

ELABORAÇÃO DE QUESTÕES DE MÚLTIPLA ESCOLHA

BERNARDO BUCHWEITZ*

1 - INTRODUÇÃO

No meio educacional, e mesmo fora dele, há um consenso geral sobre a importância da avaliação da aprendizagem dos estudantes. No entanto, uma análise dos instrumentos de avaliação que freqüentemente chegam às nossas mãos, permite-nos afirmar que na maioria dos casos há pouca preocupação com uma adequada elaboração das questões desses instrumentos, independente de serem objetivas ou discursivas. São freqüentes os descuidos relacionados com a inclusão de questões que envolvem conteúdos ausentes nos programas ou que estão fora dos objetivos estabelecidos. Também é comum encontrarmos questões extremamente difíceis ou totalmente óbvias. Outras vezes aparecem falhas técnicas na elaboração das questões.

Sem pretensão alguma de esgotar o assunto, com este trabalho pretendemos trazer um pouco da nossa experiência sobre planejamento, elaboração e análise de questões, numa tentativa de contribuir para o aperfeiçoamento dessas atividades quando forem executadas por outros interessados. Como esse assunto sobre os vários

* Professor da Faculdade de Educação da Universidade Federal de Pelotas.

tipos de questões tornaria o texto muito longo, limitaremos o trabalho a apenas um tipo de questão, o de múltipla escolha.

2 - FASES DE ELABORAÇÃO DE UMA PROVA

Para a organização e elaboração de uma prova normalmente são seguidos alguns passos, dos quais destacamos os seguintes.

1. Definição dos objetivos gerais da prova
2. Elaboração do programa
3. Escolha da equipe de elaboração e revisão
4. Distribuição das questões por conteúdo e habilidade
5. Elaboração das questões
6. Revisão das questões
7. Datilografia
8. Revisão da datilografia
9. Impressão da prova

Concluídos esses passos, desenvolvem-se as várias fases de aplicação da prova, seguidas da correção da prova e da análise dos dados obtidos.

As fases acima enumeradas podem facilmente ser entendidas pelas pessoas que já estiveram envolvidas na organização e na elaboração de uma prova para algum concurso ou habilitação. Embora se possa fazer descrições e comentários sobre cada uma dessas fases, neste trabalho vamos apenas apresentar algumas recomendações gerais sobre distribuição de questões por conteúdo, habilidades específicas a serem testadas, tipos de itens de múltipla escolha, elaboração e revisão das questões, validade e fidedignidade, número de itens, e análise dos dados obtidos das provas aplicadas.

3 - DISTRIBUIÇÃO DE QUESTÕES POR CONTEÚDO

Não vamos fazer uma análise do conteúdo envolvido ou não pelas questões, da sua adequação aos objetivos da prova, de estar relacionado com o programa elaborado para a prova. No entanto, uma recomendação pode ficar registrada: o conteúdo das questões não deve envolver tópicos que não fazem parte do programa. Se houver um tópico ou assunto importante a ser avaliado que não conste no programa distribuído, convém primeiramente mudar o programa.

As questões da prova devem envolver todo ou a maior parte possível do conteúdo. Com isso se pretende garantir a inclusão de matéria representativa de todo o programa. Questões de múltipla escolha geralmente facilitam uma abordagem abrangente do conteúdo, permitindo ter uma amostra equilibrada e representativa dos assuntos do programa em cada prova. Quando não é possível envolver todos os tópicos do programa, os responsáveis pela elaboração da prova devem escolher aqueles que julgarem mais relevantes, significativos e úteis.

A determinação da facilidade ou da dificuldade das questões é revelada a partir das respostas dadas às questões, ou seja, depende dos examinandos. Embora o índice de dificuldade não necessariamente corresponda às expectativas do elaborador, é muito importante tentar estimar a dificuldade antecipadamente para planejar e elaborar uma prova que apresente questões bem distribuídas para o grupo de candidatos, ou seja, de dificuldade pequena, grande e média, geralmente com predominância das últimas.

Recomenda-se, então, realizar o planejamento de uma prova a partir dessas considerações sobre o conteúdo e a facilidade das questões. A organização de uma tabela de distribuição de itens (um modelo será apresentado mais adiante) é útil nessa tarefa. Nesse planejamento também deve-se pretender uma distribuição equilibrada das habilidades requeridas para a solução das questões.

4 - HABILIDADES ENVOLVIDAS NAS QUESTÕES

Cada questão de uma prova exige certas habilidades do domínio cognitivo, necessárias para respondê-la corretamente. A seguir apresentamos uma breve classificação, elaborada com base no trabalho de Bloom e outros (1973).

CONHECIMENTO

Recordação de termos, fatos, convenções, categorias, métodos e conceitos anteriormente aprendidos.

COMPREENSÃO

Percepção, transformação, interpretação, explicação e extrapolação de idéias, significados, conceitos e relações.

APLICAÇÃO

Uso de informações, idéias, conceitos, relações para resolver questões e problemas concretos.

ANÁLISE

Identificação e compreensão das partes, relações, idéias e princípios envolvidos na situação.

SÍNTESE

Agregação dos elementos envolvidos na situação para estabelecer uma conclusão ou produzir uma comunicação, um plano ou um resumo.

AVALIAÇÃO

Julgamento e atribuição de um valor a uma informação, texto, fato, idéia, método, trabalho, pesquisa, resultado ou conclusão.

Ao realizar o planejamento de uma prova, recomenda-se levar em consideração esta série de habilidades ou outra que seja mais familiar ao elaborador de questões. Elaboradores de questões de língua estrangeira talvez prefiram usar itens como interpretação de texto, gramática, tradução, versão, compreensão da fala, entre outros. É uma outra classificação de questões, embora cada uma delas também possa ser considerada como envolvendo uma ou mais das habilidades acima apresentadas.

Juntamente com a distribuição das questões por conteúdo, pode ser feita a classificação dessas questões por habilidades, numa tabela de distribuição de questões semelhante ao modelo que apresentaremos em seguida. O preenchimento dessa tabela normalmente é feita na seguinte seqüência: 1^ª) número total de questões da prova (célula inferior da direita), 2^ª) relação dos conteúdos (coluna da esquerda), 3^ª) habilidades exigidas (linha superior), 4^ª) número de questões por conteúdo (coluna da direita), 5^ª) número de questões por habilidade (linha inferior) e, finalmente, 6^ª) número de itens por conteúdo e habilidade envolvida (células internas).

Esse planejamento inicial tende a sofrer algumas alterações logo após o preenchimento da tabela e, posteriormente, ao serem elaboradas as questões. Essas alterações em geral são motivadas por algum esquecimento de conteúdo no planejamento inicial ou por uma melhor adequação de uma habilidade não prevista para uma questão sobre determinado assunto. Desde que não representem um número demasiadamente expressivo, de modo a alterar substancialmente o planejamento inicial, essas alterações não devem trazer preocupações para os elaboradores.

Os dados já preenchidos na tabela apresentada referem-se à prova de Física do concurso vestibular da Universidade Federal do Rio Grande do Sul (UFRGS) realizado em 1984.

CONCURSO VESTIBULAR DA UFRGS					
1984					
PROVA DE FÍSICA			Data de entrega: 15.10.83 Data da revisão: 22.10.83 às 9 hs		
TABELA DE DISTRIBUIÇÃO DE QUESTÕES					
ASSUNTO	HABILIDADE EXIGIDA				TOTAL
	Conheci- mento	Compre- ensão	Aplica- ção	Análise Síntese	
1. Grandezas e medidas			1		1
2. Cinemática	1	1	2	1	5
3. Forças e leis de Newton	2	1	1		4
4. Energia e quantidade de movimento	1		1	1	3
5. Estática dos fluidos		1		1	2
6. Calor e Termodinâmica	2	1	1	1	5
7. Ondas mecânicas e eletromagnéticas	2	2	1		5
8. Ótica	1	1		2	4
9. Eletricidade	1	2	2	1	6
10. Magnetismo e Eletromagnetismo	1	1		1	3
11. Física Moderna	1	1			2
12.					
13.					
14.					
TOTAL	12	11	9	8	40

5 - TIPOS DE QUESTÕES

1) AFIRMAÇÃO INCOMPLETA

Quando um trem percorre 20 km, subindo uma ferrovia com declividade de 0,5%, eleva-se aproximadamente

- (A) 10 m.
- (B) 50 m.
- (C) 100 m.
- (D) 400 m.
- (E) 1000 m.

2) INTERROGAÇÃO

Qual o autor da obra literária Vila dos Confins ?

- (A) Mário Palmério
- (B) Fernando Sabino
- (C) Josué Guimarães
- (D) Jorge Amado
- (E) João Guimarães Rosa

3) FALSO X VERDADEIRO

Analise cada uma das afirmações relacionadas com a estrutura atômica e indique se é verdadeira (V) ou falsa (F).

- () A energia de um elétron ligado ao átomo não pode assumir um valor qualquer.
- () O núcleo de um átomo é composto por prótons, nêutrons e elétrons.
- () Na maioria dos átomos o número de elétrons é igual à soma dos prótons e dos nêutrons.

Quais são, respectivamente, as indicações corretas ?

- (A) V - V - F
- (B) V - F - V
- (C) V - F - F
- (D) F - V - V
- (E) F - V - F

4) AFIRMAÇÕES CORRETAS

Analise as seguintes afirmações relacionadas com a reforma partidária brasileira de 1979.

I. O PDS foi a legenda que recebeu a filiação da maior parte das oposições da época.

II. A legenda do PTB era pretendida pelo grupo liderado por Leonel Brizola, mas acabou ficando com o de Ivete Vargas, por decisão do Poder Judiciário.

III. O Partido dos Trabalhadores (PT) foi articulado a partir do novo sindicalismo do ABC paulista.

Quais estão corretas ?

- (A) Apenas I
- (B) Apenas II
- (C) Apenas I e III
- (D) Apenas II e III
- (E) I, II e III

5) LACUNA

Selecione a alternativa que apresenta as palavras que completam corretamente as lacunas no seguinte texto:

O clima é encontrado em áreas situadas nas proximidades da linha do Equador, onde ocorre a maior intensidade solar que atinge a superfície da Terra. Em virtude disso, nessas regiões registram-se altos índices de evaporação e, conseqüentemente, de precipitação de chuva. Em nosso país, encontramos uma extensa área dominada por esse tipo de clima na região

- (A) tropical - Nordeste
- (B) tropical - Norte
- (C) temperado - Nordeste
- (D) equatorial - Norte
- (E) equatorial - Nordeste

Seleccione a alternativa que apresenta a palavra que completa corretamente a lacuna no seguinte texto.

No es posible realizar una reforma educacional con eficacia sin un personal docente plenamente competente. Así la formación de es preocupación capital de todas las organizaciones de personal docente.

- (A) ése
- (B) éste
- (C) aquel
- (D) éstos
- (E) ésos

6) ASSOCIAÇÃO

Associe cada obra (coluna da direita) com o seu respectivo autor (coluna da esquerda).

- | | | |
|-------------------------------|-----|----------------------|
| 1. Carlos Drummond de Andrade | () | Ai de ti, Copacabana |
| 2. Fernando Sabino | () | O Homem Nu |
| 3. Luís Fernando Veríssimo | () | Comédias da Vida |
| 4. Millôr Fernandes | | Privada |
| 5. Rubem Braga | | |

A seqüência dos números que estabelece as associações corretas na coluna da direita, quando lida de cima para baixo, é

- (A) 1 - 5 - 4
- (B) 2 - 1 - 3
- (C) 1 - 3 - 4
- (D) 5 - 2 - 3
- (E) 5 - 2 - 4

Associe o instrumento de medida (coluna da direita) com a grandeza física (coluna da esquerda) que pode ser medida diretamente com esse instrumento.

- | | | |
|-------------------------------------|-----|-------------|
| 1. intensidade de corrente elétrica | () | amperímetro |
| 2. potência elétrica | () | voltímetro |
| 3. resistência elétrica | | |
| 4. diferença de potencial elétrico | | |

A seqüência de números que estabelece as associações corretas na coluna da direita, quando lida de cima para baixo, é

- (A) 1 - 3 (B) 1 - 4 (C) 2 - 3 (D) 2 - 4 (E) 3 - 4

7) COMPARAÇÃO

Comparando ondas luminosas com sonoras, podemos verificar que ambas

- (A) são transversais.
(B) têm frequências da ordem de 10^{14} Hz.
(C) apresentam a mesma velocidade de propagação no ar.
(D) podem ser difratadas.
(E) podem ser polarizadas.

O quadro mostra comparações de alguns aspectos relacionados com dois países, A e B.

Aspectos	País A	País B
Hemisfério	Sul	Norte
Maior fronteira marítima	Atlântico	Pacífico
Clima	Temperdo e frio	Quente e temperado
Idioma oficial	Espanhol	Espanhol
Povos principais	Branco	Mestiços e índios
Religião principal	Católica	Católica

Os países A e B podem ser, respectivamente,

- (A) Argentina e México.
(B) Uruguai e Espanha.
(C) Uruguai e Austrália.
(D) Argentina e Espanha.
(E) Austrália e México.

8) INTERPRETAÇÃO / ANÁLISE / SÍNTESE

Leia atentamente o texto abaixo.

A modernização da mulher também refletia a alteração que se processava no modo de vida e, em decorrência, enfrentava severa oposição das camadas mais tradicionais, que se recusavam a aceitar o avanço do progresso, escudando-se nos problemas dele decorrentes. Assim como a sociedade tradicional era exaltada, a mulher educada à antiga era valorizada, pois a civilização tendia a masculinizar a mulher. Se

a mulher das grandes cidades se deixava atrair por hábitos masculinos que lhe manchavam a candura, como freqüentemente aparece nos romances da época, a campestre era apontada como o tipo ideal feminino porque não estava ainda conspurcada pelos costumes ditos modernos.

Dos títulos propostos nas alternativas, qual o que melhor sintetiza o conteúdo do texto ?

- (A) Como reagiram os conservadores à modernização da mulher
- (B) Por que as mulheres se masculinizaram e se conspurcaram
- (C) As razões do fracasso do feminismo diante do machismo
- (D) A vitória da tradição sobre o progresso da mulher
- (E) A modernidade desconsidera os riscos sócio-culturais

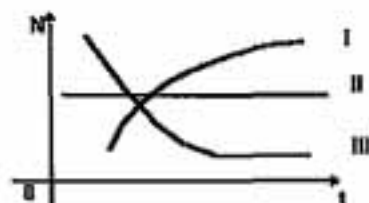
Considere o seguinte texto:

Certos raios luminosos, como os do Sol, ao se transmitirem de um meio transparente para outro, podem dar origem a vários raios refratados de cores diferentes, cada um com uma inclinação própria. Muitas vezes esse fenômeno não é percebido pelo olho humano porque esses raios, estando muito próximos uns dos outros ou superpostos, dão a sensação de uma só cor, geralmente branca. Entretanto, para certos meios, com uma incidência adequada da luz, esse fenômeno de separação de cores é perfeitamente visível.

O fenômeno abordado no texto acima é conhecido como

- (A) reflexão.
- (B) difração.
- (C) dispersão.
- (D) interferência.
- (E) polarização.

O gráfico mostra o número de indivíduos (N) em função do tempo (t) de três diferentes populações (I, II e III) de um mesmo habitat.



Analisando as curvas de crescimento dessas três populações, é possível supor que

- (A) I é predador de III.
- (B) III é predador de II.
- (C) I é presa de II.
- (D) I e II são presas de III.
- (E) I e III são presas de II.

9) CORREÇÃO

O período apresenta cinco segmentos grifados, um dos quais contém um erro. Assinale a alternativa correspondente ao segmento *incorreto*.

Rice came originally from wild plants found in Asia. Some

- (A) archeologists believe that it was first cultivate at least
- (B) seven centuries ago.
- (C)
- (D)
- (E)

O período apresenta cinco segmentos grifados, um dos quais contém um erro. Assinale a alternativa correspondente ao segmento *incorreto*.

Há várias maneiras de ele procurar um novo emprego. Uma

- (A) delas é encaminhar o seu currículo para firmas que trabalham
- (B) no setor de informática. Porque não fazê-lo ?
- (C)
- (D)
- (E)

10) NEGATIVO

Entre as amostras de matéria apresentadas nas alternativas, qual não é uma mistura ?

- (A) Ar
- (B) Leite
- (C) Ozônio
- (D) Água do mar
- (E) Gás de cozinha

11) ASSERÇÃO E RAZÃO

Considere o seguinte período, formado por duas afirmações, denominadas *asserção* e *razão*.

Asserção

Numa mesma latitude, a temperatura em regiões próximas do nível do mar é maior do que nas montanhas de grande altitude

porque

Razão

a intensidade solar diminui à medida que a distância ao Sol aumenta.

Nesse caso,

- (A) a asserção e a razão são verdadeiras, e a razão é uma justificativa da asserção.
- (B) a asserção e a razão são verdadeiras, e a razão **não** é uma justificativa da asserção.
- (C) a asserção é verdadeira e a razão, falsa.
- (D) a asserção é falsa e a razão, verdadeira.
- (E) a asserção e a razão são falsas.

12) SUBSTITUIÇÃO

Considere o seguinte período, corretamente redigido:

Não tenho como provar que estão errados, embora esteja convicto disso.

Sem alterar a idéia contida nesse período, construa um novo período iniciando com:

Estou convicto...

Qual o elemento que melhor se encaixa no novo período ?

- (A) por isso
- (B) quando
- (C) mas
- (D) então
- (E) porque

13) TRADUÇÃO

Na frase, *An empirical research was designed to investigate concept learning*, a melhor tradução para *was designed* é

- (A) foi designada
- (B) estava designada
- (C) foi desenhada
- (D) estava desenhada
- (E) foi planejada

Selecione a alternativa que apresenta a melhor tradução da frase *Il me faut acheter un cadeau pour ma nièce.*

- (A) Ele me faz comprar um presente para minha neta.
- (B) Preciso comprar um presente para minha sobrinha.
- (C) Falta-me comprar um presente para minha neta.
- (D) Ele me fala em comprar um presente para minha sobrinha.
- (E) É necessário que eu compre um presente para minha prima.

14) VERSÃO

Selecione a alternativa que apresenta a melhor versão alemã da frase *Ele perdeu o trem das cinco horas por minha causa.*

- (A) Er hat meinetwegen die Bahn um 5 verloren.
- (B) Er versäumte aus meinem Grund die Eisenbahn um 5.
- (C) Er hat wegen mir den 5 Uhr Zug verpasst.
- (D) Er hatte den Zug um 5 wegen meiner Sache verpasst.
- (E) Er verlor die 5 Uhr Bahn auf meine Veranlassung.

6 - RECOMENDAÇÕES PARA A ELABORAÇÃO DE QUESTÕES DE MÚLTIPLA ESCOLHA

1. Definir os objetivos da prova e identificar as habilidades necessárias para resolver as questões.

2. Distribuir um número adequado de questões para os diversos tópicos do programa, visando a uma avaliação ampla do conhecimento dos candidatos sobre a matéria.

3. Usar informações, dados e conhecimentos sobre os quais, de maneira geral, não haja controvérsia na comunidade científica especializada.

4. Elaborar questões que em conjunto apresentem uma dificuldade média para os candidatos, todas adequadas aos objetivos e aos conteúdos envolvidos.

5. Ordenar as questões de acordo com uma seqüência lógica, agrupando-as por conteúdo ou tipo de questão.

6. Na elaboração das questões, observar o que segue:

- a) O tópico escolhido deve ser realmente importante e fazer parte do programa.

• b) O enunciado de cada questão deve abordar uma situação ou problema completo e bem definido, que permita uma interpretação clara e inequívoca.

• c) Tudo que não for essencial ao enunciado deve ser eliminado, evitando-se que a leitura se prolongue desnecessariamente.

• d) Elaborar enunciados e perguntas na forma afirmativa. Evitar as formulações negativas. Se usadas, devem ser destacadas.

• e) Quando um enunciado, um texto, uma figura, um gráfico ou uma tabela for comum a mais de uma questão, isso deve ficar claro por meio de uma formulação inicial.

• f) Figuras, mapas, gráficos devem conter apenas os elementos essenciais para a questão. Também devem ser apresentados na forma definitiva, com adaptações e reduções necessárias ao caso.

• g) Sempre que o tipo de questão repetir algum modelo já consagrado, padronizar ordens e instruções de acordo com as formulações usadas em provas anteriores.

• h) Nas questões do tipo *associação* (coluna da esquerda com coluna da direita) o número de itens em cada coluna deve ser diferente para evitar o acerto por exclusão, ou seja, que o candidato acerte a questão conhecendo poucas ou apenas uma associação correta. Pares de números de itens recomendados: 4 e 2, 5 e 3, 6 e 2. O número de vezes que cada item aparece nas alternativas deve ser equilibrado para evitar o acerto a partir das indicações mais frequentes.

• i) Nas questões do tipo *afirmações corretas e falso x verdadeiro* evitar um número excessivo de afirmações, que geralmente facilitam obter a resposta correta conhecendo poucas ou apenas uma afirmação correta. Recomenda-se 3 afirmações por questão. Além disso, o número de vezes que cada afirmação aparece indicada como correta ou verdadeira deve ser equilibrado para evitar o acerto a partir das indicações mais frequentes.

• j) As questões devem ser independentes uma das outras, isto é, respostas ou pistas para a solução de uma questão não devem aparecer em outra.

• k) Os elementos iguais das alternativas devem ser colocados no enunciado.

• l) Deve haver uma única alternativa com a resposta correta.

- m) As alternativas incorretas devem ser relativamente atraentes e plausíveis para candidatos sem o conhecimento essencial sobre o assunto.
- n) Pistas devem ser evitadas. Por exemplo: alternativa correta mais detalhada, alternativas erradas com termos absolutos (sempre, nunca, todos, nenhum), alternativas sinônimas, alternativas que excluem outras.
- o) Evitar formulação ou dica que induza o candidato ao erro, ou dificulte intencionalmente a solução da questão.
- p) Todas as alternativas devem completar corretamente o enunciado, a pergunta.
- q) Apresentar as alternativas em ordem lógica. Por exemplo: seqüência cronológica, numérica.
- r) Analisar e revisar as questões da prova em conjunto para fazer uma avaliação global, evitar superposições e estimar a dificuldade.
- s) Restando alguma dúvida sobre a questão, reformule-a ou, permanecendo a dúvida, substitua-a por outra.

7 - ELABORAÇÃO E REVISÃO DAS QUESTÕES

Para a redação da prova é recomendada a participação de, pelo menos, dois elaboradores, um revisor de conteúdo, um revisor de técnicas de elaboração de itens e um revisor de linguagem. Eventualmente, as duas últimas tarefas podem ser executadas por um único especialista.

Os elaboradores devem se reunir para planejar o número de questões e a sua distribuição por conteúdo e habilidades. Para esse planejamento sugere-se a preparação de uma tabela de distribuição de questões por assunto e habilidades, conforme anteriormente proposto.

Na fase da elaboração convém seguir uma série de recomendações, entre as quais destacamos as apresentadas em seção anterior. Para os revisores se orientarem nas suas tarefas são feitas as mesmas recomendações.

Elaboradas as questões e feitas as revisões, cabe realizar uma reunião entre todos os elementos envolvidos nessas tarefas para a discussão do conteúdo, das dificuldades, dos problemas técnicos e da linguagem de cada questão. Em decorrência disso, normalmente

ocorrem modificações de redação e substituição de questões visando ao aprimoramento da prova. Como resultado, no final dessa reunião, deve-se pretender, de comum acordo, chegar à redação definitiva das questões, ou seja, à versão final da prova que deve seguir para a datilografia.

8 - ANÁLISE SOBRE VALIDADE, FIDEDIGNIDADE E NÚMERO DE QUESTÕES

As provas geralmente são usadas para fazer julgamentos sobre o desempenho dos indivíduos que as realizam. O valor desse julgamento se baseia na confiança que se tem sobre os escores obtidos na realização de uma prova. Essa confiança aparece quando a prova é fidedigna e válida.

Os resultados ou escores obtidos da aplicação de uma prova estão sujeitos a erros. Alguns deles, como os grosseiros (enganos) e os sistemáticos, podem ser eliminados à medida que forem detectados. No entanto, os erros acidentais ou aleatórios acompanham qualquer mensuração. Eles podem ser diminuídos mas não totalmente eliminados, ou seja, temos que conviver com eles. Quanto menor esse erro, maior a precisão ou fidedignidade da prova. Portanto, o importante é minimizar a influência do erro sobre os escores de uma prova.

Vamos fazer alguns comentários sobre a validade, a fidedignidade e o número de questões de uma prova.

VALIDADE

Distinguem-se quatro diferentes tipos de validade: conteúdo, concorrente, preditiva e construto (Vianna, 1982). A análise da validade de conteúdo é a que mais interessa ao presente trabalho.

Considera-se que uma prova tem validade de conteúdo quando constitui uma amostra representativa de conhecimentos envolvidos no conteúdo de uma certa área que se pretende avaliar. Como esses conteúdos em geral estão listados num programa, uma prova com validade deve avaliar uma amostra representativa dos conhecimentos envolvidos nos tópicos desse programa. Há programas que apresentam, além dos conteúdos, objetivos educacionais. Nesse caso, a validade é determinada pela existência de questões em número

suficiente para se ter uma avaliação de todos os conteúdos e objetivos, ou de uma quantidade representativa deles.

A validade de conteúdo não é determinada estatisticamente, mas é o resultado de um julgamento de examinadores ou especialistas sobre a representatividade das questões em relação aos conteúdos e/ou objetivos a medir.

Normalmente deve-se tomar precauções em relação à validade já durante o planejamento da prova, ao elaborar a distribuição do número de questões a partir dos conteúdos e dos objetivos. Se os tópicos que aparecem no programa são importantes, devem ser avaliados e, caso não o sejam, retirados do programa.

FIDEDIGNIDADE

A fidedignidade de uma prova está relacionada com a estabilidade e consistência interna dos seus resultados. Uma prova fidedigna é uma prova precisa, com os erros decorrentes do instrumento de medida (prova) e da variabilidade dos examinandos minimizados, ou seja, pequenos.

A estatística permite estabelecer relações para o cálculo do coeficiente de fidedignidade. Quando os escores resultam de uma única aplicação da prova, as fórmulas KR_{20} e KR_{21} desenvolvidas por Kuder e Richardson permitem determinar a fidedignidade por meio de um coeficiente de consistência interna (Vianna, 1982). Embora o coeficiente KR_{21} seja sempre uma subestimativa do KR_{20} , vamos apresentar apenas a fórmula para calcular o primeiro, que envolve a média dos escores.

A sua expressão é

$$KR_{21} = \frac{k}{k-1} \left[1 - \frac{\bar{X}(1 - \bar{X}/k)}{s^2} \right]$$

onde,

KR_{21} : coeficiente de fidedignidade

k : número de itens da prova

\bar{X} : média dos escores do grupo

s : desvio padrão dos escores

s^2 : variância dos escores

Para uma prova já aplicada, com

$$k = 200$$

$$\bar{x} = 105,83$$

$$s = 16,73$$

temos, substituindo esses valores na relação acima,

$$KR_{21} = 0,83$$

Esse valor alto do coeficiente de fidedignidade é muito bom para os propósitos da prova. Para a maioria dos autores e especialistas um coeficiente mínimo de 0,70 é considerado aceitável para fins de decisão.

Há fatores relativos à prova e aos examinandos que podem afetar a fidedignidade.

Entre os fatores relativos ao examinando podem ser mencionados a motivação, o cansaço, a fraude e perturbações (barulho, perguntas, acidentes, interferência imprópria) durante a realização da prova.

Entre os fatores relativos à prova vamos destacar os seguintes:

- Quanto maior o número de itens, tanto maior a fidedignidade.
- Quanto mais objetiva a correção, tanto maior a fidedignidade.
- Quanto mais homogêneo o conteúdo dos itens, tanto maior a fidedignidade.
- Questões extremamente fáceis ou difíceis não contribuem para a fidedignidade.
- Enunciados duvidosos ou demasiadamente longos, palavras imprecisas ou desconhecidas, frases com estruturas defeituosas, instruções inadequadas, apresentação inadequada, provas muito longas, são alguns elementos que diminuem a fidedignidade da prova porque podem levar os examinandos a falsas interpretações.

O NÚMERO DE QUESTÕES

Quando se define o número de questões de uma prova, deve-se levar em conta, principalmente, a validade e a fidedignidade.

Um número de 200 questões por prova, à primeira vista, pode parecer alto, enquanto que 20 pode ser considerado baixo. No entanto, antes de qualquer julgamento, é preciso fazer uma análise da validade e da fidedignidade.

NÚMERO DE QUESTÕES X VALIDADE

Como já mencionamos anteriormente, uma prova apresenta validade quando as questões avaliam todos ou uma amostra representativa dos conteúdos e/ou objetivos educacionais programados. Portanto, para realizar um julgamento sobre a validade da prova em questão, é necessário que especialistas da matéria relacionem as questões com as várias partes do conteúdo para então verificar se há um número suficiente ou representativo de questões para cada tipo de conteúdo. A partir dessa análise pode chegar a uma conclusão do tipo *considerando a necessidade de ter-se uma prova que apresente validade, 50 é um número de questões adequado para essa prova.*

O número pode ser diminuído se houver várias questões sobre um certo tópico do conteúdo, ou seja, uma avaliação repetitiva. Por outro lado, se houver partes do conteúdo que não estão sendo avaliadas pelas questões, deve-se pensar em incluir ou redistribuir questões para também avaliar essas partes. Nesses dois casos, bem como no caso em que o número de questões for julgado adequado, o importante é manter um número representativo de questões sobre as várias partes do conteúdo programado.

NÚMERO DE QUESTÕES X FIDEDIGNIDADE

Foi mencionado anteriormente que *quanto maior o número de questões, tanto maior a fidedignidade.* De imediato, isso pode parecer que sempre é desejável ter um elevado número de questões para ter uma maior fidedignidade. No entanto, um estudo mais detalhado mostra que o crescimento da fidedignidade com o número de questões não é linear, mas é maior quando o número de questões é pequeno, tendendo a estabilizar à medida que esse número aumenta, ou seja, quando se atinge um alto coeficiente de fidedignidade, o aumento do número de questões muda muito pouco esse índice. Por outro lado, há fatores como *provas muito longas e cansaço* que contribuem negativamente para a fidedignidade. Sabe-se também que a concentração e o rendimento do examinando começa a diminuir depois de cerca de duas horas em uma mesma atividade.

Diante dessas e de outras considerações, recomenda-se procurar um equilíbrio. Uma vez atingida uma fidedignidade mínima aceitável

(além da existência de validade), não convém aumentar mais o número de questões por prova.

Partindo de alguns pressupostos, foi desenvolvida uma fórmula (ver Vianna, 1982) que permite determinar a fidedignidade esperada de uma prova:

$$KR_{\text{esp}} = \frac{k}{k-1} \left[1 - \frac{9(N+1)}{k(N-1)} \right]$$

onde,

KR_{esp} : coeficiente de fidedignidade esperado

k: número de itens da prova

N: número de alternativas das questões

Aplicando-se essa fórmula para uma prova com cinco alternativas ($N = 5$) em cada questão, temos para k igual a 20, 30, 40, 50, 100, 120, 140, ou 160 questões, os coeficientes apresentados na tabela abaixo.

Número de itens	Fidedignidade esperada
20	0,34
30	0,57
40	0,68
50	0,74
100	0,87
120	0,89
140	0,91
160	0,92

A partir desses dados podemos estabelecer algumas possibilidades. Por exemplo, se esperamos um coeficiente de fidedignidade de 0,85, uma prova com 100 questões pode ser planejada. Se formos um pouco mais exigentes, esperando uma fidedignidade de 0,90, então 140 questões são suficientes. Em ambos os casos, deve ser verificado se a validade de conteúdo fica atendida.

Digamos que 120 questões atendem nossas expectativas sobre a fidedignidade e validade. Estimando uma média de 3 minutos por

questão, podemos programar a prova com 6 horas de duração, recomendando-se, neste caso, dois períodos, cada um de 3 horas. Para um outro caso, em que conseguirmos avaliar uma parte representativa do conteúdo com 40 questões, a fidedignidade (cerca de 0,70) ainda é aceitável, e podemos tranquilamente programar a prova para um único período de 2 horas.

9 - ANÁLISE DE UMA PROVA

Após a aplicação de uma prova convém analisar dados como média, desvio padrão e fidedignidade da prova, bem como o índice de discriminação e a facilidade de cada questão. Esses dados permitem a identificação de problemas educacionais, particularmente os relacionados com a área de aprendizagem. Além disso, constituem indicadores valiosos para o aprimoramento da avaliação, considerando que possibilitam detectar deficiências técnicas na elaboração das questões da prova.

Nessas análises a média e o desvio padrão dos escores de uma prova são informações úteis sobre a dificuldade geral da prova e a dispersão dos escores obtidos pelos candidatos, respectivamente. A precisão ou fidedignidade pode ser estimada por coeficientes obtidos de relações como as de Kuder e Richardson: KR_{21} (anteriormente apresentada) e KR_{20} (Vianna, 1982).

Já a facilidade de uma questão pode ser expressa pela proporção P de candidatos que optaram pela alternativa correta. Um índice alto indica que a questão foi muito fácil, ao passo que um índice baixo revela uma dificuldade elevada. Nem sempre a estimativa feita pelos elaboradores sobre o grau de dificuldade de uma questão corresponde à sua dificuldade real para os candidatos, determinada após a aplicação da prova.

O estabelecimento da discriminação de uma questão ocorre com base na comparação entre os acertos do grupo superior com os do grupo inferior na questão. Para as análises que serão apresentadas neste trabalho, considera-se como grupo superior (Sup) o formado por 27% dos examinandos que alcançaram os escores totais mais elevados na prova. O grupo inferior (Inf) é composto por 27% dos candidatos com escores totais inferiores. Os restantes 46% constituem o grupo de desempenho médio (Med). Pode-se utilizar, como será feito mais adiante, o coeficiente ϕ (ϕ) de discriminação, definido como

$$\phi = \frac{P_s - P_i}{2 \sqrt{pq}}$$

onde,

P_s : proporção do grupo superior que acertou a questão,

P_i : proporção do grupo inferior que acertou a questão,

pq : variância da questão, sendo p a proporção do grupo total que acertou a questão e $q = 1 - p$.

Com os recursos da informática, hoje em dia disponíveis não apenas em grandes centros de processamento de dados, a correção da prova e o processamento de dados se tornou uma tarefa fácil e acessível para a maioria dos casos em que se pretende fazer uma análise dos resultados de uma prova.

ALGUNS EXEMPLOS

A partir de um levantamento de dados relacionados com as provas de Física de cinco dos últimos concursos vestibulares organizados pela Universidade Federal do Rio Grande do Sul, foi possível construir a tabela 1. Esses dados mostram que as médias têm sido relativamente baixas (entre 31 e 36% de acertos) e uniformes ao longo desses anos. Já o coeficiente de fidedignidade foi constante e acima de 0,70, um valor que é desejável porque revela uma boa precisão das provas.

TABELA 1
DADOS SOBRE AS PROVAS DE FÍSICA DO CONCURSO
VESTIBULAR DA UFR

k : número de questões, \bar{X} : média, s : desvio padrão,

KR_{21} : coeficiente de fidedignidade, n : número de examinandos

Ano	k	\bar{X}	s	KR_{21}	n
1991	35	12,527 (35,79%)	5,361	0,74	26069
1992	35	12,688 (36,25%)	5,339	0,74	24634
1993	35	12,455 (35,59%)	5,345	0,74	26733
1994	35	11,849 (33,85%)	5,316	0,74	26982
1995	35	11,160 (31,89%)	5,175	0,74	28733

Para ilustrar a análise que pode ser feita sobre cada questão de uma prova, vamos apresentar dados, conclusões e comentários sobre três questões de provas de concursos realizados pela UFRGS.

Questão de número 14 da prova de Geografia (1991):

Kuwait e Iraque situam-se

(A) no Oriente Médio.

(B) no Extremo Oriente.

(C) no Leste Europeu.

(D) no Norte da África.

(E) na Região dos Bálcãs.

Os dados obtidos sobre esta questão estão na tabela 2.

TABELA 2
DISTRIBUIÇÃO DE RESPOSTAS DA QUESTÃO 14 DA PROVA DE
GEOGRAFIA DO CONCURSO VESTIBULAR DA UFRGS (1991)

Alternativa	Sup (%)	Med (%)	Inf (%)	P	ϕ
A*	96,88	91,16	80,33	0,90	0,28
B	1,57	4,23	8,11		
C	0,59	2,35	6,68		
D	0,52	1,37	2,87		
E	0,44	0,86	1,97		

Como 90% dos candidatos acertaram a questão, ela pode ser considerada muito fácil. Isso revela que a grande maioria (inclusive o grupo inferior) conhecia muito bem o assunto no nível que foi avaliado pela questão. Esse fato deixou o número de acertos do grupo inferior (80,33%) pouco distante do número de acertos do grupo superior (96,88%), ou seja, eles não ficaram muito discriminados nesta questão, o que é revelado pelo índice de discriminação ($\phi = 0,28$) relativamente baixo.

Questão de número 17 da prova de Literatura Brasileira (1992):
Considere as seguintes afirmações.

I - O romance Canaã, de Graça Aranha, tem como personagens centrais Lentz e Milkau, dois imigrantes que discutem ao longo do texto suas teses antagônicas sobre os objetivos e as perspectivas da imigração no Brasil.

II - Os romances de Lima Barreto Recordações do Escrivão Isafas Caminha e Clara dos Anjos seguem a trilha de escritores como

Manuel Antônio de Almeida na sua proposta de retratar as classes suburbanas e desprivilegiadas do Rio de Janeiro.

III - A obra de Monteiro Lobato, sob o ponto de vista temático, focaliza a decadência econômica e social da região produtora de café no interior paulista, e, do ponto de vista estrutural e lingüístico, traz inovações que servirão de modelo aos modernistas de 22.

Quais estão corretas ?

- (A) Apenas I
- (B) Apenas II
- (C) Apenas I e II
- (D) Apenas I e III
- (E) I, II e III

Os dados obtidos sobre esta questão estão na tabela 3.

TABELA 3
DISTRIBUIÇÃO DE RESPOSTAS DA QUESTÃO 17
DA PROVA DE LITERATURA BRASILEIRA DO
CONCURSO VESTIBULAR DA UFRGS (1992)

Alternativa	Sup (%)	Med (%)	Inf (%)	P	ϕ
A	7,74	10,11	14,10		
B	4,33	9,39	16,31		
C*	32,95	26,22	22,73	0,27	0,12
D	16,98	24,83	27,81		
E	38,00	29,42	19,00		

Os dados mostram que foi uma questão muito difícil para os candidatos: apenas 27% acertaram a questão. Além disso, é possível verificar que a distribuição das respostas nas cinco alternativas não apresenta grandes diferenças, particularmente para o grupo inferior, onde a escolha de cada alternativa não ficou longe de 20%, que corresponde ao acerto casual. Ou seja, esses candidatos não souberam resolver a questão e, então, provavelmente assinalaram uma alternativa de forma aleatória. Quando uma questão é difícil para todos os candidatos, como neste caso, a diferença entre o número de acertos dos grupos superior e inferior tende a ser pequena. Em virtude disso, a questão não foi adequada para discriminar bem os candidatos, o que é revelado pelo seu pequeno índice de discriminação $\phi = 0,12$.

Questão de número 69 da prova de Física (1991):

Assinale a alternativa que apresenta as palavras que preenchem corretamente as duas lacunas no texto abaixo, respectivamente.

Em 1989 os noticiários destacaram por um certo período a realização de pesquisas sobre maneiras alternativas de obter a fusão nuclear. Tais alternativas, contudo, não se confirmaram. O que se sabe comprovadamente hoje é o que já se sabia até aquela época: a fusão nuclear é obtida a temperaturas tão altas quanto as existentes e, ao contrário da fissão nuclear utilizada nas centrais nucleares, dejetos radioativos.

- (A) na superfície da Terra - produz
- (B) na superfície da Lua - produz
- (C) na superfície da Lua - não produz
- (D) no centro do Sol - não produz
- (E) no centro do Sol - produz

A tabela 4 apresenta os dados que foram obtidos das respostas dadas pelos estudantes.

TABELA 4
DISTRIBUIÇÃO DE RESPOSTAS DA QUESTÃO 69
DA PROVA DE FÍSICA DO CONCURSO
VESTIBULAR DA UFRGS (1991)

Alternativa	Sup (%)	Med (%)	Inf (%)	P	ϕ
A	3,81	12,75	24,48		
B	0,82	3,31	7,42		
C	2,20	4,64	7,83		
D*	70,76	48,14	27,32	0,49	0,43
E	22,39	31,04	32,89		

Analisando esses dados, pode-se verificar que a questão apresentou uma dificuldade média para os examinandos (49% acertaram a questão) e uma boa discriminação ($\phi = 0,43$), satisfazendo, com isso, uma das finalidades de provas classificatórias. A análise ainda permite observar que as alternativas B e C foram pouco atraentes, ao contrário da alternativa E. Isso deixa evidente que uma parte considerável dos candidatos entendeu que a fusão nuclear ocorre a altíssimas temperaturas (corretamente) e produz dejetos radioativos (erradamente).

Há alguns textos destinados especificamente para relatos de análises de uma prova ou conjunto de provas semelhantes às apresentadas acima e até mais detalhadas. Buchweitz et al. (1984) apresentam uma análise a partir das respostas de cerca de 20 mil candidatos que participaram do Concurso Vestibular de 1984, promovido pela Universidade Federal do Rio Grande do Sul. Nesse trabalho são apresentados os dados relacionados com cada uma das matérias que compuseram o elenco de provas desse concurso. Além disso, é feita uma análise geral dos dados de cada prova e uma análise específica de algumas questões de cada prova.

Em outro trabalho (Chassot et al., 1976), quatro professores apresentam um trabalho detalhado que realizaram sobre as questões de uma prova de Química e de uma de Física. Nesse trabalho foi feita uma análise do desempenho dos estudantes a partir dos dados obtidos das respostas dos estudantes em 40 questões de múltipla escolha de cada prova, semelhante à descrita acima. Além dessa análise, os professores apresentaram a solução de cada questão, acompanhada de comentários que julgaram pertinentes.

Axt (1986) aproveita os dados obtidos das respostas dos candidatos em questões de Física de vários concursos vestibulares da UFRGS para detectar conceitos intuitivos relacionados com as leis de movimento de Newton. O autor destaca que os dados sugerem a existência de uma estrutura conceitual espontânea, na forma de "leis", as quais frequentemente prevalecem sobre o formalismo da dinâmica, quando a compreensão deste formalismo é solicitada, como nos casos das questões selecionadas no trabalho.

Em outro artigo, Buchweitz e Silveira (1983) analisam os dados obtidos das respostas dos estudantes às questões de Física, que foram classificadas em quatro áreas de conteúdo: Mecânica, Termodinâmica, Eletricidade-Magnetismo e Ondas. A análise estatística desses dados, obtidos de nove vestibulares seguidos, revelou que as médias dos estudantes nessas quatro áreas não foram significativamente diferentes. Esse desempenho equivalente nas quatro áreas se verificou para o grupo total de vestibulandos e para os subgrupos de desempenho superior, médio e inferior. Considerando que os escores obtidos nas provas revelam o grau de conhecimento (ou aprendizagem) dos vestibulandos, foi possível concluir que este não é significativamente diferente nas quatro áreas de conteúdo escolhidas.

Os exemplos de análises de provas aqui apresentados reforçam a importância, anteriormente destacada, dessas análises na identificação de problemas educacionais relacionados com a aprendizagem e na reformulação e melhoria do currículo, do ensino e da própria avaliação.

10 - COMENTÁRIOS E CONCLUSÕES

Neste trabalho procuramos trazer um pouco da nossa experiência sobre planejamento, elaboração e análise de questões de múltipla escolha. Essa experiência foi adquirida ao longo de mais de duas décadas, durante as quais periodicamente dedicamos algum tempo para a realização desses trabalhos, juntamente com outros colegas.

Queremos deixar claro que não estamos manifestando uma preferência pela utilização de questões de múltipla escolha em detrimento de questões discursivas. São dois tipos de questões amplamente conhecidos e utilizados, cada um com suas vantagens e desvantagens. Na sua comparação podem ser considerados aspectos como dificuldades ou facilidades na sua elaboração e na correção das respostas, a presença ou ausência de subjetividade na correção, a avaliação mais ou menos representativa dos conteúdos do programa, a maior ou menor liberdade e criatividade em formular a própria resposta, entre outros. Entendemos que os resultados da avaliação dos conhecimentos de um mesmo grupo, sobre o mesmo conteúdo, nas mesmas condições, feita por meio de dois tipos de instrumentos de avaliação diferentes, devem ser os mesmos, como já tivemos oportunidade de constatar (Buchweitz, 1976). Caso esses resultados sejam significativamente diferentes, provavelmente há algum problema com um, ou os dois instrumentos. Esses problemas geralmente são técnicos, decorrentes de questões mal elaboradas ou sem validade de conteúdo, aplicação inadequada, ou correção deficiente.

Mitos e preconceitos existentes sobre certos tipos de questões podem muitas vezes ter sido causados por problemas dessa natureza ou por falta de esclarecimentos ou oportunidades de melhor conhecer o assunto. Daí surge a importância de ter-se professores bem preparados para as tarefas de planejamento, elaboração, correção e análise de questões. Artigos, textos, relatos de experiências e cursos sobre o assunto certamente podem contribuir para essa preparação.

Referências Bibliográficas

- AXT, R. Conceitos intuitivos em questões objetivas aplicadas no concurso vestibular unificado da Universidade Federal do Rio Grande do Sul. *Ciência e Cultura* **38**(3): 444-452, 1986.
- BLOOM, B. S. e outros. **Taxionomia de objetivos educacionais e domínio cognitivo**. Porto Alegre, Globo, 1973.
- BUCHWEITZ, B. Testes de múltipla escolha e resposta livre em Física Geral. *Cadernos de Pesquisa*, **16**: 3-6, 1976.
- BUCHWEITZ, B., MOTTA, M. H. S. E LEWIN, Z. G. **Análise de Itens das Provas do Concurso Vestibular Unificado da UFRGS**. Porto Alegre: Editora da Universidade, 1984.
- BUCHWEITZ, B., e SILVEIRA, F.L. Comparação de desempenhos dos vestibulandos em quatro áreas de conteúdo de Física. *Educação e Seleção*, **7**: 43-48, 1983.
- CHASSOT, A.I., SCHIFINO, J., BUCHWEITZ, B. e AXT, R. **Análise de desempenhos em provas de Química e Física**. São Paulo: Fundação Carlos Chagas, 1976.
- VIANNA, H. M. **Testes em Educação**. 4^a ed. São Paulo: IBRASA, 1982.