CO32
VALIDADE EM TESTES E QUESTIONÁRIOS

VERSÃO PRELIMINAR

Maurício Urban Kleinke

kleinke@ifi.unicamp.br

http://sites.ifi.unicamp.br/kleinke/snef/

XXII SNEF SÃO CARLOS | SP
SIMPÓSIO NACIONAL DE ENSINO DE FÍSICA
de 23 a 27 de Janeiro de 2017

Sumário

Sumário 0
1. Medidas em Física 1
1.1. Conceitos básicos sobre medidas 1
1.2. Confiabilidade de uma medida 2
1.3. Exercício #1 4
1.4. Experimento sobre calorimetria 5
1.5. Exercício #2 5
1.6. Exercício #3 7
1.7. Sugestões de Leitura 7
2. Instrumentos de Avaliação 8
2.1. Constructo e variáveis latentes 9
2.1.1. Radiação solar e blocos 10
2.1.2. Aprendizado de mecânica e satisfação do aluno 11
2.2. Exercício #4 12
2.3. Alfa de Cronbach 12
2.4. Estimativa de Spearman Brown 13
2.5. Planilha Araraquara/São Carlos 13
2.6. Exercício #5 14
2.7. Exercício #6 14
2.8. Sugestões de Leitura 15
3. Como Elaborar Questionários 16
3.1. Introdução 16
3.2. Definições Gerais 16
3.3. Dicas para desenhar um bom questionário 17
3.4. Sete Regras de Ouro para Desenhar Questionários 18
3.5. Exercício #7 20
3.6. Sugestões de Leitura 20
4. Como Elaborar Testes de Múltipla Escolha 21
4.1. Algumas observações sobre elaboração de itens 21
4.2. Preparação para Escrever Itens 23
4.2.1. Crie itens de teste enquanto prepara as suas aulas: 23
4.2.2. Vantagens: 23
4.2.3. Limitações: 23
4.2.4. Dicas: 24
4.3. Exercício #8 25
4.4. Sugestões de Leitura 27

"Todas as medidas ... são medidas sociais.
As medidas em física são realizadas com objetivos sociais"

Notes on Social Measurement
O. D. Duncam

1. Medidas em Física

1.1. Conceitos básicos sobre medidas

A formação típica de um físico tem como um de seus eixos a análise e a compreensão das medidas experimentais. As tentativas de laboratórios básicos para formação de alunos com visão de ciência experimental em geral apresentam dificuldades para professores e alunos, apesar do esforço desprendido por ambos os grupos.

Para introduzir as questões sobre medidas, vamos recapitular alguns conceitos básicos, para os quais vou utilizar uma apostila que está disponível na rede para consulta[1]. Vamos rapidamente procurar o que nos interessa, que é a medida do valor correto ou valor verdadeiro de uma variável. Nem todos os conceitos estão referenciados na apostila, alguns são construções necessárias para compreender nossas necessidades.

Toda a medida (vamos definir como X) está associada a dois valores, o valor verdadeiro (V) e o erro (e) associado à medida X

$$X = V + e \qquad (1)$$

[1] Guia para Física Experimental Caderno de Laboratório, Gráficos e Erros Instituto de Física, Unicamp.

A estimativa de um valor mais próximo do verdadeiro em uma variável costuma estar associada ao valor médio dessa variável obtido em uma série de medidas

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N} \qquad (2)$$

onde X_i representa a iésima medida de um conjunto total de N medidas.

Para obter informação sobre a variação das medidas em torno da média, se utiliza um indicador que independa do erro associado estar acima ou abaixo do valor médio, então par isso se utiliza a variância (SS)

$$SS = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N - 1} \qquad (3)$$

A qual também define o valor do desvio padrão (S)

$$S = \sqrt{SS} = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N - 1}} \qquad (4)$$

Uma das dificuldades para se tratar com alguns desses resultados é que a notação varia muito de autor para autor, estou buscando uma notação que seja mais próxima de nosso universo de conhecimento comum.

Tanto a variância quanto o desvio padrão são medidas associadas à dispersão em torno da média. Porém, para avaliar qual é o valor da medida final, utilizamos o erro padrão da média (SE), associado ao valor médio obtido

$$V = \bar{X} \mp SE \qquad (5)$$

onde o valor do erro padrão da média é expresso por

$$SE = \frac{S}{\sqrt{N}} \qquad (6)$$

1.2. Confiabilidade de uma medida

Em geral temos dois tipos de erros associados a medidas experimentais em laboratórios, os erros aleatórios e os erros sistemáticos. Como compreender a diferença entre esses dois erros? Uma das formas mais simples é analisar o que ocorre com um atirador que busca acertar um alvo à distância, com uma mira que utiliza equipamentos óticos. Se a mira estiver descalibrada, os tiros serão todos desviados na mesma direção. Ainda que a mira esteja estritamente calibrada, variações de temperatura, vento e do próprio atirador, com um caráter de pouca previsibilidade, poderão ser consideradas como aleatórias, com os tiros se espalhando em torno do centro do alvo.

Esses dois tipos de erros nos conduzem à dois indicadores de erro, a precisão e a exatidão. Você irá perceber na continuidade do texto que existem diferenças e semelhanças entre medidas em física e medidas em ciências humana, mas que as similaridades permitem utilizar essa abordagem para introduzirmos os conceitos a partir de exemplos mais conhecidos.

Segundo nosso manual de laboratório *"Erros sistemáticos fazem com que as medidas feitas estejam consistentemente acima ou abaixo do valor real, prejudicando a exatidão ("accuracy") da medida";* enquanto *"os erros aleatórios são flutuações, para cima ou para baixo [...] afeta[ndo] a precisão ("precision") da medida".* Uma representação do efeito da precisão e exatidão é apresentada na Figura 1.

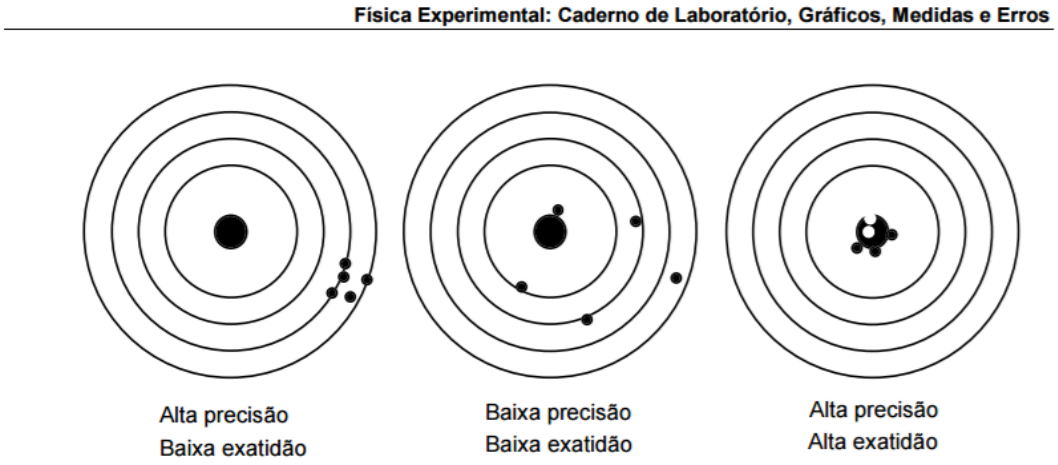


Figura 1. Precisão e exatidão de uma medida.

Vamos inicialmente observar a exatidão da medida. A exatidão da medida está associada à medida ser confiável, ser repetitiva, apresentar resultados próximos em situações similares. É importante reconhecer quando uma medida é confiável, então vamos criar um indicador de confiabilidade de uma medida, o índice de confiabilidade (IC). Uma medida confiável é aquela que é mais próxima do valor verdadeiro $X \cong V$, ou, visto por outro ângulo, o erro deve ser o menor possível, $e \cong 0$. Para nosso IC vamos utilizar a razão entre o valor verdadeiro e o valor medido

$$IC = \frac{V}{X} \qquad (7)$$

Se o valor medido X for igual ao valor verdadeiro V, o valor do IC será igual a um. Como não temos acesso ao valor verdadeiro V, podemos reescrever a equação acima como

$$IC = \frac{X - e}{X} = 1 - \frac{e}{X} \qquad (8)$$

Novamente, se a medida for isenta de erro, temos que o valor de IC é também um. Com o aumento do erro experimental, o valor de IC diminui, apresentando um limite inferior igual a zero, em uma situação improvável onde a medida é igual ao erro.

A variância de uma medida está associada à quantidade de informação que essa medida transporta, quanto maior a variância mais informação está sendo fornecida sobre o sistema[2]. Portanto, para ganharmos em informação sobre o sistema, podemos substituir as razões entre as variáveis por suas razões entre as variâncias, reescrevendo a equação (1)

$$SS_X = SS_V + \ SS_e + \ SS_{Ve} \qquad (9)$$

Nessa expressão temos um termo a mais do que na eq. 1, que é associado à covariância entre a medida verdadeira e do erro da medida, SS_{Ve}. Como não é razoável que exista correlação entre o erro e a medida verdadeira, podemos definir como nula essa covariância e reescrever a expressão para o índice de confiabilidade em função das variâncias

$$IC = \frac{SS_V}{SS_X} = \frac{SS_X - \ SS_e}{SS_X} = 1 - \frac{SS_e}{SS_X} \qquad (10)$$

Redefinimos o conceito do índice de confiabilidade, em função das variâncias associadas ao erro e às medidas realizadas. Vamos buscar aplicar os conceitos apresentados em diferentes situações.

1.3. Exercício #1

A partir dos conceitos apresentados no tópico anterior, calcule qual é o índice de confiabilidade para o resultado de uma média de N medidas. Justifique as aproximações necessárias.

[2] No caso de medidas experimentais em física, esse ganho da variância pode ser traduzido como indicador dos limites dos modelos utilizados para explicar os resultados ou dos limites experimentais dos equipamentos de medidas.

1.4. Experimento sobre calorimetria

Um professor de física resolveu fazer uma atividade experimental para estudar efeitos de capacidade térmica e da radiação solar. A atividade consistia em medira a temperatura com um termômetro digital, de um conjunto de oito amostras de mesmo tamanho, massa e composição, colocadas no chão da quadra de esportes, sob o Sol. A temperatura inicial era a mesma para todos os oito blocos, próxima de 15 ^{0}C, sendo a temperatura medida com um termopar acoplado a um "termômetro digital", para o qual o fabricante indica uma precisão de 0,05 ^{o}C.

Os oito blocos foram expostos ao mesmo ao mesmo tempo e foi realizada uma única medida, estabilizada depois de dezenas de segundos. Os alunos juntamente com o professor definiram que iriam tratar os resultados como se fosse um único bloco com oito medidas.

Foram obtidos os seguintes valores para a temperatura final:

Amostra	Temp. (^{o}C)
1	32.4
2	31.6
3	32.2
4	31.8
5	24
6	22.6
7	21.4
8	20

Tabela 1. Temperatura final dos corpos.

A partir desses resultados, calcularam a média, a variância, o desvio padrão e o erro padrão médio da temperatura, para caracterizarem o que ocorreu. Os resultados estão apresentados na Tabela 2.

	T ^{0}C
$\bar{M}$	27,0
SS	29.9
S	5.5
SE	1.9
N	8

Tabela 2. Valores de média, variância, desvio padrão, erro padrão e número de amostras para a temperatura.

Calculamos um erro padrão de 2,0 ^{o}C, sem problemas com a precisão de nosso termômetro.

1.5. Exercício #2

Calcule o valor para o índice de confiabilidade dessa medida.

O que podemos afirmar sobre nossa medida?

- ✓ Temos um modelo de que a radiação solar iria aquecer os corpos idênticos da mesma forma, e essa proposta foi confirmada;
- ✓ Em vez de repetir a experiências oito vezes, a experiência foi realizada uma única vez com oito amostras consideradas iguais, e esses resultados foram tratados como se as respostas viessem de uma única amostra;
- ✓ Apesar dos bons resultados, temos um erro padrão próximo de 3 °C, o que nos parece um valor elevado.

Sempre que possível, olhe também gráficos dos dados, eles fornecem mais informações que as tabelas.

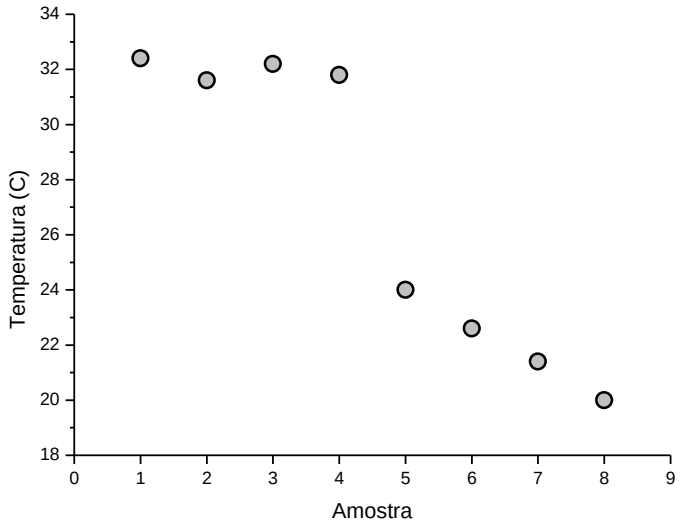


Figura 2. Temperatura das amostras.

Observando o gráfico observamos dois grupos distintos de medidas. O que fazer? Voltar para o laboratório, isto é para a quadra e olhar novamente para as amostras. Ao chegar à quadra, os alunos e o professore perceberam que agora, seis das amostras estavam na sombra! O movimento relativo do sol sombreou quatro amostras, de forma que temos um grupo com maior temperatura e outro com menor temperatura.

Vamos olhar agora para os nossos dados não como um conjunto único, mas como se fossem dois conjuntos, e calcular separadamente os valores para os dois grupos de amostras, as amostras que ficaram exclusivamente no sol (1 a 4) e as demais que estiveram parcialmente sob irradiação solar (5 a 8), conforme apresentado na Tabela 3.

	Todas	Sol	Sombra
$\bar{M}$	27,0	30,0	24,0
SS	29.9	0,19	0,22
S	5.5	0,43	0,47
SE	1.9	0,11	0,12
N	8	4	4

Tabela 3. – Medidas da temperatura para todas as amostras e para os dois conjuntos de amostras (sol e sombra).

Com essas novas condições de análise, o que podemos afirmar?

- ✓ O modelo continua valendo para a nova situação problema, temos que o tempo de exposição à radiação solar define a temperatura que a amostra irá atingir;
- ✓ Temos (no mínimo) dois grupos distintos de amostras: um com insolação constante, e outro com insolação variável;
- ✓ A metodologia proposta para o cálculo do índice de confiabilidade (expressa pela eq. 10) não é mais satisfatório para esse cálculo, pois temos agora dois conjuntos de respostas distintos!
- ✓ Para os conjuntos separados, o erro padrão é muito menor que o obtido para o conjunto completo.

Não temos mais uma única fonte de erro em nosso sistema, mas sim duas! Necessitamos rever a metodologia proposta. O erro não está associado à uma variável única, mas a dois conjuntos dessa mesma variável. Se retomamos nossa equação inicial, vemos que o resultado não é mais expresso em função de um único erro, mas de duas fontes de erro. Nosso exemplo pode ser reescrito como

$$IC = 1 - \frac{SS_e}{SS_X} = 1 - \frac{SS_{eSOL} + SS_{eSOMBRA}}{SS_{XTOTAL}} \qquad (11)$$

onde temos agora três medidas separadas, a variância total das medidas, juntamente com as variâncias das amostras no sol e na sombra.

1.6. Exercício #3

Calcule o índice de confiabilidade para a nova situação, com uma visão mais completa do experimento.

1.7. Sugestões de Leitura

ALLEN, Mary J.; YEN, Wendy M. **Introduction to measurement theory**. Waveland Press, 2001.

DE BRITO CRUZ, C. H.; FRAGNITO, Hugo Luis. Guia para Física Experimental Caderno de Laboratório, Gráficos e Erros. **Campinas, UNICAMP**, 1997. http://www.ifi.unicamp.br/~brito/graferr.pdf

DEVELLIS, Robert F. **Scale development: Theory and applications**. Sage publications, 2016.

NOLL, Victor Herbert. **Introdução às medidas educacionais**. Liv. Pioneira, 1957.

"As inúmeras citações a meu artigo de nenhuma maneira
indicam que a pessoa que o citou o tenha lido,
e nem sequer demonstram que ele ou ela
tenha sequer olhado para ele."

My current thoughts on coefficient alpha and successor procedures
Educational and Psychological Measurement
Lee J. Cronbach

2. Instrumentos de Avaliação

Vamos considerar como instrumentos de avaliação (IA) tanto os testes quanto os questionários. Porém, antes de iniciarmos a discussão é necessário compreender o significado dos nossos IA. Qual o objetivo de se desenvolver um IA? Tipicamente, os pesquisadores estão mais interessados no construto do que nos itens ou nas escalas utilizados para sua caracterização. Os IA são um caminho necessário para conhecer uma informação que não pode ser obtida/é muito difícil de ser obtida de forma direta; isso implica em que medidas diretas devem ser realizadas e os resultados dessas medidas combinados para criar as variáveis latentes que podem estar associadas ao constructo.

A melhor metodologia para se obter informações sobre um grupo de pessoas é acompanhar esse grupo em relações de trabalho, vivência ou aprendizado, interagindo e transformando o processo de avaliação em uma reflexão mútua sobre quais seriam as melhores possibilidades de conduta em diferentes situações. Essa estrutura com realimentação (feedback) constante é também conhecida como uma das formas da avaliação formativa, sendo a melhor metodologia para educar, formar e avaliar uma pessoa.

Porém essas avaliações são limitadas no que diz respeito ao custo e à comparabilidade dos resultados. Sim, é possível preparar os avaliadores para que todos tenham um padrão similar de comportamento em suas medidas, porém o custo dessa operação e o tempo que se levaria, tanto na formação dos avaliadores quanto na avaliação e na transcrição dos resultados tornaria esse processo inviável para média ou larga escala.

As avaliações em larga escala (seja por questionários ou por testes, os instrumentos de avaliação) costumam apresentar um caráter geográfico (regional, estadual ou nacional); geralmente são apresentadas em uma estrutura de itens de múltipla escolha; e no caso dos testes existe um controle rígido de horário e de condições para a realização dos testes, impedindo qualquer discussão ou esclarecimento que não se encontre escrito na documentação fornecida aos participantes. Esse grupo de avaliações inclui o Enem, pelos vestibulares e pelos concursos públicos com maior concorrência.

O Enem apresenta também um exemplo de questionário aplicado em larga escala, associado ao questionário socioeconômico que os candidatos respondem durante o processo de inscrição. É uma fonte muito importante de informações sobre o perfil de nossos estudantes, porém ainda pouco utilizada pelos pesquisadores brasileiros.

Existem alguns pressupostos ao se tratar os resultados das aplicações dos exames de larga escala, espera-se que:

- ✓ Exista um número limitado de perfis associados ao constructo que você está buscando em sua pergunta de pesquisa;
- ✓ Seu instrumento de avaliação apresenta itens que permitem descrever esse perfil;
- ✓ Os perfis obtidos a partir da aplicação desse IA são comparáveis aos perfis que você observaria em uma análise individual de cada participante da amostra.

O que se espera de um bom IA aplicado em larga escala é que o grande número de participantes do processo forneça informações muito próximas às que você obteria em uma interação pessoal com os participantes. A arte de se obter bons resultados em avaliações de larga escala é construir indicadores que sejam adequados à criar o constructo necessário para responder à sua pergunta de pesquisa.

2.1. Constructo e variáveis latentes

Os constructos são os indicadores formados por variáveis latentes que esperamos que nossas medidas indiretas (através das respostas de variáveis (supostamente) mensuráveis) possam permitir ao pesquisador acessar, por exemplo: o grau de conhecimento em alguns tópicos de ensino; grau de satisfação com uma compra/viagem/hotel/serviço; perfil das expectativas dos ingressantes na universidade; expectativas sobre os filhos; nível de dependência de drogas/álcool; satisfação com os livros didáticos, etc.

No sentido de nosso trabalho, o construto se refere à uma variável latente intangível – refere-se à um conceito, uma ideia, um comportamento, ao domínio de um conteúdo. Esse domínio de conhecimentos, habilidades, competências ou hábitos é o que se busca ao realizar uma pesquisa em ciências humanas, ensino ou educação. A mediação entre o constructo e a realidade é feita utilizando um conjunto de variáveis latentes.

As variáveis latentes são as diversas possibilidades de reconhecimento do fenômeno associado ao construto. Vamos analisar dois exemplos

2.1.1. Radiação solar e blocos

No caso do aquecimento dos blocos na quadra de esportes, em primeira aproximação tínhamos somente efeito da insolação e da capacidade térmica dos blocos. A capacidade térmica não é uma variável mensurável, ela pode ser considerada uma variável latente. O constructo é a combinação das leis físicas de absorção e emissão de calor, conjugada com a radiação solar. Para acessar esse conjunto de conhecimentos abstratos – o constructo – realizamos medidas (temperatura e tempo de exposição ao sol, por exemplo) e utilizamos esses resultados para inferir características de variáveis latentes (quantidade de radiação recebida e emitida, por exemplo) que são associadas à descrição do construto.

A Figura 4 apresenta um esquema sobre o construto, as variáveis latentes e as medidas.

Figura 4. O construto é representado pela estrutura verde, a qual contém as variáveis latentes, as quais estão todas correlacionadas entre sí, enquanto as entradas externas associada à insolação e ao tempo são os parâmetros externos.

Essa foi a proposta original do experimento, porém, o que se percebe é que não é possível associar à um construto mais elaborado sobre calorimetria sem uma medida da massa dos blocos. Note que existe uma resposta independente da massa dos blocos, porém, dependendo do nível de elaboração que se deseje do construto, podem ser necessárias mais medidas externas para que as variáveis latentes do construto possam fornecer a informação que solicitamos.

Por exemplo, no caso de se buscar qual será a temperatura do corpo em função do tempo de exposição à radiação solar, temos que ter mais entradas para buscarmos uma resposta. Inclusive quem sabe seja necessário reconhecer qual é o material do qual o bloco é

constituído. Um mesmo conjunto de medidas pode estar relacionado a diferentes constructos, como é o caso observado nesse exemplo

2.1.2. Aprendizado de mecânica e satisfação do aluno

Ao buscar avaliar o impacto da satisfação do aluno no aprendizado de mecânica, necessitamos obter uma série de medidas – questionários[3] – que serão associadas à outros indicadores – prédio, carteiras, laboratório, biblioteca – e essas medidas irão se combinar em variáveis latentes visando construir um construto associado à satisfação do aluno com a escola. Além disso, ainda necessitamos de outras informações sobre a satisfação ou não do aluno com os diferentes atores da escola (colegas, professores, orientadores, diretores).

Na avaliação do aprendizado de mecânica, essa informação pode ser obtida de diferentes maneiras:

- ✓ Apresentação de um problema aberto em lápis e papel para um grupo de alunos resolverem;
- ✓ Um teste com itens de múltipla escolha similares ao Enem;
- ✓ O aluno propõe um conjunto de problemas e sugere/indica/elabora a sua solução;
- ✓ Uma montagem experimental onde alguns conceitos deveriam ser testados; sendo refutados, aperfeiçoados ou abandonados em função dos resultados.

O ideal é sempre ter o máximo de informação possível sobre qualquer grupo que se vá analisar, porém na prática nem sempre isso é possível. Os resultados seriam combinados para criar as variáveis latentes que conduziriam ao constructo, o qual permitiria estabelecer de forma ampla as relações entre satisfação e aprendizagem de mecânica.

Ao serem analisadas algumas das variáveis propostas, vemos nitidamente que será muito difícil obter uma resposta que seja significativa para todos esses questionários e avaliações. As relações entre as medidas possíveis e as variáveis latentes que irão formar o construto são bastante fluidas, provavelmente com correlações muito baixas entre sí.

O que vemos nesses dois exemplos é uma diferença enorme entre relações físicas e relações humanas no que se refere aos constructos. As relações físicas são sustentadas por um arcabouço de medidas que conduzem à uma rede simples e correlacionada de variáveis latentes que conduzem diretamente à um constructo; enquanto as medidas que podem ser quantificadas em ciências humanas são poucas, em geral associadas à intensidade de concordância com propostas ou frases ou ainda à frequência de atitudes. Essas medidas irão se correlacionar com um grande conjunto de variáveis latentes, sendo que os modelos que são possíveis de serem estabelecidos para o constructo são muito simplificados, com respostas quase óbvias, ou são tão complexos que a quantidade de informação recebida das medidas não é suficiente para estabelecer os perfis a partir das variáveis latentes e do constructo.

---

[3] Normalmente essas medidas estão associadas à itens em escala Lickert, onde são apresentadas cinco alternativas ao entrevistado, normalmente solicitando seu posicionamento perante alguma frase, entre concordar muito/discordar muito.

Essa dificuldade em se estabelecer uma uniformidade entre as medidas, as variáveis latentes e o constructo é que fazem com que seja muito importante analisar se existe coerência entre as medidas utilizadas nessa área. Uma das formas de se conferir a existência ou não dessa coerência é calcular o índice de confiabilidade desses sistemas, a partir das medidas realizadas, sendo que dentre as várias possibilidades de se medir a confiabilidade, a mais utilizada no mundo é o alfa de Cronbach. É o indicador de coerência das medidas em ciências humanas mais utilizado no mundo.

2.2. Exercício #4

Em função do segundo exemplo proposto, procure indicar três variáveis latentes que poderiam vir a formar o constructo, bem como quatro ou cinco medidas (perguntas) para conduzir às variáveis latentes, e logo, ao constructo.

2.3. Alfa de Cronbach

O alfa de Cronbach é a proposta para análise da confiabilidade mais bem-sucedida até hoje. Em um artigo de 1951, Cronbach propôs uma generalização de um índice de confiabilidade proposto por Kuder e Richardson para situações com variáveis dicotômicas, de forma a ser possível sua aplicação em um maior conjunto de situações. Apesar de ser um artigo árido, com uma discussão psicométrica complexa, esse artigo apresenta perto de 31 mil citações no Google Acadêmico. Em 2004 Cronbach comentou que "As numerosas citações de meu artigo de forma alguma indicam que a pessoa que o citou tenha lido, e nem sequer demonstra que ele ou ela tenha olhado para ele" [4]. A proposta para o cálculo da confiabilidade pelo alfa de Cronbach (α_C) estabelece uma relação entre as somatórias das variâncias, sendo calculado pela expressão a seguir:

$$\alpha_C = \frac{n}{n-1}\left(1 - \frac{\sum_{i=1}^{n} SS_i}{SS_t}\right) \qquad (12)$$

onde n é o número de itens; SS_t é a variância total e SS_i é a variância para cada item.

A confiabilidade de uma avaliação depende fortemente do número de perguntas que você realiza. Se um candidato responder a uma centena de itens dissertativos, a avaliação de sua competência na área será muito mais precisa e confiável do que se o candidato responder a 50 itens de múltipla escolha. Mesmo o número de itens presentes no exame modifica o resultado do índice de confiabilidade (ou fidedignidade), como observa Fletcher "Tipicamente, uma prova curta, de apenas 15 itens, poderia ter um coeficiente de fidedignidade de 0,50; uma prova de 30 itens bem elaborados, uma fidedignidade de 0,80; e 85 itens, num teste aprimorado ao longo de vários anos, um coeficiente de 0,90".

[4] The numerous citations to my article by no means indicate that the person who cited it had read it, and does not even demonstrate that he or she had looked at it.

2.4. Estimativa de Spearman Brown

Um dos primeiros IC de uma prova desenvolvidos foi a estimativa da meia prova de Spearman-Brown, desenvolvida no início do século passado. Seu procedimento é simples: prova aplicada era dividida em duas metades, sendo calculada a correlação de Pearson entre as duas meias-provas, e a partir dessa correlação se estima a confiabilidade para a prova completa. Quanto mais alta for a correlação, maior será a confiabilidade do teste, pois é esperado que se o indivíduo for bem na primeira metade, também irá sair-se bem na outra metade. A expressão para o cálculo da profecia sobre a confiabilidade de Spearman-Brown (IC_{SB}) é apresentada a seguir:

$$IC_{SB} = \frac{n \propto}{1 + (n-1) \propto} \qquad (13)$$

onde $\propto$ é a correlação entre as meias provas ou qualquer outro índice de confiabilidade, n é a razão entre o número de itens da nova prova e o número de itens da prova atual.

No caso da meia prova de Spearman Brown, utiliza-se a correlação entres as duas metades, ρ, e o valor de n é igual a dois, pois dobram o número de itens, sendo o resultado expresso abaixo

$$IC_{SB} = \frac{2\,\rho}{1 + \rho} \qquad (14)$$

Uma regra de ouro – ou regra do dedão – para correlações em medidas de desempenho ou similares é que ela passa a ser relevante acima de 0,7, o que implicaria em um índice de confiabilidade igual a 0,82. Esse resultado fornece uma indicação de caráter geral para valores ótimos esperados para a confiabilidade de provas ou questionários, próximos ou acima de 0,8.

2.5. Planilha Araraquara/São Carlos

Para exercitarmos um pouco a análise desses conceitos são apresentados dados da prova de Ciências da Natureza do Enem de 2014, restritos aos candidatos de São Carlos e Araraquara, para não ficar um arquivo muito grande. Alguns indicadores socioeconômicos também são apresentados, tais como o tipo de escola onde concluiu o ensino médio e o sexo d@ candidat@.

Alguns resultados característicos são apresentados nas Tabelas 5 e 6.

	Ciências da Natureza	Sexo		Rede Escolar	
	Geral	**Feminino**	**Masculino**	**Pública**	**Privada**
N	3507	1930	1577	2229	1278
M	14.4	13.6	15.4	12.4	18.0
DP	5.5	5.1	5.9	3.9	6.1

Tabela 5. Número de candidatos, média e desvio padrão para os 45 itens da prova de Ciências da Natureza.

	Física	Sexo		Rede Escolar	
	Geral	**Feminino**	**Masculino**	**Pública**	**Privada**
N	3507	1930	1577	2229	1278
M		4.0	4.7	3.7	5.5
DP		2.0	2.4	1.8	2.5

Tabela 6. Número de candidatos, média e desvio padrão para os 15 itens de Física da prova de Ciências da Natureza.

Alguns resultados característicos são apresentados nas Tabelas 7 e 8.

	Ciências da Natureza	Sexo		Rede Escolar	
	Geral	**Feminino**	**Masculino**	**Pública**	**Privada**
N	3507	1930	1577	2229	1278
α		0.68	0.75	0.45	0.76

Tabela 7. Número de candidatos e alfa de Cronbach para os 45 itens da prova de Ciências da Natureza.

	Física	Sexo		Rede Escolar	
	Geral	**Feminino**	**Masculino**	**Pública**	**Privada**
N	3507	1930	1577	2229	1278
α Cronbach		0.37	0.50	0.17	0.52
IC_{SB}		0.41		0.22	

Tabela 8. Número de candidatos, alfa de Cronbach e estimativa do índice de confiabilidade de Spearman Brown para os 15 itens de Física da prova de Ciências da Natureza.

2.6. Exercício #5

Nas Tabelas para a Física não estão calculados os valores para a média e o desvio padrão dos itens de Física. A partir da planilha fornecida calcule esses valores. Repita a operação para os valores de alfa de Cronbach nos dois casos.

2.7. Exercício #6

Utiliza a expressão da profecia de Spearman Brown para calcular qual seria o valor estimado para os itens de Física a partir dos itens de Ciências da Natureza.

2.8. Sugestões de Leitura

ALLEN, Mary J.; YEN, Wendy M. **Introduction to measurement theory**. Waveland Press, 2001.

CRONBACH, Lee J. Coefficient alpha and the internal structure of tests. **psychometrika**, v. 16, n. 3, p. 297-334, 1951.

CRONBACH, Lee J.; SHAVELSON, Richard J. My current thoughts on coefficient alpha and successor procedures. **Educational and psychological measurement**, v. 64, n. 3, p. 391-418, 2004.

DROST, Ellen A. Validity and reliability in social science research. **Education Research and Perspectives**, v. 38, n. 1, p. 105, 2011.

FLETCHER, Philip Ralph. Da Teoria Clássica dos Testes para os Modelos de Resposta ao Item. 2010. http://biblioteca.ibge.gov.br/visualizacao/livros/liv49324.pdf

MAROCO, João; GARCIA-MARQUES, Teresa. Qual a fiabilidade do alfa de Cronbach? Questões antigas e soluções modernas?. **Laboratório de Psicologia**, v. 4, n. 1, p. 65-90, 2013.

NOLL, Victor Herbert. **Introdução às medidas educacionais**. Liv. Pioneira, 1957.

"Pensar claramente sobre o conteúdo
da escala a ser construída requer
pensar claramente sobre o constructo
que está sendo medido"

Scale Development Theory and Applications
Robert F. DeVellis

3. Como Elaborar Questionários

3.1. Introdução

Essa é uma adaptação livre do site https://wiki.ecdc.europa.eu/fem/w/wiki/questionnaire-design sobre o desenvolvimento de questionários para a área de epidemiologia. Curiosamente, a área de saúde pública costuma ser muito próxima da área de educação e ensino no que se refere às metodologias de pesquisa e de survey com populações variadas.

Não se pretende que seja possível elaborar um questionário com alta qualidade a partir somente dessas informações, mas em função do tempo restrito que temos, são os pontos principais para serem discutidos

3.2. Definições Gerais

Questionários são uma das ferramentas mais importantes para pesquisadores em diferentes campos do conhecimento. Uma das vantagens dos questionários é a possibilidade de atingir um grande número de pessoas ao mesmo tempo. Por outro lado, utilizar questionários apresenta a desvantagem de ter pouca informação sobre cada sujeito e, em alguns casos, não

perguntar a questão correta da qual se busca a resposta. O sucesso de uma investigação ou projeto de pesquisa depende fortemente de uma alta resposta da população alvo e da quantidade de informação válida que foi obtida.

Existem diferentes tipos de questionários, porém vamos enfatizar os questionários que podem são preenchidos pelos respondentes e no formato de múltipla escolha em escala Lickert. Essa escolha torna similar em alguns aspectos a discussão sobre a elaboração de questões em provas de múltipla escolha.

Um questionário bem construído é aquele que pode fornecer as respostas adequadas à sua pergunta de pesquisa. Um questionário bem estruturado é muito mais fácil de ser completado, apresentando menos dúvidas aos participantes. Finalmente, criar um bom questionário é crucial para o sucesso de seu projeto.

A principal vantagem de utilizar um instrumento de avaliação (seja ele um teste ou um questionário) é o fato de que um grande número de pessoas pode ser alcançado de forma relativamente simples e econômica. Outro fator importante é a relativa facilidade de analisar os resultados.

Questionário não são necessariamente a melhor forma de se obter uma informação. O risco das perguntas em um questionário não ampliarem o conhecimento sobre o constructo desejado é elevado, além do problema que um questionário sempre apresenta um número limitado de itens a serem respondidos.

Outro revés dos questionários é a variação das respostas em função da interpretação das perguntas. Os entrevistados às vezes entendem mal ou interpretam mal as perguntas. Se este for o caso, será muito difícil corrigir esses erros.

3.3. Dicas para desenhar um bom questionário

Um questionário bem concebido tem uma boa aparência, é curto e simples e abrange temas relevantes para o problema de pesquisa. Tem uma estrutura lógica e um layout agradável. Os questionários bem concebidos podem atingir uma elevada taxa de resposta e permitir uma fácil síntese e análise dos dados. As Sete Regras de Ouro (apresentadas a seguir) são úteis para desenhar perguntas apropriadas.

Apresentando a Pesquisa

O questionário deve ser sempre acompanhado por uma carta de apresentação, sendo também necessário a aprovação prévia do comitê de ética em pesquisa em ciências humanas. A carta de apresentação deve incluir informações sobre:

- ✓ Quem você é e para quem trabalha;
- ✓ Por que você está investigando;
- ✓ Como e onde você pode ser contatado;
- ✓ O(s) endereço(s) do pesquisador.

A confidencialidade deve ser garantida. O tempo solicitado para preencher o questionário ou a duração da entrevista devem ser indicados corretamente. Mais importante ainda, a introdução deve esclarecer a utilidade do estudo para os entrevistados e convencê-los a participar.

A primeira página deve conter o título da pesquisa em negrito. Todas as páginas devem ter uma marca de identificação ou um identificador único, números de página e orientações para os entrevistados. Os itens devem ser numerados. Se você optar por enviar um questionário por correio, o envio deve sempre incluir um envelope endereçado e pré-pago para facilitar uma resposta.

Ordem das Perguntas

A ordem das perguntas feitas pode influenciar as respostas. Recomenda-se agrupar as perguntas por tópico. As perguntas iniciais devem ser simples e não ameaçadoras, a fim de colocar os participantes à vontade e capturar o seu interesse. Apesar de muitos questionários iniciares com questões demográficas ou pessoais, esse não é um bom início. As primeiras perguntas devem cativar e deixar "de bom humor" os participantes pare responderem sobre as questões principais vinculadas ao seu constructo.

O maior desafio na elaboração de um questionário consiste em manter o foco na pergunta de pesquisa. O desvio deve ser evitado a todo custo. Evite coletar informações desnecessárias que não ajudem a responder seus principais tópicos. No entanto, as informações demográficas (idade, sexo, local de moradia) devem ser coletadas.

Concluir a Pesquisa

Conclua o questionário agradecendo aos participantes por sua participação e assegurando-lhes que sua participação foi realmente útil. Caso esteja conduzindo a pesquisa oralmente, nesta fase os participantes devem ter a oportunidade de fazer perguntas sobre o estudo ou o assunto em que foram entrevistados. Se necessário, peça permissão para fazer mais perguntas e registre os contatos daqueles que concordarem em fazê-lo.

3.4. Sete Regras de Ouro para Desenhar Questionários

É muito importante ajustar o estilo das perguntas ao público-alvo. A formulação das perguntas pode variar substancialmente dependendo se o público-alvo é composto por professores, alunos ou o público em geral.

1. Peça uma informação de cada vez

— Você possui um cão?

Em vez de

— Você possui um cão ou tem contato frequente com cães?

2. Faça perguntas precisas

— Quantas vezes você agradou um cachorro nos últimos 3 meses?

Em vez de

— Você costuma agradar cachorros?

3. Faça perguntas apropriadas, sem julgamento

— Quantas vezes você se ausentou da escola nos últimos 6 meses?

Em vez de

— Você falta às suas aulas com frequencia?

4. Evite perguntas sugestivas

— Qual bebida você consumiu?

Em vez de

— Você bebeu a estranha bebida rosa?

5. Seja tão simples quanto possível

— Você fumou uma média de 2 pacotes de cigarros / semana nos últimos 5 anos?

Em vez de

— Você fumou não menos do que uma quantidade média de 7 cigarros / 2 dias a partir de 1999?

6. Evite o jargão, as abreviaturas ou a gíria

— Você participa de quantas reuniões pedagógicas mensais?

Em vez de

— Você participa de quantas reuniões mensais com o PCOP?

7. Usar opções de resposta mutuamente exclusivas

Coloque as opções em uma ordem vertical ou horizontal.
Não se esqueça da opção "não sei", se aplicável.

Sim	
Não	
Não sei	

Ou

Sim	Não	Não sei

Preencha todas as alternativas,
facilitando a escolha do seu colaborador.

Todos os dias	
Várias vezes por semana	
Uma vez por semana	
Algumas vezes por mês	
Nunca	

Frequência

Nunca	
Raramente	
Às vezes	
Muitas vezes	
Sempre	

Quantidade

Muito pouco	
Pouco	
Médio	
Muito	
Bastante	

Avaliação

Péssimo	
Mau	
Razoável	
Boa	
Excelente	

3.5. Exercício #7

Vamos desenvolver um questionário com entre dez e quinze itens, buscando averiguar quais são os instrumentos que os pesquisadores mais utilizam para estudos sobre ensino de Física.

3.6. Sugestões de Leitura

BREMER, Viviane. Questionnaire Design
https://wiki.ecdc.europa.eu/fem/w/wiki/questionnaire-design

DEVELLIS, Robert F. **Scale development: Theory and applications**. Sage publications, 2016.

HILL, Manuela Magalhães et al. A construção de um questionário. **Lisboa: Dinâmia**, 1998.

GÜNTHER, Hartmut. Como elaborar um questionário. **Brasília: UnB, Laboratório de Psicologia Ambiental**, 2003. http://psi-ambiental.net/pdf/01Questionario.pdf

"Uma das influências preeminentes, com uma função normativa mais poderosa do que os programas oficiais, livros didáticos, propostas curriculares ou os atuais parâmetros, sempre foi o exame vestibular"

Myriam Krasilchik
Reformas e realidade: o caso do ensino das ciências

4. Como Elaborar Testes de Múltipla Escolha

4.1. Algumas observações sobre elaboração de itens

Tanto os itens de questionários quanto os itens de uma prova apresentam um problema em comum: devem estabelecer uma relação clara, precisa e unívoca com o constructo que se busca avaliar. Essa relação se dá através da elaboração do instrumento de avaliação – em última instância, na elaboração dos itens.

Logo, a resposta de cada item deve ser objetiva e precisa. Da mesma forma que um item em um questionário não pode perguntar ao aluno se ele "*gosta e estuda álgebra*", não podemos perguntar "*qual a altura máxima que uma pedra em lançamento oblíquo alcança e em que direção e sentido está a aceleração na altura máxima*[5]" em uma prova.

Em geral os pesquisadores buscam maximizar a informação que desejam obter em cada instrumento de avaliação desenvolvido, e isso passa por apresentar ou muitos itens, ou muita informação em cada item.

[5] Essa discussão vale para itens de múltipla escolha, pode ser parcialmente extrapolada para itens dissertativos, com alguns cuidados.

Em minha visão, eu prefiro um menor número de itens bem respondidos que um grande número com uma taxa de respostas reduzida. Os manuais clássicos indicam números de itens elevados, seja para questionários, seja para provas.

No caso de itens de provas, existe um erro comum que são as questões conhecidas na literatura como “multisteps”, com múltiplas etapas de solução. Um erro aritmético em qualquer etapa intermediária inviabiliza o resultado final, inclusive não fornecendo ao pesquisador informação sobre onde ocorreu o erro...

Outro ponto importante é o tempo que o seu colaborador que está sendo pesquisado irá investir na resposta do instrumento de avaliação. Para itens de prova, costuma-se estimar que um item deva ser resolvido por um aluno com a formação esperada em três minutos. Muitas vezes existem textos longos que inviabilizam essa estimativa devido simplesmente ao tempo necessário para a leitura!

Empatia é outra necessidade ao elaborar um instrumento de avaliação, devemos nos colocarmos na posição de nosso público alvo; qual é sua linguagem, sua história de vida, qual a condição onde vai responder ao instrumento, se é eletrônico, impresso ou oral. Mas principalmente deve ocorrer uma relação de respeito com o colaborador, não podem ser solicitadas informações fora do seu universo ou que coloquem o colaborador em uma situação incômoda por motivos pessoais ou profissionais.

É necessário “trocar de lado” como o avaliado! Temos que abandonar parcialmente nosso capital cultural e científico acumulado para pensarmos como os respondentes de nossos instrumentos.

Se formos comparar as provas de Ciências da Natureza do Enem com as demais provas, iremos perceber que o nível de contextualização dos itens em ciências nos anos iniciais do novo Enem é muito maior que nas outras áreas. Esse resultado está associado com as expectativas dos elabores de itens de muitos pesquisadores que desejam estabelecer uma relação mais próxima da ciência com o cotidiano das pessoas, em uma visão muito próxima da área de pesquisa associada à Ciência, Tecnologia, Sociedade e Ambiente, CTSA.

Uma das dificuldades em física é que o cotidiano é muito mais complicado do que parece, e que realizar as aproximações necessárias para que seja possível aplicar as equações mais triviais às situações reais, é extremamente difícil, senão impossível em muitos casos. Essa dificuldade de buscar uma transposição didática do cotidiano para uma avaliação é um dos pontos que a comunidade de ensino de ciências irá ter que tomar uma posição mais clara, pois, com o formato que temos hoje – tanto no que se refere ao número de horas de aula nas escolas públicas quanto a quantidade de conteúdo – não existe como os alunos aprenderem os conceitos gerais de Física.

Elaborar itens não é um trabalho solitário. Em minha experiência de elaborador de itens de diferentes exames, sempre o resultado foi melhor ao trabalhar em grupos de pessoas, preferivelmente com formações distintas, ou pelo menos que trabalhem em áreas de conhecimento distintas, se tiverem a mesma formação.

Um dos melhores resultados que tínhamos com itens era quando não era o mesmo elaborador que realizava a mesma etapa do processo inteiro, isto é: elaborador A tem a ideia do item; elaborador B pensa na ideia e escreve uma proposta para o item enquanto o elaborador C irá resolver e criticar o texto e as alternativas. Temos muitos subentendidos quando pensamos em um item, não escrevemos todos os passos necessários.

Se eu fosse simplificar a elaboração de itens em uma única frase, seria “*siga o constructo!* ”.

4.2. Preparação para Escrever Itens

Esse tópico é uma adaptação do website do Zimmaro.

4.2.1. Crie itens de teste enquanto prepara as suas aulas:

1. Anote as perguntas que os alunos fazem com maior frequência durante a aula.
2. Anote os equívocos comuns (concepções prévias, alternativas ou não científicas) que os alunos fazem durante as aulas ou no dever de casa.
3. Peça aos alunos que elaborem itens ao final da aula ou em outras ocasiões.

4.2.2. Vantagens:

1. Podem ser avaliados resultados de aprendizagem desde situações mais simples até mais complexas.
2. As tarefas que devem ser realizadas podem ser estruturadas, precisas e objetivas.
3. Uma ampla amostra de conhecimentos e habilidades pode ser medida.
4. Alternativas incorretas fornecem informações diagnósticas sobre problemas de compreensão dos alunos.
5. As pontuações são mais conviáveis do que os itens pontuados subjectivamente, como alguns itens dissertativos e redações, por exemplo.
6. A correção e a pontuação são fáceis, objetivas e confiáveis.
7. A análise de itens pode revelar a dificuldade para cada item; além de permitir avaliar quanto o resultado do ítem é capaz de diferenciar entre os grupos de alunos com maior e menor desempenho.
8. O desempenho pode ser comparado de uma classe a outra e de um ano para outro, de um estado para outro e até entre países distintos.
9. É possível avaliar muito mais tópicos do que em uma avaliação dissertativa, pois se estima que cada item demore em torno de três minutos para ser respondido.

4.2.3. Limitações:

1. Elaborar bons itens é demorado, difícil e, consequentemente, caro.
2. É muito difícil encontrar alternativas erradas (distractores) que pareçam plausíveis.
3. Itens de múltipla escolha não medem algumas resoluções de problema, nem a capacidade de organizar e expressar idéias, essas devem ser avaliadas por outros formatos de testes.
4. Uma situação problema do mundo real apresenta outras nuancias, deve se propor uma solução, enquanto em multipa escolha basta selecionar uma alternativa.
5. A escolha das alternativa é influenciada pela capacidade de leitura e pelo treinamento anterior na resolução de itens de múltipla escolha.

6. Não é uma avaliação formativa, existe uma falta de feedback sobre os processos de pensamento individuais - é difícil determinar por que os alunos selecionaram respostas incorretas[6].
7. Os candidatos se concentram em utilizar níveis mais básicos do processo cognitivo, evitando ou não atingindo níveis mais elaborados.
8. Distratores dúbios podem causar muitos problemas, sugerindo a existência de mais de uma resposta "correta" defensável.
9. Uma resposta correta depende muito da capacidade de leitura e análise do aluno.
10. Não avalia a capacidade de escrever ou de se esforçar.
11. Pode encorajar a adivinhação das respostas.

4.2.4. Dicas:

1. Basear cada item em um objetivo educacional coerente com o grupo que deve ser avaliado.
2. Tente escrever itens em que haja uma e apenas uma resposta correta ou uma resposta nitidamente mais correta que as demais.
3. A frase que introduz o item (enunciado) deve indicar claramente o problema.
4. Testar apenas uma ideia em cada item.
5. Certifique-se de que escolhas de alternativas erradas (distratores) são – pelo menos – plausíveis.
6. Incorpore os erros mais comuns dos alunos em seus distratores. Esse é um ponto de discórdia entre distintos avaliadores. Muitos acreditam que utilizar, por exemplo, concepções alternativas, é “uma pegadinha” contra os alunos que estão realizando a prova. Outros afirmam que dessa forma, temos condições de separar os níveis de conhecimento.
7. A posição da resposta correta deve variar aleatoriamente de item para item. Inclua de três a cinco opções para cada item. Lembrar que alguns anos atrás em um vestibular paulista um programador gerou uma ordem com um padrão de respostas que muitos alunos descobriram e gabaritaram na prova.
8. O comprimento (texto) das alternativas deve ser aproximadamente o mesmo em cada item (preferencialmente baixo).
9. Evite o uso de "Todas as anteriores" e "Nenhuma das anteriores"

[6] Esse é exatamente a área em que eu trabalho, que é buscar nas respostas coletivas dos distratores informações para que os professores possam aperfeiçoar o seu trabalho em sala de aula.

4.3. Exercício #8

Observe os itens propostos a seguir e discuta sua elaboração.

	2009
Item	**05**
Grupo	Média
Geral	0.58
Pública	0.55
Privada	0.74

Questão 5

Na linha de uma tradição antiga, o astrônomo grego Ptolomeu (100-170 d.C.) afirmou a tese do geocentrismo, segundo a qual a Terra seria o centro do universo, sendo que o Sol, a Lua e os planetas girariam em seu redor em órbitas circulares. A teoria de Ptolomeu resolvia de modo razoável os problemas astronômicos da sua época. Vários séculos mais tarde, o clérigo e astrônomo polonês Nicolau Copérnico (1473-1543), ao encontrar inexatidões na teoria de Ptolomeu, formulou a teoria do heliocentrismo, segundo a qual o Sol deveria ser considerado o centro do universo, com a Terra, a Lua e os planetas girando circularmente em torno dele. Por fim, o astrônomo e matemático alemão Johannes Kepler (1571-1630), depois de estudar o planeta Marte por cerca de trinta anos, verificou que a sua órbita é elíptica. Esse resultado generalizou-se para os demais planetas.

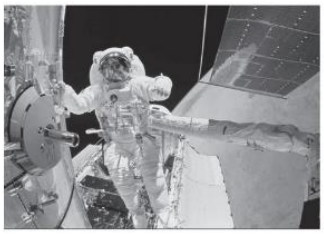
A respeito dos estudiosos citados no texto, é correto afirmar que

- **A** Ptolomeu apresentou as ideias mais valiosas, por serem mais antigas e tradicionais.
- **B** Copérnico desenvolveu a teoria do heliocentrismo inspirado no contexto político do Rei Sol.
- **C** Copérnico viveu em uma época em que a pesquisa científica era livre e amplamente incentivada pelas autoridades.
- **D** Kepler estudou o planeta Marte para atender às necessidades de expansão econômica e científica da Alemanha.
- **E** Kepler apresentou uma teoria científica que, graças aos métodos aplicados, pôde ser testada e generalizada.

	2009
Item	**27**
Grupo	Média
Geral	0.14
Pública	0.14
Privada	0.14

Questão 27

O ônibus espacial *Atlantis* foi lançado ao espaço com cinco astronautas a bordo e uma câmera nova, que iria substituir uma outra danificada por um curto-circuito no telescópio *Hubble*. Depois de entrarem em órbita a 560 km de altura, os astronautas se aproximaram do *Hubble*. Dois astronautas saíram da *Atlantis* e se dirigiram ao telescópio. Ao abrir a porta de acesso, um deles exclamou: "Esse telescópio tem a massa grande, mas o peso é pequeno."



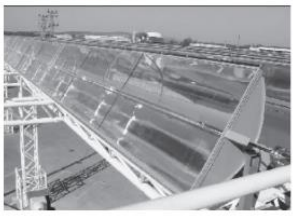
Considerando o texto e as leis de Kepler, pode-se afirmar que a frase dita pelo astronauta

- **A** se justifica porque o tamanho do telescópio determina a sua massa, enquanto seu pequeno peso decorre da falta de ação da aceleração da gravidade.
- **B** se justifica ao verificar que a inércia do telescópio é grande comparada à dele próprio, e que o peso do telescópio é pequeno porque a atração gravitacional criada por sua massa era pequena.
- **C** não se justifica, porque a avaliação da massa e do peso de objetos em órbita tem por base as leis de Kepler, que não se aplicam a satélites artificiais.
- **D** não se justifica, porque a força-peso é a força exercida pela gravidade terrestre, neste caso, sobre o telescópio e é a responsável por manter o próprio telescópio em órbita.
- **E** não se justifica, pois a ação da força-peso implica a ação de uma força de reação contrária, que não existe naquele ambiente. A massa do telescópio poderia ser avaliada simplesmente pelo seu volume.

	2009
Item	**35**
Grupo	Média
Geral	0.13
Pública	0.13
Privada	0.13

Questão 35

O Sol representa uma fonte limpa e inesgotável de energia para o nosso planeta. Essa energia pode ser captada por aquecedores solares, armazenada e convertida posteriormente em trabalho útil. Considere determinada região cuja insolação — potência solar incidente na superfície da Terra — seja de 800 watts/m^2. Uma usina termossolar utiliza concentradores solares parabólicos que chegam a dezenas de quilômetros de extensão. Nesses coletores solares parabólicos, a luz refletida pela superfície parabólica espelhada é focalizada em um receptor em forma de cano e aquece o óleo contido em seu interior a 400 °C. O calor desse óleo é transferido para a água, vaporizando-a em uma caldeira. O vapor em alta pressão movimenta uma turbina acoplada a um gerador de energia elétrica.



Considerando que a distância entre a borda inferior e a borda superior da superfície refletora tenha 6 m de largura e que focaliza no receptor os 800 watts/m^2 de radiação provenientes do Sol, e que o calor específico da água é 1 cal g^{-1} °C^{-1} = 4.200 J kg^{-1} °C^{-1}, então o comprimento linear do refletor parabólico necessário para elevar a temperatura de 1 m^3 (equivalente a 1 t) de água de 20 °C para 100 °C, em uma hora, estará entre

- **A** 15 m e 21 m.
- **B** 22 m e 30 m.
- **C** 105 m e 125 m.
- **D** 680 m e 710 m.
- **E** 6.700 m e 7.150 m.

	2010
	59
Grupo	Média
Geral	0.55
Pública	0.50
Privada	0.73

Questão 59

As cidades industrializadas produzem grandes proporções de gases como o CO_2, o principal gás causador do efeito estufa. Isso ocorre por causa da quantidade de combustíveis fósseis queimados, principalmente no transporte, mas também em caldeiras industriais. Além disso, nessas cidades concentram-se as maiores áreas com solos asfaltados e concretados, o que aumenta a retenção de calor, formando o que se conhece por "ilhas de calor". Tal fenômeno ocorre porque esses materiais absorvem o calor e o devolvem para o ar sob a forma de radiação térmica.

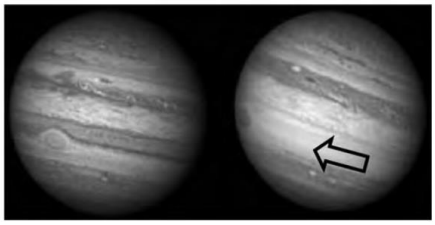
Em áreas urbanas, devido à atuação conjunta do efeito estufa e das "ilhas de calor", espera-se que o consumo de energia elétrica

A diminua devido à utilização de caldeiras por indústrias metalúrgicas.
B aumente devido ao bloqueio da luz do sol pelos gases do efeito estufa.
C diminua devido à não necessidade de aquecer a água utilizada em indústrias.
D aumente devido à necessidade de maior refrigeração de indústrias e residências.
E diminua devido à grande quantidade de radiação térmica reutilizada.

	2010
Item	**54**
Grupo	Média
Média	0.30
Pública	0.29
Privada	0.37

Questão 54

Júpiter, conhecido como o gigante gasoso, perdeu uma das suas listras mais proeminentes, deixando o seu hemisfério sul estranhamente vazio. Observe a região em que a faixa sumiu, destacada pela seta.



Disponível em: http://www.inovacaotecnologica.com.br. Acesso em: 12 maio 2010 (adaptado).

A aparência de Júpiter é tipicamente marcada por duas faixas escuras em sua atmosfera — uma no hemisfério norte e outra no hemisfério sul. Como o gás está constantemente em movimento, o desaparecimento da faixa no planeta relaciona-se ao movimento das diversas camadas de nuvens em sua atmosfera. A luz do Sol, refletida nessas nuvens, gera a imagem que é captada pelos telescópios, no espaço ou na Terra.
O desaparecimento da faixa sul pode ter sido determinado por uma alteração

A na temperatura da superfície do planeta.
B no formato da camada gasosa do planeta.
C no campo gravitacional gerado pelo planeta.
D na composição química das nuvens do planeta.
E na densidade das nuvens que compõem o planeta.

	2010
Item	**89**
Grupo	Média
Geral	0.39
Pública	0.34
Privada	0.60

Questão 89

Deseja-se instalar uma estação de geração de energia elétrica em um município localizado no interior de um pequeno vale cercado de altas montanhas de difícil acesso. A cidade é cruzada por um rio, que é fonte de água para consumo, irrigação das lavouras de subsistência e pesca. Na região, que possui pequena extensão territorial, a incidência solar é alta o ano todo. A estação em questão irá abastecer apenas o município apresentado.

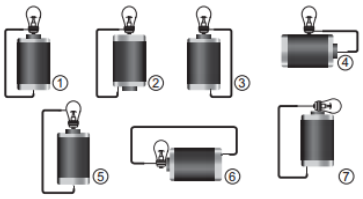
Qual forma de obtenção de energia, entre as apresentadas, é a mais indicada para ser implantada nesse município de modo a causar o menor impacto ambiental?

A Termelétrica, pois é possível utilizar a água do rio no sistema de refrigeração.
B Eólica, pois a geografia do local é própria para a captação desse tipo de energia.
C Nuclear, pois o modo de resfriamento de seus sistemas não afetaria a população.
D Fotovoltaica, pois é possível aproveitar a energia solar que chega à superfície do local.
E Hidrelétrica, pois o rio que corta o município é suficiente para abastecer a usina construída.

	2011
Item	**70**
Grupo	Média
Geral	0.36
Pública	0.35
Privada	0.40

QUESTÃO 70

Um curioso estudante, empolgado com a aula de circuito elétrico que assistiu na escola, resolve desmontar sua lanterna. Utilizando-se da lâmpada e da pilha, retiradas do equipamento, e de um fio com as extremidades descascadas, faz as seguintes ligações com a intenção de acender a lâmpada:



Tendo por base os esquemas mostrados, em quais casos a lâmpada acendeu?

A (1), (3), (6)
B (3), (4), (5)
C (1), (3), (5)
D (1), (3), (7)
E (1), (2), (5)

4.4. Sugestões de Leitura

ZIMMARO, Dawn M. Writing good multiple-choice exams. **Measurement and Evaluation Center. University of Texas at Austin. Web site**, v. 16, 2004. http://www6.cityu.edu.hk/edge/workshop/seminarseries/2010-11/Seminar03-WritingGoodMultipleChoiceExams.pdf

NOLL, Victor Herbert. **Introdução às medidas educacionais**. Liv. Pioneira, 1957

MAGER, Robert F.. **Medindo os objetivos de ensino: ou" consegui um par adequado?"** Editora Globo. 1981.

Guia de **Elaboração** e Revisão de **Itens** – INEP * o link foi retirado do INEP, será disponibilizado durante o minicurso