

REPRESENTAÇÕES GRÁFICAS NOS ITENS OBJETIVOS DA PROVA DE GEOGRAFIA NO CONCURSO VESTIBULAR DA UFMG: ELEMENTO DISCRIMINATÓRIO DO DESEMPENHO DOS VESTIBULANDOS

Bárbara Lúcia Pinheiro de Oliveira França¹
Márcia Maria Duarte dos Santos²
Alexandre Magno Alves Diniz³

Nos últimos trinta anos, houve um crescimento expressivo das tentativas de aprimoramento do sistema seletivo universitário. Mudanças significativas têm sido realizadas como estratégias para atenuar as contradições existentes no referido sistema. Muitas dessas mudanças são norteadas pelos resultados de estudos sobre as características sociais, culturais e econômicas dos candidatos aos concursos vestibulares. Outras modificações são possíveis a partir de análises do instrumento de avaliação – as provas. Nesse caso, as análises objetivam o aprimoramento desse importante instrumento de medida.

Neste trabalho, propõe-se abordar alguns aspectos concernentes ao uso de provas objetivas nos concursos vestibulares e a caracterização destas como testes padronizados de desempenho. Propõe-se também apresentar a teoria psicométrica de análise de resultados de testes padronizados. A par disso, as provas dos concursos vestibulares da UFMG serão apresentadas como testes padronizados e discutidas a partir das medidas de desempenho. A seguir, será relatado um estudo de caso referente ao desempenho dos examinandos nas provas de Geografia da primeira etapa dos concursos vestibulares, entre os anos de 1997 e 2001.

1. O uso de provas objetivas no vestibular

A introdução de provas com itens objetivos⁴ no vestibular ocorreu no final da década de 1960. Essa inovação adveio do aumento de candidatos aos concursos, o que tornou, de um lado, necessária a utilização de provas que pudessem ser corrigidas em menor tempo, através do emprego de leitoras óticas ou de cartões, e, de outro lado, possível a eliminação da subjetividade no decorrer das correções.

¹ Mestre, Curso de Pós-Graduação em Geografia, IGC/UFMG, bbarbaralucia@yahoo.com.br

² Profa. Adjunta, Dra, Departamento de Geografia, IGC/UFMG, mdsantos@ufmg.br

³ Prof., MS, PUC, MG

⁴ O termo "objetivo" está relacionado ao processo de cálculo de escores, e não propriamente à maneira como é dada a resposta. Os itens objetivos são arquitetados com a finalidade de computar os escores observando uma única palavra ou frase, ou, ainda, notando qual de várias respostas possíveis foi escolhida. (LINDEMAN, 1972, p.65)

No início de sua difusão, as provas objetivas – uma novidade no setor de verificação, porém difíceis de elaborar – não cumpriram bem a sua função de instrumento de seleção. No entanto, como afirma Silva (1987, p.15), “[...] o problema não se refere a sua natureza, mas sim a sua elaboração”. Por isso, o uso freqüente e cada vez mais extensivo dessas provas possibilitou seu aprimoramento e sua aceitação como um instrumento adequado de medida de conhecimento.

Lindeman (1972, p.77), favorável ao uso das provas objetivas, argumenta que elas “[...] são excelentes para medir a compreensão e testar a capacidade de aplicar conceitos e conhecimentos a situações especiais”; porém, ele adverte que, assim como qualquer outro método de aferição, as provas objetivas possuem limitações. Entre elas, destaca a impossibilidade de avaliar algumas capacidades da inteligência humana, como, por exemplo, criatividade, habilidades de síntese, expressão oral ou verbal.

Cabe ressaltar que a principal limitação das provas objetivas refere-se à sua construção, pois, se não forem bem preparadas, sua validade será comprometida. Nesse sentido, desde que sejam bem construídas, as principais vantagens das provas objetivas são, segundo França (2000, p.7), apoiando-se em Medeiros (1989)⁵: o julgamento impessoal e imparcial; a correção rápida das provas; a abordagem de conteúdos variados; a identificação de limitações individuais dos estudantes; a possibilidade de comparação de resultados e a avaliação do trabalho docente. Além das vantagens citadas, as provas objetivas permitem o uso de formas variadas e flexíveis de questões, a verificação de capacidades de análise e comparação e a redução da probabilidade de acerto por sorte.

Segundo Di Dio (1975, p. 103-110), citado por Dias (1977, p.13-15), os defensores da aplicação de testes objetivos nos concursos vestibulares invocam, em síntese, os seguintes argumentos:

- 1- Os testes ainda são a maneira mais prática, simples e menos dispendiosa para classificar candidatos em função de seu desempenho.
- 2- Uma educação de massa só comporta os testes como meio de avaliação, tanto mais que a tecnologia moderna facilita a correção das provas.
- 3- Os testes classificam realmente os alunos de acordo com sua capacidade, desempenho e suas aptidões e permitem abranger uma área de conhecimento muito mais ampla na formulação das

⁵ Medeiros, Ethel. *Provas objetivas, discursivas, orais e práticas: técnicas de construção*- 9 ed. Ampliada. Rio de Janeiro: Editora Fundação Getúlio Vargas, 1989XI,178p.

perguntas, que se constituem numa amostra representativa do universo de noções cujo domínio se quer aferir.

4- Os testes eliminam o fator sorte, pois não beneficiam o aluno que estudou um só ponto nem prejudicam quem se preparou em todas as áreas menos uma.

5- Se os testes objetivos não tivessem outra vantagem, têm a indiscutível superioridade de serem impessoais na atribuição dos escores, evitando que, consciente ou inconscientemente, a avaliação atenda a simpatias ou interesses.

6- As críticas que se fazem aos testes seriam válidas em relação às provas objetivas mal construídas, não atingindo os itens formulados por pessoas competentes na matéria e na tecnologia dos testes.

7- Quem combate os testes toma uma atitude negativa de crítica destrutiva, pois não oferece uma solução mais eficiente e mais satisfatória.

8- Pesquisas feitas há muitos anos têm demonstrado que os testes são melhores preditores de desempenho nos cursos superiores na maioria das matérias.

9- Não são os testes os responsáveis pelo rebaixamento do nível de ensino superior, mas sim a massificação e o critério classificatório, pelo qual são admitidos na Universidade os que obtiverem as melhores notas, desde que haja vagas e a nota do candidato seja diferente de zero.

Ressalta-se, por fim, que o uso de testes objetivos permite, ainda, assim como outros tipos de prova, a aplicação de tratamento estatístico para analisar o desempenho dos vestibulandos. A teoria psicometrista aponta que mais importante do que aplicar um teste é verificar a sua legitimidade através daquele tratamento, com vistas à interpretação das especificidades da capacidade humana. Nesse contexto, educadores vêm utilizando-se de procedimentos denominados medidas educacionais de desempenho.

2. As medidas educacionais de desempenho e os testes padronizados

Acredita-se que o sistema avaliativo do vestibular deve ser aperfeiçoado. Para tal propósito é preciso conhecer o desempenho dos vestibulandos através de medidas

educacionais e, dessa forma, realizar uma adequação entre o aparelho de avaliação e o público-alvo.

Segundo Silva (1987, p.77), ao se fazer uso de provas objetivas em concursos vestibulares, deve-se cuidar não só de sua formulação, mas também de sua análise após a aplicação. O referido autor alega que a análise deve ser feita pelo elaborador com a assessoria de especialistas em medidas educacionais para que o tipo de prova seja aprimorado. E, enfaticamente, observa que, “[...] de fato, se a questão da prova não pode ser testada antes de sua aplicação, deve ser analisada depois de aplicada, com a finalidade de aprimorar o sistema” (SILVA, 1987, p.77).

Como assegura Lindeman (1972, p.5), “[...] a medida em educação é essencial para que o processo avaliativo seja levado a termo com precisão e eficácia”. Desse modo, para o emprego das medidas educacionais como meio apropriado para avaliar o desempenho dos candidatos é necessário valer-se de observação, escalas classificatórias ou instrumentos que permitam obter informações quantitativas (MEHRENS, 1978, p.3).

É consenso entre os especialistas em medidas educacionais o fato de que só é possível utilizar as medidas educacionais como meio de avaliar e comparar desempenhos em uma prova se esta for um instrumento de mensuração confiável. E esses especialistas nomeiam, via de regra, como instrumentos confiáveis, os testes padronizados.

Os testes em questão, nas suas diversas modalidades, para serem utilizados conforme os fins descritos anteriormente, devem ser construídos e aplicados com bastante rigor, prescrito metodologicamente. Além disso, devem ser escolhidos cuidadosamente, considerando-se a consecução de objetivos específicos.

2.1 Modalidades de testes

Mehrens (1978, p.3) define os testes padronizados como sendo aqueles preparados por especialistas em medidas, e complementa a definição observando que eles são assim denominados porque fornecem métodos para obtenção de amostras comportamentais em condições uniformes, ou seja, a partir de um conjunto fixo de questões, instruções, limitações de tempo e atribuição de escores.

O autor relata que os referidos testes têm como princípio geral a aplicação deles a um grupo normalizado, possibilitando, assim, que o desempenho de um indivíduo possa ser interpretado por meio da comparação com outros. Outro princípio, não menos importante, lembrado pelo autor está relacionado à apresentação de regras claras e objetivas a propósito da aplicação e correção.

Existem inúmeros testes padronizados; porém o que atende aos interesses deste estudo são os denominados Testes Padronizados de Desempenho ou Aproveitamento. A função específica do Teste Padronizado de Aproveitamento (TPA) é auxiliar a tomada de decisões instrucionais, de orientação, administrativas e de pesquisa.

O primeiro Teste Padronizado de Desempenho foi aplicado por Rice em 1895, nos Estados Unidos. Tratava-se de um teste de ortografia realizado por mais de dezesseis mil alunos de 4^a a 8^a série, que se destacou por avaliar de maneira objetiva e científica o conhecimento do aluno. A partir de então, surgiram nos EUA vários outros tipos de testes padronizados de aproveitamento⁶, construídos em diferentes áreas de conteúdo. Esses testes passaram a ser aplicados a uma população relativamente homogênea para que se pudesse interpretar o progresso relativo dos examinandos em áreas disciplinares distintas (LINDEMAN, 1972, p.106).

Cabe destacar que, no TPA, o aproveitamento do estudante é aferido de acordo com prescrições de aplicação, o que o torna diferente dos testes construídos pelo professor. Além disso, os referidos testes diferenciam-se por amostragem de conteúdo, construção, normas, objetivos e uso (MEHRENS, 1978, p.191).

O quadro de comparação apresentado a seguir, elaborado por Mehrens (1978, p.191), demonstra as principais diferenças entre testes padronizados de aproveitamento e os testes de aproveitamento construídos pelo professor (QUADRO 1).

A partir da análise do referido quadro constata-se que existem diferenças significativas entre os dois tipos de testes. Os testes padronizados demonstram mais confiabilidade quando comparados aos testes construídos pelo professor por se apresentarem de modo impessoal. Além disso, o seu uso exige a prescrição de critérios como, por exemplo, a análise e revisão de itens. Todas essas características tornam o Teste Padronizado apto a medir com bastante precisão diversos assuntos, específicos ou não, de maneira sistemática.

O fato de a construção dos testes padronizados envolver especialistas possibilita uma discussão mais apurada sobre sua elaboração, o que não acontece no teste construído pelo professor. Dessa forma, evita-se que interesses pessoais do professor possam interferir na construção do teste e nos resultados do desempenho dos alunos.

Outra importante diferença refere-se aos objetivos e uso dos testes. Geralmente, nos testes construídos pelo professor, os objetivos curriculares medidos são

⁶ As baterias de testes padronizados de aproveitamento mais usadas logo no surgimento dos testes padronizados foram as seguintes: Califórnia Achievement Tests, Cooperative General Achievement Tests, Estential High School Content Battery, Iowa Tests of Educational Development, Sequential Tests of Educational Progress. (LINDEMAN, 1972, p.108)

menos amplos. Quando o professor constrói um teste ele não se preocupa em avaliar uma grande gama de conteúdos e disciplinas. Nesse caso, a quantidade máxima de conteúdos abordados nos testes construídos por professores refere-se àqueles lecionados no final de uma etapa ou no decorrer do ano letivo em curso. O professor não dará a seus alunos um teste que contenha uma gama de conteúdos relativas aos anos letivos anteriores. Na maioria das vezes requer, como matéria para prova, apenas os últimos conteúdos ministrados em aula.

Além disso, os testes construídos pelo professor costumam ser destinados apenas à verificação da aprendizagem, ou seja, visam somente saber se o aluno atingiu, ou não, o domínio de determinado conteúdo. Dessa forma, não é necessário realizar uma comparação criteriosa do desempenho de cada aluno com o restante da classe.

Ao contrário dos testes construídos pelo professor, o Teste Padronizado permite avaliação de um grande número de conteúdos, habilidades e competências, pois o objetivo desse teste é conhecer uma distribuição de comportamentos dos examinados com a finalidade de comparar os desempenhos. Esse aspecto é de fundamental importância quando seu uso destina-se a um concurso.

No que se refere especificamente à modalidade dos testes padronizados de aproveitamento, nota-se que são classificados de duas formas. A primeira é designada por teste baseado em norma ou teste de discriminação. Seu objetivo é distinguir indivíduos de acordo com seu grau de desempenho. A segunda classificação, denominada teste baseado em critério ou teste de domínio, tem por objetivo determinar se uma pessoa atingiu um conjunto específico de objetivos ou domínio (MEHRENS, 1978, p.14).

QUADRO 1

Comparação entre testes de aproveitamento

Características	Testes construídos pelo professor	Testes padronizados
Instrução para aplicação e correção	via de regra, não são uniformes e específicas	normas específicas padronizam os processos de aplicação e correção
Amostragem de conteúdo	determinada pelo professor	sistemática, determinada por especialistas em currículo e no assunto;

		envolve extensa investigação de compêndios, textos e programas existentes
Construção	rápida e improvisada; não há projetos de testes, itens objetivos provisórios, análise de itens objetivos ou revisão; a qualidade dos itens pode ser duvidosa	Meticulosa, incluindo enunciado de objetivos e projeto de teste, itens provisórios, análise e revisões de itens
Normas	locais, baseadas na sala de aula	da própria instituição; nacionais e/ou da região geográfica
Objetivos e uso	particulares, estabelecidos pelo professor; compara classes	mede objetivos curriculares amplos; compara classes, escolas e Estados

FONTE: MEHRENS, 1978, p.191(Adaptação)

Os testes padronizados de aproveitamento baseados em discriminação devem produzir, segundo Silva (1992, p.170), resultados variados, para que possam atingir seu objetivo principal: tornar possível a realização de comparações entre os resultados alcançados pelos indivíduos.

2.2 Modalidades de medidas

Duas variáveis estatísticas são freqüentemente utilizadas na análise⁷ psicométrica de itens objetivos de testes padronizados de aproveitamento baseados em discriminação: o índice de acerto e o índice de discriminação.

De acordo com Dias et al. (1977, p.22), o índice de acerto é assim denominado porque fornece a medida da facilidade do item conforme a percentagem de respostas corretas. Dessa forma, quanto mais elevado o valor numérico do índice de acerto, mais fácil o item será considerado. Nota-se que o valor alcançado por um item, ou sua

⁷ Nessas análises consideram-se também os distratores, que são as opções incorretas do teste de múltipla escolha. (SILVA, 1992, p.171)

classificação como fácil ou difícil, irá variar conforme a aptidão ou a capacidade do grupo a que for aplicado.

Ainda de acordo com o referido autor, os testes que compreenderem itens muito fáceis e/ou muito difíceis apresentarão um coeficiente de confiabilidade⁸ muito baixo e, conseqüentemente, discriminarão menos. Por isso, Dias et. al. (1977, p.23) recomenda: “[...] na construção de testes de desempenho, devem predominar itens de facilidade média”.

O índice de acerto pode ser calculado dividindo-se o total de acertos no item pela quantidade de examinandos que fizeram o teste. Lindeman (1972, p.89) apresenta, como forma de calcular o índice de acerto, a divisão do número de examinandos que responderam corretamente ao item pelo número total de alunos dos 27% que se saíram melhor no teste somado aos 27% dos que se saíram pior no teste. Os valores obtidos nesse cálculo poderão variar de zero a um – quanto mais alto o valor resultante, mais fácil é considerada a questão.

O índice de discriminação mede a extensão em que o item distingue os que se saíram melhor dos que se saíram pior no teste, ou seja, discrimina os estudantes de diferentes níveis de aproveitamento.

Pode-se utilizar a correlação bisserial de pontos para calcular o índice de acerto. A correlação bisserial é uma estimativa do coeficiente de correlação de Pearson. Nesse caso, exige-se que os itens objetivos sejam corrigidos de maneira a estabelecer uma dicotomia entre certo ou errado.

O índice de discriminação pode ser calculado, também, por meio de outros procedimentos. Silva (1992, p.172) e Dias et. al. (1977, p.23) descrevem as seguintes etapas para a obtenção desse índice:

- a) aplicar o teste a um mínimo de 30 sujeitos;
- b) computar o total de acertos de cada examinando no teste;
- c) arrumar o teste em ordem crescente ou decrescente de total de acertos;
- d) calcular 27% sobre o número de testes aplicados, arredondando-se o resultado para um número inteiro;

⁸ O coeficiente de confiabilidade ou precisão refere-se “[...] a propriedade de um teste que, aplicado várias vezes às mesmas pessoas, gera resultados semelhantes. Um teste confiável proporciona sempre a mesma classificação” (DIAS et. al., 1977, p.27).

- e) separar os 27% superiores, compreendendo os que tiraram os escores mais altos, e os 27% inferiores, abrangendo os que tiraram os escores mais baixos, para achar N ($N = 27\% \text{ superiores} + 27\% \text{ inferiores}$);
- f) calcular a diferença entre os que acertaram o item no grupo superior e os que acertaram o item no grupo inferior;
- g) dividir a diferença em questão pela soma do número de candidatos de cada um dos subgrupos.

O quociente obtido na última etapa do processo descrito anteriormente pode ser expresso em fração decimal. A essa fração denomina-se índice de discriminação.

Popham (1978), citado por Silva (1992, p.173), apresenta a justificativa para o uso de grupos extremos de 27% para análise de itens objetivos. Segundo esse autor, estudos estatísticos comprovam que essa porcentagem permite a formação de grupos extremos tão grandes e diferentes quanto possível. Nesse caso, o grau de confiança nos índices calculados é bem maior que para qualquer outra porcentagem.

Para analisar o índice de acerto e o índice de discriminação é necessário utilizar-se de escalas de classificação, as quais, criadas por Ebel, são de grande relevância para a interpretação desses índices. Cabe destacar que, apesar de a concepção de tais escalas ter ocorrido na década de 60, elas são utilizadas e aceitas contemporaneamente por especialistas em medidas educacionais.

A propósito, apresentam-se nos QUADROS 2 e 3 as referidas escalas, elaboradas por Ebel (1965), citado por Silva (1992, p.173).

QUADRO 2

Classificação dos itens objetivos quanto ao Índice de Acerto

Faixas do Índice de Acerto	Classificação
0,80 a 1,00	muito fáceis
0,60 a 0,79	fáceis
0,40 a 0,59	médias
0,20 a 0,39	difíceis
0,00 a 0,19	muito difíceis

FONTE: SILVA,1991, p. 173. (Adaptação)

QUADRO 3

Classificação dos itens objetivos quanto ao Índice de Discriminação

Faixas do Índice de Discriminação	Classificação
0,40 e acima	muito bons
0,30 a 0,39	razoavelmente bons, mas possivelmente sujeitos a melhoramento
0,20 a 0,29	marginais, usualmente necessitando, e estando sujeitos a melhoramento
0,19 e abaixo	deficientes, devem ser rejeitados ou melhorados por revisão
0,00 a 0,19	muito deficientes

FONTE: SILVA,1991, p. 173. (Adaptação)

De acordo com Silva (1992, p.175) e Lindeman (1972, p.91), o poder de discriminação de um teste está diretamente ligado ao grau de dificuldade do item. Sendo assim, itens objetivos extremamente fáceis ou extremamente difíceis não discriminam os estudantes, pois são errados ou acertados por todos ou quase todos os examinandos.

Os referidos autores constatam também que, se um item foi mais acertado pelos examinandos que se saíram pior no teste, ele apresenta um índice negativo de discriminação. Esse caso, considerado anômalo, demonstra que o item foi mal formulado.

Cabe destacar que a variabilidade de resultados do teste só ocorrerá se seus itens apresentarem uma discriminação positiva, ou seja, se forem mais respondidos corretamente pelos examinandos que alcançaram melhores resultados no teste do que pelos examinandos que não atingiram bons resultados no teste.

Nesse contexto, é possível afirmar que os itens objetivos que melhor discriminam são os de dificuldade média. Nesse caso, o item poderá atingir o valor **um**, ou seja, o valor máximo de discriminação. Porém, isso só ocorrerá se todos os examinandos do grupo que se saiu melhor no teste acertarem, e se todos do grupo que se saiu pior no teste errarem.

Dias (1977, p.24) confirma, como apontado anteriormente, que o índice de discriminação possui “[...] a propriedade de favorecer os itens de dificuldade média”. O autor acrescenta também que “[...] caso somente 50% dos examinandos acertem o item, é possível que todos os do grupo superior e nenhum do grupo inferior respondeu corretamente”.

O referido autor afirma ainda que os psicometristas observam que, mesmo que um item seja altamente discriminatório em uma amostra, em outra ele pode apresentar baixa ou negativa discriminação. No entanto, ressalta que, no caso do vestibular, que envolve os cálculos não só de uma amostra, mas de toda a população, os índices de discriminação alcançam máxima fidedignidade.

Nesse sentido, Dias et. al. (1977, p.24,25) comenta:

Em testes como o do exame vestibular, cuja função principal seja distinguir entre diferentes níveis de desempenho, com a maior precisão possível, é de se esperar que os itens tenham elevado índice de discriminação. [...] Entre dois testes, semelhantes nos demais aspectos, aquele cuja média de índices de discriminação for mais alta, será sempre o melhor, isto é, o mais confiável.

Cabe destacar que, apesar de as medidas de desempenho – índice de acerto e índice de discriminação – serem freqüentemente utilizadas para classificar os itens objetivos quanto à sua validade, o uso dessas medidas pode trazer contribuições de outra natureza.

Chama-se atenção para o fato de que a análise desses índices pode propiciar, por exemplo, inferências a respeito do aprendizado alcançado pelos examinandos. Ou seja, as medidas de desempenho podem revelar especificidades sobre as experiências de ensino/aprendizagem.

No caso dos concursos vestibulares, a interpretação das medidas de desempenho não precisa ficar restrita apenas às afirmações a respeito do grau de dificuldade ou qualidade do item de prova. Pode-se também atentar-se para o desempenho dos candidatos e interpretar-se se este foi influenciado pelo tipo de conteúdo abordado no item, ou ainda, pelas competências e habilidades solicitadas pelo item de prova.

Lindeman (1972, p.88) assegura a importância de avaliar os resultados alcançados pelos examinandos através da seguinte afirmação:

[...] muitas informações valiosas sobre o desempenho dos estudantes, a eficácia do ensino e as características das questões

de teste podem ser obtidas mediante uma análise das respostas dadas pelos alunos. Um exame da percentagem de estudantes que responderam todas as questões corretamente pode revelar áreas em que o ensino foi particularmente bom ou medíocre. A análise das características de questões, tais como são reveladas pelas respostas dos alunos a determinadas questões, também pode ajudar a identificar os bons itens [...] Cada um desses tipos de informação serve a um propósito diferente que pode ser útil ao professor consciencioso, melhorando não só o ensino como também os processos de testagem.

É necessário esclarecer que, neste trabalho, a definição do termo “conteúdo” não está restrita ao sentido disciplinar, ou seja, não se apresenta como conjunto de conhecimentos relacionados a matérias escolares; relaciona-se, sim, como aponta Zaballa (1998, p.30), apoiado em Coll (1986), aos conteúdos agrupados, segundo sejam conceituais, procedimentais e atitudinais. Segundo essa concepção, os conteúdos conceituais compreendem os fatos, conceitos e princípios; os procedimentais, os procedimentos, as técnicas e os métodos; os atitudinais, os valores, as atitudes e as normas.

Nesse sentido, ao considerar que um item de prova tenha sido bem elaborado e que esse mesmo item tenha obtido um elevado índice de acerto, pode-se inferir que, possivelmente, o conteúdo contemplado no item foi bem compreendido pelos estudantes nos níveis escolares anteriores. E, ainda, que as habilidades e competências requeridas pelo item foram atingidas pela maioria dos estudantes ou são do domínio da maioria deles.

Caso ocorra o contrário, ou seja, caso o item apresente um baixo índice de acerto, pode-se deduzir que o conteúdo, as habilidades e competências solicitadas pelo item não foram suficientemente trabalhados na educação básica.

Da mesma forma, ao interpretar o índice de discriminação, pode-se observar, por exemplo, que, se um item de prova discriminou bem, isso significa que o item conseguiu atingir e separar os melhores candidatos, o que pode denotar que os conteúdos, habilidades e competências requisitados pelo item foram trabalhados com ênfase e eficácia diferentes nas diversas instituições escolares. Nesse caso, talvez a maioria dos estudantes que se saíram pior no teste não conseguiu acertá-lo por não ter desenvolvido a aprendizagem necessária do que foi solicitado pelo item de prova.

Cabe, então, aos professores indagar sobre as eventuais falhas no processo de ensino/aprendizagem que levaram alguns estudantes a acertar o item e pesquisar sobre o que pode ser feito para melhorar o nível deles em relação ao que foi abordado pelo item.

Dessa forma, depois de interpretados, os resultados obtidos podem servir como um panorama do nível em que se encontravam os estudantes no período em que o teste foi realizado. Assim, o referido panorama permitirá a verificação das principais deficiências na aprendizagem. Uma vez constatadas as principais dificuldades do grupo de estudantes avaliados, educadores e professores poderão discutir sobre as possibilidades de superação dessas dificuldades durante o planejamento de suas práticas docentes.

Já para os estudantes que passaram pelo teste, a interpretação das medidas de desempenho poderão propiciar, caso haja interesse, o redirecionamento de seus estudos. Nesse caso, o estudante deverá estudar dando a ênfase necessária aos conteúdos abordados pelos itens em que teve dificuldade, na tentativa de melhorar seu aprendizado, ou para que se saia melhor em testes futuros.

3. As provas objetivas do vestibular da UFMG como Teste Padronizado

As provas objetivas aplicadas na primeira etapa dos concursos vestibulares da UFMG correspondem a uma bateria de testes padronizados de aproveitamento. Nota-se que as provas em questão, assim como os testes padronizados de aproveitamento, fundamentam-se na discriminação dos candidatos.

Assim como na maioria dos grandes concursos vestibulares, a UFMG tem entendido, até o momento, que o mérito do estudante é a principal forma de qualificá-lo para a entrada na Universidade. Nesse sentido, a referida fundamentação das provas dos concursos vestibulares da UFMG justifica-se pela sua natureza, que é seletiva. Dessa forma, os itens formulados pelas equipes de elaboração das provas objetivam a avaliação de conhecimentos e habilidades atuais dos candidatos, utilizando-se de conteúdos importantes para a *performance* universitária dos futuros alunos.

Ressalta-se também outra característica semelhante aos testes padronizados, a qual diz respeito à aplicação deles a uma população relativamente homogênea. A maioria dos vestibulandos possui o mesmo nível de escolaridade, ou seja, são portadores de certificado de conclusão do Ensino Médio.

Destaca-se ainda que as provas seguem padrões uniformes de aplicação para todos os candidatos. Isso permite a comparação e avaliação do desempenho individual e coletivo dos vestibulandos em diferentes áreas do conhecimento, através dos resultados dos testes expressos pelas medidas educacionais de desempenho.

A propósito, observa-se que, assim como os testes padronizados de aproveitamento, as provas de primeira etapa dos concursos vestibulares da UFMG possibilitam a verificação da validade de seus itens, o que, por sua vez, permite um aperfeiçoamento do processo seletivo.

Por fim, cabe ressaltar que, assim como no TPA, os resultados do desempenho dos vestibulandos dos concursos vestibulares da UFMG podem ser empregados para diagnosticar dificuldades de aprendizagem. E isso, de acordo com Lindeman (1972, p.135), poderá contribuir para a erradicação dessas dificuldades, caso suas origens sejam focalizadas de modo contundente.

3.1 As provas de Geografia de vestibulares da UFMG sob o crivo de medidas de desempenho

A cada ano, após seus concursos, a UFMG sistematicamente divulga para a comunidade e com ela discute as provas realizadas, bem como os resultados alcançados pelos candidatos, apresentados através de medidas educacionais.

Para professores do Ensino Médio das redes pública e particular, esses resultados têm suscitado muito interesse, uma vez que lhes possibilitam a oportunidade de confirmarem avaliações dos itens objetivos de provas, particularmente nos aspectos relacionados à facilidade ou à dificuldade que poderiam ter apresentado aos candidatos.

Para alguns dos professores que se encontram envolvidos no processo seletivo da UFMG, esses resultados têm sido considerados instigadores. Esses professores têm se organizado em grupos⁹ de pesquisa e se dedicam a analisar o desempenho dos vestibulandos de modo a obter resultados balizados pelo rigor científico. Nota-se que o resultado dessas pesquisas tem se tornado público por meio de monografias, dissertações e artigos.

No Departamento de Geografia, do Instituto de Geociências da UFMG, a partir de 1998 foram realizadas várias pesquisas concernentes aos itens de prova. No decorrer desse período, foram apresentadas cerca de dezesseis monografias relacionadas a diversos temas da Geografia Física e Humana, além de alguns artigos e apresentações em encontros da comunidade científica.

Entre as monografias citadas, insere-se um estudo relativo às provas de Geografia da primeira etapa dos concursos vestibulares ocorridos entre 1997 e 2001. Esse estudo, que procurou estabelecer relações entre o desempenho dos vestibulandos e os itens de prova com representações gráficas, será apresentado a seguir.

⁹ Entre os grupos de estudos, destaca-se o Grupo de Avaliação e Medidas Educacionais (GAME), ligado à Faculdade de Educação.

3.2 O desempenho dos vestibulandos nos concursos vestibulares da UFMG entre 1997 e 2001 nos itens de prova com e sem representações gráficas das provas objetivas de Geografia

O estudo em questão refere-se a uma questão abordada nos encontros¹⁰ promovidos pela COPEVE, os quais, no caso específico da prova de Geografia, acontecem geralmente no Instituto de Geociências da UFMG, quando os professores são convidados a discutir e analisar os itens presentes nessa prova.

Nessa oportunidade, observou-se que, ao analisarem o material de exame, ou seja, as provas do vestibular de Geografia, alguns dos professores tecem comentários e emitem juízos de valor a propósito da dificuldade dos itens objetivos de prova.

A presença ou ausência de representações gráficas¹¹ constitui para muitos deles um critério de análise dos itens objetivos de prova. Em geral, há uma crença por parte desses professores de que os itens objetivos da prova de Geografia que contêm representações gráficas seriam mais fáceis, proporcionando um melhor desempenho de seus alunos.

Essa crença originou o estudo de França (2000, 2001, 2002), que teve como objetivo a verificação da presença de diferenças estatisticamente significativas entre o desempenho dos vestibulandos em questões com e sem representações gráficas, nas provas de Geografia da primeira etapa dos concursos vestibulares da UFMG, concernentes ao período de 1997¹², a 2001.

Duas hipóteses foram levantadas para alcançar o objetivo proposto. Na primeira, cuja variável utilizada foi o índice de acerto, a hipótese nula, H_0 , corresponde à suposição de que a média do grupo de itens de prova com representação gráfica (M_{g1}) é igual à média do grupo dos itens de prova sem representação gráfica (M_{g2}), ou seja, as médias dos dois grupos são iguais. A hipótese alternativa, H_1 , é de que a média do grupo de itens de prova com representação gráfica (M_{g1}) difere da média do grupo dos itens de prova sem representação gráfica (M_{g2}), ou seja, as médias dos dois grupos são diferentes.

No segundo teste, cuja variável utilizada foi o índice de discriminação, a hipótese nula, H_0 , refere-se à suposição de que a média do grupo de itens de prova com

¹⁰ Anualmente, a UFMG proporciona a todos os professores do Ensino Médio da rede pública, da rede privada e de cursos pré-vestibulares um encontro com a equipe de membros da comissão de elaboração das provas do concurso vestibular de todas as disciplinas.

¹¹ O estudo considerou como representações gráficas todos os documentos geográficos apresentados sob a forma simbólica ou gráfica e matemático-estatística que constam nos enunciados das questões das provas de Geografia. Dessa forma, as representações gráficas compreendem mapas, fotografias aéreas, imagens de satélite, croquis, blocos -diagrama, quadros, tabelas, diagramas.

¹² Até 1997, os concursos vestibulares de primeira etapa da UFMG aconteciam pela manhã e à tarde; por isso, havia dois cadernos de prova para cada disciplina.

ilustração (M_{g1}) é igual à média do grupo dos itens de prova sem ilustração (M_{g2}), ou seja, as médias dos dois grupos são iguais. A hipótese alternativa, H_1 , é de que a média do grupo de itens de prova com ilustração (M_{g1}) difere da média do grupo dos itens de prova sem ilustração (M_{g2}), ou seja, as médias dos grupos são diferentes.

A realização da pesquisa em questão exigiu, na primeira fase do trabalho, a classificação dos itens objetivos de provas, segundo a presença ou ausência de representações gráficas. Em seguida, os itens objetivos, assim classificados, foram relacionados aos dados estatísticos, fornecidos pela COPEVE, que indicam o desempenho dos vestibulandos.

Na população estudada, a participação percentual dos itens objetivos com e sem representações gráficas no total de itens objetivos, por prova, é variável, como pode ser observado no GRÁFICO 1. Destaca-se, nessas provas, a participação dos itens objetivos com representações gráficas, cuja variação ocorre do mínimo de 35%, referente ao ano de 1997, ao máximo de 67%, correspondente à prova de 1999. No período considerado, as provas compreendem, em média, 39,6% de itens objetivos com representações gráficas.

Nas provas em questão, os mapas correspondem às representações gráficas mais utilizadas, seguidos pelos diagramas e croquis, como pode ser observado na TABELA 1. Nessas provas, encontra-se também de um a três outros tipos de representações gráficas que, mesmo não se destacando nas participações absolutas e relativas, contribuem para diversificar as representações gráficas dos itens de provas.

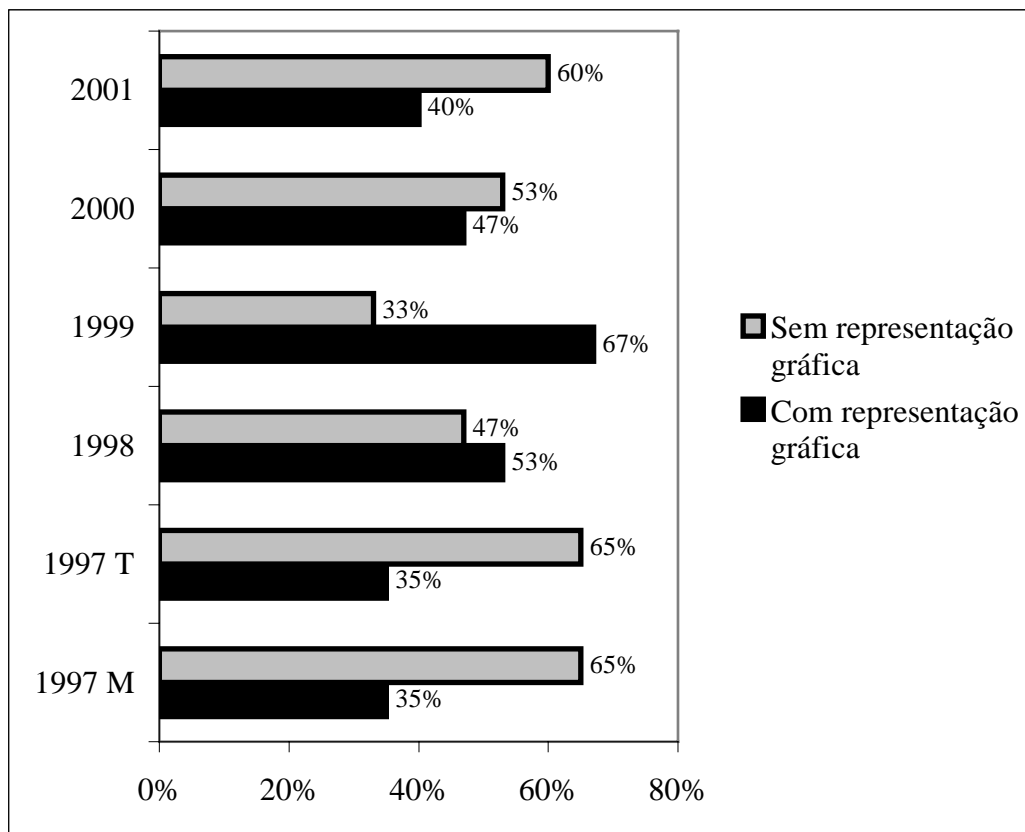


GRÁFICO 1- Distribuição da frequência dos itens objetivos com e sem representações gráficas nas provas objetivas de Geografia dos concursos vestibulares de 1997 a 2001.

FONTE: FRANÇA et al., 2001.

TABELA 1

Freqüência dos tipos de representações gráficas presentes nos itens objetivos de prova de Geografia da 1ª etapa dos concursos vestibulares da UFMG realizados entre 1997 e 2001

Tipos	Frequência Abs. (por prova*)						Freq. Abs.	Freq. %
	1997 manhã	1997 tarde	1998	1999	2000	2001	97 -01	Total
Mapa.....	4	5	5	6	3	2	25	51,0
Diagrama.....	1	-	2	1	2	-	6	12,24
Croqui.....	-	1	1	2	1	1	6	12,24

Bloco-Diagrama.	-	-	-	1	1	1	3	6,12
Foto aérea...	1	1	-	-	-	-	2	4,08
Perfil.....	7	1	-	-	-	1	2	4,08
Fluxograma..	-	-	-	1	1	-	2	4,08
Charge.....	-	-	1	-	-	-	1	2,04
Planta.....	1	-	-	-	-	-	1	2,04
Tabela.....	-	-	-	-	-	1	1	2,04
TOTAL.....	14	8	9	11	8	6	49	100,0

* um item de prova pode compreender mais de um tipo de representação gráfica.

FONTE: FRANÇA et al., 2001.

Quanto ao desempenho dos vestibulandos, observa-se que o grupo de questões sem representações gráficas apresenta-se, na TABELA 2, com maior porcentagem de questões oscilando entre difíceis e fáceis. Já o grupo de questões com representações gráficas, em que a maior porcentagem de questões manteve-se entre as moderadas, apresentou, apenas em 1998 e 2000, o correspondente a 14% de questões consideradas fáceis, como se observa na TABELA 3. Isso evidencia que os itens com representações gráficas aproximam-se do que os teóricos afirmam, ou seja: na construção de testes de aproveitamento, devem predominar itens de dificuldade média.

No grupo de questões sem representações gráficas, apresentado na TABELA 4, há maior porcentagem de questões com baixa de discriminação e menor porcentagem com boa discriminação. Já no grupo de questões com representações gráficas, percebe-se, com exceção de 1997 manhã, que, em todos os anos, houve um grande percentual de questões com boa discriminação, destacando-se o ano de 1999, que atingiu 90% das questões com representações gráficas, como se demonstra na TABELA 5.

Os dados referentes às variáveis **índice de acerto** e **índice de discriminação** foram sistematizados para a aplicação do teste das hipóteses, utilizando a técnica estatística Teste “T de Student”. O valor estipulado do nível de significância para os testes de hipóteses foi 0,05.

No primeiro teste de hipótese, cuja variável utilizada foi o índice de acerto, a média para as questões com representações gráficas foi 0,4861 com desvio padrão de 0,1245. A média para as questões sem representações gráficas foi 0,4696, com desvio padrão de 0,1779.

Esses resultados demonstram que não existe diferença estatisticamente comprovada em relação à variável índice de acerto, indicando a ausência de desempenho

diferenciado por parte dos vestibulandos entre os itens de prova com e sem representações gráficas. Isso leva à afirmação de que, considerando-se o período analisado, os itens de prova com representações gráficas não são mais fáceis do que os itens de prova sem representações gráficas.

TABELA 2

Desempenho dos vestibulandos nas provas de Geografia da 1ª etapa dos concursos vestibulares da UFMG segundo o índice de acerto por ano

Itens sem representações gráficas

Índice de Acerto	97 M %	97 T %	98 %	99 %	00 %	01 %
0,10-0,40..... (difícil)	53,8	38,5	42,9	60,0	12,5	16,6
0,41-0,70..... moderada	46,2	46,2	28,6	20,0	87,5	66,8
0,71-1,00..... (fácil)	0	15,4	28,6	20,0	0	16,6
TOTAL.....	100,0	100,0	100,0	100,0	100,0	100,0

FONTE: FRANÇA, et. al., 2001.

TABELA 3

Desempenho dos vestibulandos nas provas de Geografia da 1ª etapa dos concursos vestibulares da UFMG segundo o índice de acerto por ano

Itens com representações gráficas

Índice de Acerto	97 M %	97 T %	98 %	99 %	00 %	01 %
0,10- 0,40 (difícil)	14,3	28,6	42,9	20,0	28,6	33,3
0,41- 0,70 moderada	85,7	71,4	42,9	80,0	57,1	66,6
0,71-1,00 (fácil)	0	0	14,3	0	14,3	0
TOTAL	100,0	100,0	100,0	100,0	100,0	100,0

FONTE: FRANÇA, et. al., 2001.

TABELA 4

Desempenho dos vestibulandos nas provas de Geografia da 1ª etapa dos concursos vestibulares da UFMG segundo o índice de discriminação por ano

Itens sem representação gráfica						
Índice de discriminação	97 M	97 T	98	99	00	01
	%	%	%	%	%	%
<0,4 baixo	76,9	61,5	28,6	60,0	75,0	44,4
>OU = 0,4 ... bom	23,1	38,5	71,4	40,0	25,0	55,6
TOTAL.....	100,00	100,00	100,00	100,00	100,00	100,00

FONTE: FRANÇA, et. al., 2001.

TABELA 5

Desempenho dos vestibulandos nas provas de Geografia da 1ª etapa dos concursos vestibulares da UFMG segundo o índice de discriminação por ano

Itens com representação gráfica						
Índice de Discriminação	97 M	98 T	98	99	00	01
	%	%	%	%	%	%
<0,4 baixo	71,4	28,6	14,3	10,0	28,6	50,0
>OU = 0,4 bom	28,6	71,4	85,7	90,0	71,4	50,0
TOTAL	100,00	100,00	100,00	100,00	100,00	100,00

FONTE: FRANÇA, et. al., 2001.

No segundo teste de hipótese, cuja variável utilizada foi o índice de discriminação, a média para as questões com representações gráficas foi 0,4162, com desvio padrão de 0,1162. A média para as questões sem representações gráficas foi 0,3563, com desvio padrão de 0,1148.

O resultado do segundo teste de hipótese, relativo à variável índice de discriminação, demonstrou uma diferença estatisticamente comprovada entre o desempenho dos vestibulandos nos itens de prova com e sem representações gráficas. Isso

permite a conclusão de que, no período considerado, os itens de prova com representações gráficas discriminam mais os vestibulandos que os itens de prova sem representações gráficas.

Concluiu-se, então, que os resultados estatísticos obtidos nessa pesquisa representaram um forte argumento contrário à crença de que as questões com representações gráficas, presentes nas provas objetivas de Geografia dos concursos vestibulares da UFMG, seriam mais fáceis que as questões sem representações gráficas.

Considerando-se a população analisada, os resultados evidenciaram que os problemas concernentes a cada um dos itens – as atividades cognitivas que encerram ou os objetivos educacionais que procuram alcançar – não se resumem à simples identificação de dados, armazenados nas representações gráficas, nem à definição da veracidade ou falsidade das alternativas que os exploram. Pode-se considerar aceitável que os itens estejam exigindo muito mais do que aquisições de conhecimento em nível elementar, armazenados na memória representada pelas representações gráficas.

Finalmente, cabe ressaltar que os resultados alcançados sugerem a inferência de que, além do bom conhecimento do conteúdo divulgado no programa, os examinandos devem estar habilitados para resolver itens de prova com representações gráficas, pois a qualidade cognitiva desses itens requer o uso de diferentes configurações cognitivas por parte dos estudantes.

ANEXO – Exemplos de itens de prova

QUESTÃO 27

Com relação aos novos países industriais, é **INCORRETO** afirmar que, atualmente, eles

- A) correspondem a economias atingidas fortemente pela saída de capitais estrangeiros de caráter especulativo.
- B) demonstram dificuldades na reconversão de suas economias, exigida pelas novas tecnologias de produção e gestão, crescentemente valorizadas.
- C) abandonam a política de privatizações das empresas estatais, a fim de proteger suas metas de desenvolvimento sócioeconômico da atual crise mundial.
- D) enfrentam obstáculos crescentes para ampliar as exportações e assegurar mercados para seus produtos.

MEDIDAS EDUCACIONAIS

Alternativa correta : C

Índice de Acerto : 0,76

Índice de Discriminação: 0,54

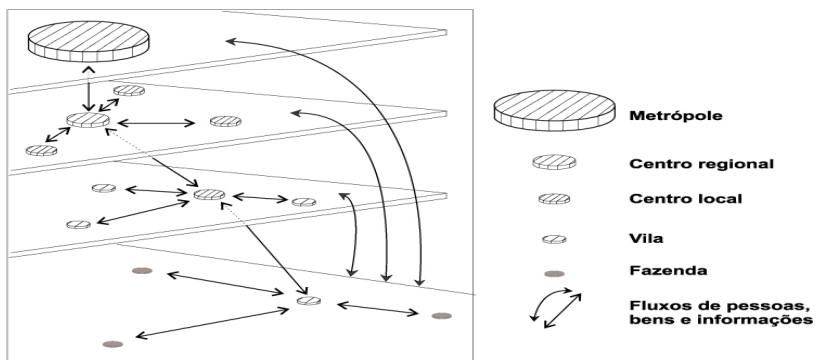
Atratividade da alternativa (%): A (12,0); B (8,8); C(76,2); D (3,0).

FIGURA 1- Questão 27 da prova de Geografia aplicada na 1ª etapa do concurso vestibular da UFMG de 1999 e resultados concernentes.

FONTE : COPEVE, UFMG.

QUESTÃO 16

A figura representa um modelo de organização do espaço geográfico.



FONTE: Adaptado de DINIZ, J.A.F. *Geografia da Agricultura*.

São Paulo: DIFEL, 1984. p.232

Todas as alternativas contêm aspectos da organização do espaço representada nesse modelo, **EXCETO**

- A) Hierarquia da rede urbana
- B) Relações interurbanas
- C) Funções urbanas
- D) Interdependência rural-urbana

MEDIDAS EDUCACIONAIS

Alternativa correta: C

Índice de Acerto: 0,47

Índice de Discriminação: 0,58

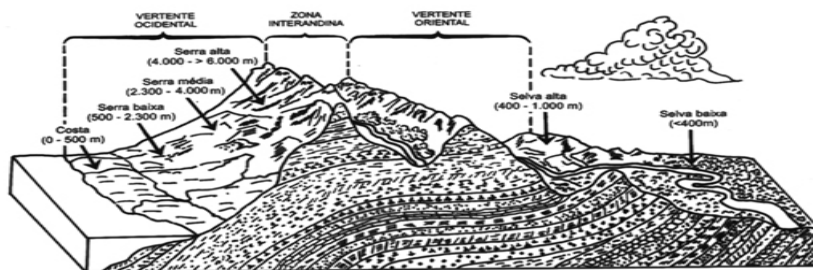
Atratividade da alternativa (%): A (22,8); B (12,7); C (47,0); D (17,4%).

FIGURA 2- Questão 16 da prova de Geografia aplicada na 1ª etapa do concurso vestibular da UFMG de 1999e resultados concernentes.

FONTE : COPEVE, UFMG.

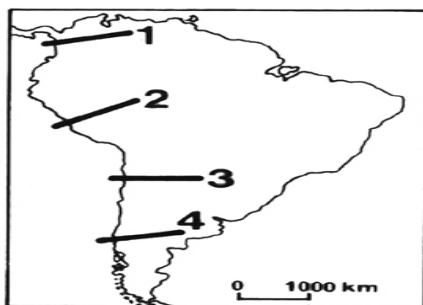
QUESTÃO 18

Analise o bloco-diagrama.



FONTE: ONERN

Agora, observe, no mapa, a localização das seções 1, 2, 3 e 4.



Com base na observação desse mapa, é **CORRETO** afirmar que a paisagem representada no bloco-diagrama é encontrada ao longo da seção

A) 2.

B) 4.

C) 1.
D) 3.

MEDIDAS EDUCACIONAIS

Alternativa correta : A

Índice de Acerto : 0,42

Índice de Discriminação: 0,41

Atratividade da alternativa (%): A(42,0); B (13,2); C (15,1); D (29,8).

FIGURA 3- Questão 18 da prova de Geografia aplicada na 1ª etapa do concurso vestibular da UFMG de 1999 e resultados concernentes.

FONTE : COPEVE, UFMG.

REFERÊNCIAS

AZZI, Sandra. *Relação entre resultados do concurso vestibular e desempenho acadêmico*. Tese de Mestrado - Faculdade de Educação. Universidade Federal de Minas Gerais. Belo Horizonte, 1972.

BRADFIELD, James M. & MOREDOCK H. Stewart. *Medidas e testes em educação*. The Macmillan Company, 1957.

COPEVE, *Manual do candidato*. Edição única, Belo Horizonte: Editora UFMG, 2004,75 p. Disponível em < [http:// www.ufmg.br/copeve](http://www.ufmg.br/copeve)> acesso em 15 de dezembro de 2003.

CUNHA, Nadia Franco da. *Vestibular na Guanabara*. MEC, INEP, Cbpe, RJ 1968. 406p.

DIAS, José Augusto. *Caracterização dos candidatos ao vestibular*. FEUSP, São Paulo, 1978.

DIAS, José Augusto et. al. *Testes objetivos versus provas analíticas dissertativas no vestibular Fuvest*. FEUSP, São Paulo, 1977.

EBEL, Robert L. *Essentials of educational measurement*. New Jersey, USA: Prentice-Hall, Inc. Englewood, 1972.

ESTEVES, Oyara Petersen. *Testes e Medidas em Educação*. Rio de Janeiro, 1967.

FRANÇA, Bárbara Lúcia Pinheiro Oliveira. *O desempenho dos vestibulandos nos concursos vestibulares da UFMG nos anos 1997,1998,1999 e 2000 nos itens de prova com e sem ilustração das provas objetivas de Geografia*. 2000. 105 f. Monografia (Graduação em Geografia) - Instituto de Geociências, Universidade Federal de Minas Gerais, Belo Horizonte.

LINDEMAN, Richard H. *Medidas educacionais, testes objetivos e outros instrumentos de medida para a avaliação da aprendizagem*. Trad. Leonel Vallandro. Porto Alegre, Editora Globo, 1972.

MAGER, Robert Frank. *Medindo os objetivos de ensino*. Trad. Maria Ângela V. Almeida. Porto Alegre, Globo, 1970.

MEDEIROS, Ethel. *Provas objetivas, discursivas, orais e práticas: técnicas de construção – 9 ed. Ampliada*. Rio de Janeiro: Editora Fundação Getúlio Vargas, 1989 XI, 178p.

MEHRENS, Willian A. *Testes padronizados em educação*. São Paulo,1978.

SEMINÁRIO SOBRE O VESTIBULAR. *Dilemas do acesso ao ensino superior no Brasil*. Pontifícia Universidade Católica do Rio de Janeiro, 1979. Anais, Rio de Janeiro ,PUC,1980. 310p. Conselho de Reitores das Universidades Brasileiras, Secretaria de Ensino Superior, PUC/RJ.

SILVA, Oscar V. *Manual do vestibular: guia para planejamento, execução e avaliação do concurso vestibular*. Brasília, Secretaria da Educação Superior, 1987.

SILVA, Ceres Santos da. *Medidas e avaliação em educação*. São Paulo: Editora Vozes, 1992.

SUND, Robert B. *Objetivos comportamentais e medidas de avaliação*. Trad. Célia Augusta Teixeira. et. Al. SP, EPU: Editora da Universidade de São Paulo, 1978.