,4010	0,1558	5,1050	5,3960	5,7090
ρ_1	-0,1222	0,0398	-0,2011	-0,1225	-0,0452
η_1	0,4255	0,1098	0,2147	0,4247	0,6408
η_2	1,9330	0,7898	0,2250	1,9860	3,3680
Q_{11}	0,0475	0,0452	0,0116	0,0351	0,1559
Q_{12}	-0,0808	0,2668	-0,6395	-0,0471	0,3333
Q_{22}	2,8850	3,1720	0,4733	1,9610	11,4000
ρ_2	-0,1832	0,4292	-0,8578	-0,2268	0,6904

5. Conclusões

Sob um cenário educacional de baixo desempenho escolar, distorção idade/série e evasão escolar, foi implementado o Projeto de Certificação Ocupacional, que se iniciou com o exame para Dirigentes Escolares.

Esse projeto trouxe o benefício de estabelecer conhecimentos e habilidades necessários para o exercício da função do gestor escolar, deixando claro o padrão do dirigente requerido para as novas demandas da gestão participativa e autônoma da escola. A partir daí, foram desenvolvidos vários cursos e seminários para, principalmente, professores que assumem a função de diretor sem ter oportunidades de desenvolver as competências requeridas por esse cargo, como cita Reis (2003):

“Após essa definição (dos padrões de Dirigente Escolar), realizaram-se cursos presenciais com carga horária de 20 a 96 horas, objetivando subsidiar os dirigentes para liderar a equipe escolar, planejar, acompanhar, avaliar os resultados, implementar o projeto político-pedagógico e solucionar os conflitos a que as escolas estão sujeitas. Realizaram-se, também, encontros mensais para aprofundamento de temas específicos e seminários regionais com o propósito de divulgar experiências exitosas na gestão escolar.”

Além disso, a certificação utiliza critérios de mérito na escolha dos profissionais, ao invés de outro tipo de seleção, como eleição, que frequentemente leva à escolha de profissionais pela simpatia ou pela desinibição, e não pela competência, como afirma o professor Clemenceau Chiabi Saliba em depoimento a Nossa Gente (2005). O professor, que é um dos maiores especialistas brasileiros em educação e que preside a Câmara de Certificação, também declara que os resultados positivos na educação pública do Estado, observados por pesquisa do Ministério da Educação, estão diretamente relacionados com a Certificação Ocupacional.

A Certificação Ocupacional também se mostra importante na vida dos profissionais aprovados no exame que passa a ter um crescimento profissional, ganho salarial e o reconhecimento público de suas competências. Ainda em Nossa Gente (2005), a diretora Antônia Cazumbá deu o seguinte depoimento:

“A certificação é uma seleção criteriosa. Por isso o profissional fica mais valorizado e a comunidade passa a nos ver com outros olhos, como um profissional capacitado para a função. Logo, somos também mais cobrados. Os pais, os alunos, toda a equipe que aqui trabalha tem uma relação de confiança com a direção da escola. E eu me sinto mais segura, confiante e valorizada para estar à frente.”

Pensando na importância da Certificação Ocupacional e no esforço dos profissionais em serem aprovados para obterem o seu benefício, esse artigo traz um estudo de como os candidatos reprovados se comportam em versões subsequentes do TCE – Teste de Conhecimentos Específicos e do QP – Teste de Questões Práticas. Esses dois testes avaliam os seis padrões estabelecidos para o cargo de Dirigente Escolar.

Pelos resultados das análises apresentadas, é possível concluir que o TCE e o QP possuem perfis de repetentes diferentes:

- No TCE, com exceção do 2º exame, os candidatos novatos tiveram, em média, um melhor desempenho que os repetentes. Já no QP, não se verifica um padrão, pois, no 4º exame, os candidatos novatos tiveram melhores resultados, no 2º e 7º exames, os reprovados se saíram melhor e, já nos outros 4 exames restantes, não houve diferença significativa entre os desempenhos médios dos novatos e repetentes;
- Apenas 26,7% dos candidatos que responderam ao TCE repetiram pelo menos uma vez esse teste. Enquanto que, no QP, esse percentual foi de 52,1%;
- A maioria dos candidatos que respondeu ao TCE (60,1%) foi aprovada na primeira tentativa. Ao passo que, apenas 30,5% dos candidatos foram aprovados na primeira vez que responderam ao QP;
- Os candidatos que foram aprovados na primeira tentativa apresentaram um conhecimento maior no TCE que os demais candidatos aprovados na segunda tentativa em diante. Já no QP, os candidatos que foram aprovados na 2ª e 3ª tentativas apresentaram, em média,

melhores desempenhos que os demais candidatos aprovados em outras tentativas.

Em ambos os testes, os candidatos apresentaram, em média, um crescimento nos conhecimentos/habilidades ao responder às versões. Porém, apesar de existir essa melhora no desempenho, ela não é suficiente para que o candidato tenha boas chances de ser aprovado, visto que, no TCE, cai pela metade a probabilidade do candidato ser aprovado de uma versão para outra e, no QP, cai em cerca de 12%. Dessa forma, seria interessante que a Secretaria de Educação do Estado da Bahia oferecesse mais capacitação/cursos/seminários, de modo que esses reprovados pudessem ter base para se preparar melhor para o próximo teste.

Mediante o modelo hierárquico bivariado, pôde-se concluir que os candidatos inscritos no 2º exame tiveram o menor crescimento entre as versões no TCE; isso também aconteceu com os candidatos inscritos 4º e 5º exames, em relação ao QP. Já os candidatos do 6º exame apresentaram maior crescimento ao responder às diferentes versões do QP.

Ao analisar as características demográficas dos candidatos, percebeu-se que tanto os professores quanto os mais jovens apresentaram uma maior tendência em adquirir mais conhecimentos entre as versões de ambos os testes. Essa maior tendência também foi observada entre os candidatos que têm no mínimo o nível superior completo no TCE e entre os candidatos do sexo feminino no QP.

REFERENCIAS BIBLIOGRÁFICAS

- Hoxby, C.M. (2000). The effects of class size on student achievement: new evidence from population variation. *Quarterly Journal of Economics*, 115, pp.1239-1285.
- Andrade, D. F., Tavares, H. R. y Valle, R.C. (2000). Teoria da resposta ao item: conceitos e aplicações. 14º *SINAPE, ABE - Associação Brasileira de Estatística*, São Paulo.
- Arter, J. y McTigh E.J. (2001). *Scoring rubrics in the classroom: using performance criteria for assessing and improving student performance*. Thousand Oaks: Corwin Press.
- Box, G.E.P. y Jenkins, G.M. (1976). *Time Series Analyses: forecasting and control*. San Francisco: Holden-Day.
- Bryk, S.A. y Raudenbush, W. (1992). *Hierarchical Linear Models*. Califórnia: Newbury Park.
- Congdon, P. (2001). *Bayesian Statistical Modelling*. London: Wiley.
- Crocker, L. y Algina, J. (1986). *Introduction to Classical & Modern Test Theory*. Florida: Harcourt Brace Jovanovich-HBJ.
- D'almeida, M. C. (2003). *Certificação ocupacional de profissionais da educação na Bahia: análise do alinhamento entre os resultados do Teste de Conhecimentos Específicos e da auto-avaliação*. Salvador: UFBA.
- Ferreira, S.M.M. (2003). *O impacto do Programa "Educar para Vencer" no desempenho de dirigentes na escola pública baiana: estudo de caso sobre avanços e possibilidades*. Salvador: UFBA.
- Fundação Luís Eduardo Magalhães (2000). *Padrões de competências: dirigente escolar*. Disponível em: <http://www.flem.org.br>. Acesso em: 02 de fevereiro de 2007.
- Gamerman, D. y Lopes, H.F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. 2ª edição. Londres: Chapman and Hall/CRC.
- Goldstein, H. (1995). *Multilevel Statistical Models*. New York: John Wiley and Sons.
- Goldstein, H. *Multilevel Models*. London: Arnold.

- Goldstein H. y Rasbash J. (1992). Efficient computational procedures for the estimation of parameters in multilevel models based on iterative generalised least squares. *Computational Statistics and Data Analysis*, 13(1), p. 63-71.
- Grupo de Avaliação E Medidas Educacionais (2002). *Escola Eficaz: um estudo de caso em três escolas da rede pública de ensino do Estado de Minas Gerais*. Belo Horizonte: UFMG.
- Gelman, A. et al. (1995). *Bayesian Data Analysis*. New York: Chapman and Hall.
- Hambleton, R.K., Swaminathan, H. y Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. North Caroline: Sage Publications.
- Hashimoto, R. (2003). Certificação ocupacional em educação: garantia de educação continuada e ascensão salarial. *VIII Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública. Panamá, 2003*.
- Hewson, P. (2005). Multilevel Modelling for Environmental Applications. *S.I.S. 2005 – Estadística y Ambiente. Italy*.
- Kolen, M.J. y Brennan, R. L. (1995). *Teste Equating: methods and practices*. New York: Springer.
- Kreft, I. y De Leeuw, J. (1998). *Introducing Multilevel Modeling*, London: Sage.
- Lück, H. (2000). Indicadores para a Qualidade na Gestão Escolar e Ensino. *Gestão em Rede*, 25, pp. 15-18.
- Marras, J.P. (2002). *Administração de Recursos Humanos: do operacional ao estratégico*. São Paulo: Futura.
- Migon, H.S. y Gamerman, D. (1999). *Statistical Inference: an Integrated Approach*. Londres: Arnold.
- Morettin, P.A. y Toloi, C.M.C. (2004). *Análise de series temporais*. São Paulo: Edgard Blücher.
- R Development Core Team (2004). *R: A language and environment for statistical computing*. Austria: R Foundation for Statistical Computing.
- Rabash, J., Prosser, R. y Goldstein, H. (1991). *ML3. Software for Three-Level Analysis. User's Guide for V-2*. London: University of London.
- Reis, N.M.V. (2003). Projeto fortalecimento da gestão escolar: reflexões sobre os desafios e possibilidades vivenciados na construção progressiva da gestão democrática e da autonomia escolar a partir de uma experiência baiana. *VIII Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública. Panamá, 2003*.
- Sammons, P., Hillman, J. y Mortimore, P. (1995). *Key Characteristics of Effective Schools: A Review of School Effectiveness Research*. London: Office for Standards in Education [OFSTED].
- Spiegelhalter, D., Thomas, A. y Best, N. (1998). *WinBUGS version 1.1.1 user manual. Technical report*. Cambridge: MRC Biostatistics Unit.
- SPSS (1995). *SPSS: Statistical Package for the Social Sciences. Versão 7*. New York: McGraw-Hill Book.