

Validação de instrumentos de medida aplicados à pesquisa em ensino de Física

Fernando Lang da Silveira – Instituto de Física da UFRGS

lang@if.ufrgs.br

Validação de instrumentos de medida aplicados à pesquisa em Ensino de Física

*Fernando Lang da Silveira - Instituto de Física da UFRGS
lang@if.ufrgs.br*

Material preparado para a II Escola Latino-Americana sobre
Pesquisa em Ensino de Física, Porto Alegre (Canela), 05 a 16/07/1993.

Silveira, F. L. Validação de testes de lápis e papel. In: Moreira, M.A. e
Silveira, F.L. Instrumentos de pesquisa em ensino e aprendizagem. Porto
Alegre: EDIPUCRS, 1993.

Introdução

Do ponto de vista da pesquisa, tem se argumentado, provavelmente com razão, que a entrevista clínica é a melhor técnica para investigar o conhecimento do aluno, suas concepções, suas crenças, seus modelos não científicos, por exemplo. Contudo, da ótica do ensino em sala de aula, a entrevista clínica parece ser inviável porque requer muita prática, consome muito tempo e não pode ser aplicada a um grande número de alunos. Além disso, o professor na maioria das vezes, não é um pesquisador. O que lhe interessa, por exemplo, é saber, rapidamente, se seus alunos (que, geralmente, são muitos) têm ou não têm determinadas concepções alternativas que os pesquisadores detectaram, em outros alunos, através de entrevistas clínicas.

Para isso, a entrevista clínica é não só inviável como, também, desnecessária. O professor pode facilmente elaborar um teste de papel e lápis, geralmente de escolha múltipla, usando como indicadores os resultados das pesquisas. Este teste, uma vez validado, dá a ele em pouco tempo a informação.

Este é apenas um exemplo. Existem inúmeras outras situações em que os professores – e também os pesquisadores, é claro, fazem uso de testes de papel e lápis (provas de rendimento escolar, escalas de atitude, opiniários, questionários de avaliação do desempenho do professor são outros exemplos). Tais testes, podem, no entanto, ser sempre validados. Precisamente aí se enquadra o objetivo deste trabalho: explicar como se validam instrumentos dessa natureza.

Medida da intensidade da relação entre duas variáveis

A Tabela 1 apresenta, para 32 alunos de FIS 182 (UFRGS), os escores em um teste sobre concepções em relação à corrente elétrica (Silveira, Moreira e Axt, 1989) e a média final na disciplina.

Há relação entre C e F?

Uma simples inspeção visual da Tabela 1 revela a tendência de que aumentando C, aumenta F. Também constata-se facilmente que essa tendência está longe de se constituir em uma relação perfeita; observa-se que a perfeita ordenação crescente dos escores C é acompanhada por uma ordenação imperfeita de F.

Um **coeficiente de correlação** quantifica a intensidade da relação entre duas variáveis. Esta quantificação pode ser extremamente útil, pois, permitirá dizer até quanto uma variável se relaciona com outra e possibilitará comparar a intensidade da relação entre pares de variáveis. Apresenta-se a seguir uma equação que permite calcular o coeficiente de correlação de Pearson (Wherry, 1984), simbolizado por (r_{XY}) .

$$r_{XY} = \frac{\Sigma XY - \Sigma X \Sigma Y/n}{((\Sigma X^2 - (\Sigma X)^2/n)(\Sigma Y^2 - (\Sigma Y)^2/n))^{1/2}}$$

De acordo com essa equação, o cálculo do coeficiente de correlação envolve cinco somatórios com os valores observados de X e Y; n é o número de pares de valores observados, ou o número de termos em cada somatório.

Para os dados da Tabela 1 estes cinco somatórios são os seguintes:

$$\Sigma C = (1 + 2 + 3 + \dots + 14 + 14) = 345$$

$$\Sigma C^2 = (1^2 + 2^2 + 3^2 + \dots + 14^2 + 14^2) = 4151$$

$$\Sigma F = (0,0 + 1,5 + 3,0 + \dots + 8,5 + 9,5) = 169,5$$

$$\Sigma F^2 = (0,0^2 + 1,5^2 + 3,0^2 + \dots + 8,5^2 + 9,5^2) = 1153,25$$

Tabela 1

Relação dos escores no teste sobre concepções relativas à corrente elétrica (C) e a média final em FIS 182 (F).

Aluno	C	F
1	1	0,0
2	2	1,5
3	3	3,0
4	5	1,0
5	6	2,0
6	7	0,5
7	9	1,5
8	9	5,5
9	10	6,5
10	11	2,0
11	11	4,5
12	11	5,0
13	11	7,0
14	11	7,5
15	11	8,0
16	12	7,0
17	12	8,0
18	13	2,5
19	13	6,0
20	13	6,0
21	13	6,5
22	13	8,0
23	13	8,0
24	13	8,5
25	14	0,5
26	14	6,0
27	14	7,0
28	14	7,0
29	14	7,0
30	14	8,0
31	14	8,5
32	14	9,5

$$\Sigma CF = ((1 \times 0,0) + (2 \times 1,5) + (3 \times 3) + \dots + (14 \times 8,5) + (14 \times 9,5)) = 2055$$

Calculando-se o coeficiente de correlação de Pearson entre C e F obtém-se

$$r_{CF} = \frac{2055 - (345 \times 169,5)/32}{((4151 - 345^2/32) \times (1153,25 - 169,5^2/32))^{1/2}}$$

$$r_{CF} = +0,69$$

O coeficiente de correlação de Pearson é uma medida da **relação linear** entre X e Y , sendo também denominado coeficiente de correlação linear. O seu valor está necessariamente compreendido no intervalo fechado de -1 a +1. Os valores extremos identificam uma relação linear perfeita entre X e Y ; qualquer outro valor indica uma relação imperfeita, sendo que o valor nulo mostra ausência de relação linear entre X e Y .

O sinal do coeficiente informa sobre o sinal da declividade da reta que descreve a tendência entre X e Y .¹ Desta forma um sinal positivo indica que quando X aumenta (diminui), Y tende a aumentar (diminuir); um sinal negativo indica que quando X aumenta (diminui), Y tende a diminuir (aumentar).

A interpretação para $r_{CF} = +0,69$ é a seguinte: há uma tendência imperfeita de que aumentando C, aumenta F.

Dois propriedades importantes do coeficiente de correlação de Pearson são as seguintes:

1. Ele é invariante em módulo frente a transformações lineares em X e/ou Y .
2. Ele é quase-invariante em módulo frente a qualquer transformação monotônica em X e/ ou Y .

¹O coeficiente de correlação é a declividade da reta dos mínimos quadrados para ambas as variáveis padronizadas. A padronização de uma variável é uma especial transformação linear que consiste em subtrair desta variável a sua média e depois dividir pelo desvio padrão. A variável padronizada é adimensional – portanto livre de efeitos de escala de medida –, tem média nula e desvio padrão unitário.

Uma transformação monotônica é uma função que tem a sua derivada constante em sinal; as transformações monotônicas não alteram a ordenação dos valores observados, podendo apenas produzir uma inversão da ordem quando a derivada primeira da função transformante é negativa.

A primeira propriedade torna inócua qualquer mudança de escala nas variáveis sobre o valor do coeficiente de correlação. Ou seja, é completamente irrelevante para o estudo da relação entre C e F na Tabela 1 o fato de que os escores C vão de 1 até 14, enquanto os escores F vão de 0,0 até 9,5.

A segunda propriedade é extremamente importante pois confere o status para o coeficiente de correlação de Pearson de **medida quase-universal** de relação entre duas variáveis. Mesmo que a real relação seja não-linear, o coeficiente de correlação de Pearson detectará o relacionamento entre X e Y caso ele seja monotônico em todo o espaço, ou na maior parte do espaço de variação de X e Y . Ele poderá falhar na detecção de relação entre X e Y quando a tendência real for descrita por uma função que não tenha a sua derivada primeira constante em sinal. Em qualquer caso de não-linearidade, o coeficiente de correlação de Pearson **subestimar**á a intensidade da relação entre as duas variáveis.

Validação das medidas em ciências humanas

Quando se realiza uma operação empírica com o objetivo de medir alguma coisa, sempre é pertinente a questão se **realmente está sendo medido aquilo que se pretende**. Este questionamento é relevante para medidas em qualquer domínio; por exemplo, quando se mede a diferença de potencial elétrico com um voltímetro analógico, observa-se a posição de um ponteiro sobre uma escala. Ou seja, na verdade se mede uma posição e se infere sobre uma diferença de

potencial elétrico. O que permite transformar esta medida de posição em medida de diferença de potencial são diversas teorias (teoria sobre a relação entre o torque aplicado em uma mola com a deformação da mola; teoria que relaciona o torque magnético em uma bobina com corrente e o campo magnético sobre ela aplicado; etc.). Como tem sido enfatizado por diversos filósofos da ciência, todo o nosso conhecimento está impregnado de teorias, inclusive nossas observações e medidas (Popper, 1975).

No caso referido da medida da diferença de potencial elétrico, todas essas teorias permitem supor uma correlação (não necessariamente linear) quase-perfeita entre a posição do ponteiro e a diferença de potencial elétrico. A intensidade dessa relação dependerá inclusive do grau de sofisticação tecnológica envolvida na construção do voltímetro.

No domínio das ciências humanas o questionamento sobre a validade das medidas é igualmente importante. A situação se complica principalmente porque, de um modo geral, não se dispõem de “boas” teorias que relacionam aquilo que pode ser observado com aquilo que se deseja medir.

Um indicador é uma variável que pode ser observada, medida, e que se relaciona com a variável que se deseja medir (o construto). A posição do ponteiro do voltímetro (variável medida) é um indicador que guarda altíssima correlação com a diferença de potencial elétrico (construto). A resposta que um aluno dá em um item do teste sobre concepções relativas a corrente elétrica elaborado por Silveira, Moreira e Axt (1989) (vide Anexo 2) é um indicador da concepção (científica ou alternativa) deste aluno; muito certamente um único item será um indicador que guarda baixa correlação com a concepção do aluno. O aluno até pode não ter a concepção científica sobre corrente elétrica e dar uma resposta compatível com esta e vice-versa. É por isto que se pode, a priori, esperar uma correlação baixa entre um único item e o construto a ser medido. Um único item se constitui, teoricamente, em uma medida pouco confiável da concepção.

No domínio das ciências humanas destacam-se três formas de se discutir a validade das medidas (Ghisselli, 1964; Nunnally, 1967; Cronbach, 1973):

1. **Validade de conteúdo.**
2. **Validade em relação a critério.**
3. **Validade de construto.**

Essas três formas não são mutuamente excludentes e, rigorosamente, as duas primeiras se constituem em aspectos particulares da última.

Um instrumento (teste, questionário, etc.) possui **validade de conteúdo** se os itens que o constituem são representativos do universo que ele pretende representar, ou ainda, se os itens são, do ponto de vista teórico, indicadores da variável a ser medida, do construto. O teste de matemática do concurso vestibular à universidade tem validade de conteúdo se os itens (questões) forem representativos dos conteúdos de matemática do 2º grau. Os itens de um teste sobre concepções alternativas em física tem validade de conteúdo se eles forem coerentes com os achados das entrevistas clínicas relativas a essas concepções.

A validade de conteúdo é estabelecida através de uma análise do instrumento e do confronto dos itens com os pressupostos (teoria) que lhe deram origem. Muitas vezes se recorre ao julgamento de diversos juízes (especialistas no conteúdo do instrumento), procurando-se um consenso intersubjetivo.

A **validade em relação à critério** é estabelecida por se demonstrar empiricamente a existência de relação (através de um coeficiente de correlação ou de outra medida de associação) entre a medida do construto, realizada com instrumento a validar, e uma outra variável relevante (critério). Por exemplo, se medidas de QI têm um coeficiente de correlação de +0,60 com medidas de rendimento escolar (típica correlação entre inteligência e rendimento), então as medidas de QI tem validade em relação à critério (critério: rendimento escolar).

O critério pode ser a medida de uma variável diferente daquela que se deseja validar ou pode ser uma medida independente da mesma variável. No primeiro caso, exemplificado pelo QI e rendimento escolar, tem-se **validade preditiva**. No segundo caso, tem-se **validade concorrente**.

Um exemplo real de validade concorrente no domínio das ciências biomédicas é a alta correlação no “Teste dos 12 minutos” de Cooper com a medida em laboratório da capacidade aeróbica humana (quantidade de O₂ consumida por unidade de tempo e por unidade de massa corporal). Os estudos, com um grande número de pessoas, demonstraram haver um coeficiente de correlação de +0,90 entre a distância percorrida no “Teste dos 12 minutos” e a medida em laboratório da capacidade aeróbica (Cooper, 1976). Aliás, é esta alta correlação que permite tomar a distância como uma medida bastante precisa da capacidade aeróbica.

A construção de um argumento de **validade de construto** começa por derivar da teoria sobre a variável medida (o construto) hipóteses que expressam relações ou ausência de relações desta com outras variáveis, outros construtos. O passo seguinte é a testagem empírica das hipóteses. Caso as hipóteses venham a ser verificadas empiricamente, tem-se uma indicação indireta de que a medida do construto é válida.

O “método dos grupos conhecidos” é freqüentemente utilizado como parte da argumentação de validade de construto (Shaw e Wright, 1967). Este método começa por identificar grupos que diferem entre si no construto. Se o instrumento que mede este construto for válido, então aplicando-se o instrumento aos grupos deverá ser constatada empiricamente diferença entre os resultados nos mesmos. Silveira (1982) utilizou esse processo na validação de uma “Escala de atitude em relação à solução de problemas”; as respostas de alunos da engenharia, arquitetura e pedagogia na escala de atitude foram comparados através de uma análise de variância (Guilford, 1973), confirmando-se a hipótese de que os três grupos diferiam.

A construção do argumento de validade de construto para o “Questionário de avaliação do desempenho docente” pelos alunos (Silveira e Moreira, 1984; Silveira, Moreira e Nunes, 1985; Silveira, 1989) envolveu a testagem das seguintes hipóteses:

1. Um mesmo professor avaliado por turmas diferentes, em diferentes disciplinas e universidades tem resultados semelhantes.
2. Professores diferentes quando avaliados por seus alunos têm resultados diferentes.
3. A avaliação que os alunos fazem do seu professor não se relaciona com o grau de importância que eles atribuem à disciplina, ministrada por esse professor, em relação aos seus cursos.
4. A avaliação que os alunos fazem do professor não se relaciona com a auto-avaliação dos alunos.
5. A avaliação que os alunos fazem do professor se relaciona com o quanto os alunos consideram que aprendem nas aulas deste professor.
6. A avaliação que os alunos fazem do professor não se relaciona com a importância atribuída pelos alunos ao questionário de avaliação.

A validade em relação à critério, onde se busca verificar empiricamente a correlação do construto com outros ou com medidas independentes do mesmo, pode ser encarada como uma parte de um argumento de validade de construto. A validade de construto é configurada através da testagem de diversas hipóteses de relação do construto com outros e freqüentemente também de hipóteses de ausência de relação (Campbell e Fiske, 1959); quando a teoria permite prever que o construto se relaciona mais com uma variável do que com

outra, tal hipótese pode ser testada. Fazem parte também os estudos da estrutura interna ou da consistência interna da medida (Cronbach e Meehl, 1955) que serão abordados na próxima seção.

A fidedignidade das medidas

Na seção anterior notou-se que os coeficientes de correlação entre variáveis podem se constituir em evidências de validade para as medidas das mesmas. Os estudos correlacionais são muito comuns em outras pesquisas que não têm o objetivo de validar instrumentos. Entretanto, para se observar correlação entre duas variáveis, cada uma delas deve se correlacionar consigo mesma (esta afirmação se constitui em um teorema da teoria da medida e o leitor interessado poderá encontrar a sua demonstração, por exemplo, em Ghiselli (1964), Nunnally (1967), Lord e Novick (1968)). Por isso, nos estudos correlacionais deve haver preocupação com a correlação das medidas consigo mesmas; a autocorrelação é denominada **coeficiente de fidedignidade**.

Outro importante teorema da teoria da medida é a afirmação de que o máximo coeficiente de correlação observável entre duas variáveis é menor ou igual a raiz quadrada do produto dos coeficientes de fidedignidade de ambas as variáveis. Ou ainda, que o coeficiente de correlação observado entre duas variáveis é igual ao coeficiente de correlação verdadeiro multiplicado pela raiz quadrada do produto dos coeficientes de fidedignidade das variáveis.

Demonstra-se também que quando se comparam grupos de indivíduos através das médias de uma variável, a razão F de Snedecor utilizada na análise da variância (ou a razão t de Student no caso de dois grupos apenas) é tanto menor quanto menor for o coeficiente de fidedignidade da variável (Silveira, 1981b).

Citam-se esses teoremas para enfatizar a importância da fidedignidade das medidas. O coeficiente de fidedignidade pode ser **estimado** empiricamente de diversas maneiras. Em especial discutir-se-á um dos métodos de consistência interna para a estimativa do coeficiente de fidedignidade. O leitor poderá encontrar outras maneiras de se estimar o coeficiente de fidedignidade em qualquer dos textos referidos no primeiro parágrafo desta seção e também em Vianna (1973).

Os métodos de consistência interna são aplicáveis a instrumentos constituídos por diversos itens; a resposta a cada um dos itens deve ser, teoricamente, indicadora do construto a ser medido. Neste tipo de instrumento cada item gera um único escore para um particular indivíduo; os escores em cada item são somados, obtendo-se, assim, um escore total para o indivíduo. Os testes de rendimento escolar ou provas de verificação da aprendizagem são exemplos bastante conhecidos de instrumentos com essas características; o professor formula diversas questões (itens), as quais quando corrigidas levam a atribuição de pontos (escores) que finalmente são somados em uma única nota (escore total). Este escore total, supostamente, se constitui em uma medida de conhecimento (construto).

O escore total é uma variável composta ou um composto, isto é, tem componentes que são os escores nos itens. É possível que o escore total também seja obtido através de uma soma ponderada com pesos diferentes para diferentes itens. Para uma discussão sobre diversas maneiras de se ponderar os itens e as influências que exercem sobre o coeficiente de fidedignidade do composto pode-se consultar Silveira (1986).

Os métodos de consistência interna são apropriados para as variáveis compostas e têm como objetivo não apenas estimar o coeficiente de fidedignidade do composto, mas também identificar itens (componentes) que não devem ser mantidos neste composto (itens "ruins"). A eliminação destes itens terá como consequência aumentar o coeficiente de fidedignidade do composto.

A construção de um compósito tem como pressuposto a **homogeneidade das componentes**, ou seja, **os itens devem ser indicadores do mesmo construto**. Por exemplo, em um teste de conhecimento de física cada item deve ser um indicador do conhecimento em física. Esta obviedade parece às vezes ignorada por alguns professores; envolvem na construção das questões aspectos que nada têm a ver com o objetivo precípua de uma prova de conhecimento; parecem querer em algumas questões medir a astúcia ou a capacidade de decifrar charadas.

A suposição da homogeneidade das componentes permite demonstrar que o escore total é uma medida mais confiável, mais fidedigna do construto que os escores nos itens. Essa última afirmativa, que é mais um teorema da teoria da medida, é facilmente aceita de forma intuitiva.

A maneira de realizar uma análise de consistência interna começa por aplicar o instrumento em muitos indivíduos. Os especialistas em teoria da medida sugerem que uma análise de consistência interna, para dar reais subsídios no sentido de melhorar o instrumento, deve ser conduzida com no mínimo um número de respondentes igual a 5 vezes o número de itens.²

O Anexo 1 apresenta o desenrolar de uma análise de consistência interna com todos os passos definidos e exemplificados. Por simplicidade, este exemplo fictício se constitui em um compósito que tem inicialmente 8 componentes (itens) e apenas 31 respondentes. Cada componente é uma variável que pode assumir 5 valores (1, 2, 3, 4, 5) e, portanto, o escore total pode ser no máximo igual a 40 e no mínimo 8.

Há muitas maneiras de se estimar o coeficiente de fidedignidade do compósito ou do escore total no instrumento. No exemplo do Anexo 1, ele é estimado através do coeficiente alfa de Cronbach (Cronbach, 1951). Outra possibilidade de se es-

²A justificativa desta recomendação é bastante técnica, não cabendo detalhá-la aqui; constitui-se em uma precaução contra a obtenção de correlações entre itens e total espuriamente altas, que acontecem quando são semelhantes os números de itens e respondentes.

timar o coeficiente de fidedignidade é por calcular o coeficiente beta (Silveira, 1985).

Estes dois coeficientes aumentam quando crescem os coeficientes de correlação de cada componente com o escore total e quando aumenta o número de componentes (itens). Desta forma a medida de um construto para o qual os indicadores são "ruins" (se correlacionam fracamente com o construto) pode ser "boa" (apresentar grande fidedignidade), desde que se utilize muitos indicadores, muitos itens.

A forma de depurar um instrumento, identificando os itens "ruins", pode ser conduzida a partir do cálculo dos coeficientes de correlação de cada item com o escore total (coeficiente de correlação item-total). No Anexo 1 mostra-se como se proceder esses cálculos. Os itens que apresentarem correlações mais baixas ou negativas com o escore total deverão ser eliminados ou revisados, alterados, substituídos por outros que deverão ser aplicados em um futuro estudo de consistência interna. A versão "final" de um instrumento é usualmente fruto de diversos estudos sucessivos de consistência interna.

A Tabela 2 apresenta a síntese de uma análise de consistência interna, baseada nas respostas de 121 alunos de FIS 182 (UFRGS) após o estudo da eletrodinâmica. Os alunos responderam a um teste (Anexo 2) constituído inicialmente por 16 itens; essa versão com 16 itens foi antecedida por outra que havia sido proposta principalmente a partir de entrevistas clínicas sobre concepções relativas à corrente elétrica. O instrumento tem como objetivo identificar se o aluno possui a concepção científica sobre corrente elétrica ou possui alguma(s) concepção(ões) alternativa(s) (Silveira, Moreira e Axt, 1989).

Neste exemplo cada item gera uma variável binomial, uma variável com apenas dois valores possíveis; o aluno recebe escore 1 em um item se a sua resposta coincidir com a alternativa coerente com a concepção científica; recebe escore zero se a sua escolha é qualquer outra alternativa. O exemplo numérico do Anexo 1 é mais complicado pois ali os itens geraram va-

Tabela 2

Síntese da análise de consistência interna do instrumento sobre as concepções relativas à corrente elétrica utilizando 121 alunos de FIS 182 (UFRGS).

Item	Frequências nas alternativas			p	r_{IT}
	A	B	C		
1	13	2	106*	88%	0,46
2	83*	32	5	69%	0,74
3	31	85*	5	70%	0,73
4	75*	38	7	62%	0,61
5	9	91*	21	75%	0,75
6	8	85*	28	70%	0,71
7	49	26	46*	38%	0,43
8	33	34	54*	45%	0,56
9	105*	15	1	87%	0,56
10	4	113*	4	93%	0,25
11	3	11	107*	88%	0,46
12	16	46*	59	38%	0,66
13	101*	11	9	84%	0,18
14	36	46*	37	38%	0,66
15	95*	9	15	79%	0,53
16	16	35	70*	58%	0,58

p - proporção das respostas coerente com a concepção científica.

r_{IT} - coeficiente de correlação item-total.

* - frequência das respostas coerentes com a concepção científica.

Observação: para o item em que a soma das três frequências é menor do que 121, a diferença se deve às omissões de resposta.

riáveis com 5 valores possíveis. Este exemplo é facilmente generalizável para outras situações mais complicadas onde os itens geram variáveis com valores extremos diferentes; pode também ser particularizado para itens binomiais. A análise

de consistência é aplicável a qualquer compósito, não importando se os itens ou questões que geram os escores componentes são "objetivos" (a resposta dada é escolhida entre diversas opções oferecidas) ou de resposta livre (a resposta é formulada pelo respondente). Aplica-se também a compósitos onde as componentes são resultados de julgamentos de diversos juízes sobre os mesmos objetos, ou ainda, as componentes são medidas temporalmente afastadas, como os resultados de provas de rendimento escolar ao longo de um semestre.

A última coluna da Tabela 2 apresenta os coeficientes de correlação item-total. Todos os itens, exceto 10 e 13, apresentam correlação maior do que +0,40 com o escore total. Decidiu-se então por eliminar esses dois itens e o teste ficou constituído por 14 itens.³ A Tabela 1, que relaciona os escores totais no teste com a média final em FIS 182, faz parte dos estudos de validade do teste. O coeficiente de correlação ali encontrado, +0,69, revela validade preditiva para o teste e se constitui em uma parte do argumento de validade de construto.

A Tabela 3 apresenta as estimativas do coeficiente de fidedignidade para o escore total no teste (coeficientes alfa e beta) antes e após a eliminação dos itens 10 e 13.

Observa-se que a eliminação dos itens 10 e 13 determinou um pequeno incremento nas estimativas de fidedignidade; isto se deveu a que os itens se correlacionavam positivamente com o escore total. Se a correlação item-total fosse negativa, o incremento seria maior.

³A baixa correlação de um item com o escore total pode ter diversas causas, desde simples problemas de redação ou má formulação, até um equívoco na escolha do item como indicador do construto.

No caso dos itens 10 e 13, a baixa correlação possivelmente se deva a que a resposta baseada na concepção científica é a mesma que respostas baseadas em algumas concepções alternativas. Essa suposição decorre de um exame do conteúdo dos dois itens e é apoiada empiricamente pela grande incidência de respostas "corretas", respectivamente 93% e 84%, conforme a Tabela 2.

Tabela 3

Características do escore total no teste sobre concepções relativas à corrente elétrica utilizando 121 alunos de FIS 182 (UFRGS).

Versão do teste	Média	Variância	Coeficiente	
			Alfa	Beta
16 itens	10,8	14,9	0,86	0,87
14 itens	9,0	14,1	0,87	0,88

Caso um instrumento seja constituído por itens que, do ponto de vista teórico, sejam indicadores de **diferentes construtos** ou de **dimensões diferentes de um mesmo construto** (itens não-homogêneos), a análise de consistência interna deve ser conduzida separadamente para cada grupo de itens. Técnicas de análise fatorial (Guilford, 1971; Nunnally, 1967; Mulaik, 1972) podem ser utilizadas em casos como esses. Um exemplo relacionado com a medida das concepções alternativas e newtoniana sobre força e movimento pode ser encontrado em Silveira, Moreira e Axt (1992); outro exemplo, relacionado à medida da audibilização em crianças, pode ser encontrado em Golbert e Silveira (1988).

É importante ainda destacar que o coeficiente de fidedignidade é afetado pela **heterogeneidade entre os indivíduos na variável medida pelo instrumento**; quanto **mais** heterogêneo for o grupo de respondentes, **maior** será o coeficiente de fidedignidade. Desta forma o coeficiente de fidedignidade não é uma propriedade exclusiva do instrumento, mas uma propriedade do instrumento para um determinado grupo, para uma determinada situação; expressa a fração da variabilidade observada entre os indivíduos (variância observada) que é verdadeira (variância verdadeira), isto é, não atribuível a erros de medida. Quando o coeficiente é próximo da unidade, significa que o instrumento está sendo capaz de detectar muito bem as diferenças interindividuais no grupo investigado.

O valor mínimo aceitável para um coeficiente de fidedignidade dependerá da utilização que se fará com os escores gerados pelo instrumento. Quando se deseja discriminar acuradamente entre os diferentes indivíduos é necessário um coeficiente alto (em torno de 0,9); uma típica situação onde tal deveria acontecer é em uma prova de seleção ao estilo do concurso vestibular à universidade. Quando se deseja utilizar os escores gerados pelo instrumento para comparar grupos em **médias**, pode-se tolerar coeficientes mais baixos (da ordem de 0,7). Para maiores detalhes sobre fidedignidade na comparação entre grupos consultar Silveira (1981b).

Ao conduzir um estudo de consistência interna o pesquisador deverá optar por realizá-lo em um grupo onde, a priori, espera encontrar uma razoável variabilidade interindividual.

Considerações finais

A validade de um instrumento de medida sempre é relativa à situação na qual o estudo de validade foi conduzido. Não existe "A VALIDAÇÃO do instrumento" e sim evidências de validade para o instrumento relativamente a um determinado grupo de respondente, a uma específica situação. A validação do teste sobre concepções relativas à corrente elétrica foi conduzida com alunos universitários cursantes de disciplinas de Física Geral. Estes estudos permitem aprofundar a validade do teste para esse tipo de respondente. Certamente que o teste não é válido para detectar as concepções de indivíduos analfabetos por razão óbvia; talvez ele sirva para alunos de 2º grau.

As estatísticas utilizadas em estudos de validade, como quaisquer estatísticas, flutuam de uma amostra para outra da mesma população e podem efetivamente ser diferentes em amostras de diferentes populações. Desta forma, os estudos de validade são em princípio generalizáveis para aquela população na qual se colheu a amostra. Em qualquer ciência

não há afirmações indubitavelmente definitivas; todo o nosso conhecimento é virtualmente provisório e aberto à crítica, à reformulação (Popper, 1982).

Há necessidade de se proceder a novos estudos de validade quando o instrumento for utilizado em circunstâncias diferentes dos estudos anteriores. Isto também ocorre no domínio das ciências naturais; não existe, por exemplo, um voltímetro válido sempre. Um voltímetro que for adequado para fazer medidas na rede elétrica de uma residência, onde as diferenças de potencial elétrico são da ordem de centenas de Volt, não poderá ser utilizado na medida da diferença de potencial elétrico de Hall que é da ordem de mili ou micro Volt.

Referências bibliográficas

- CAMPBELL, D.T. e FISKE, D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, **56**: 81-105, 1959.
- COOPER, K.H. *O método de Cooper*. São Paulo, Edibolso, 1976.
- CRONBACH, L.J. Coefficient alpha and the internal structure of tests. *Psychometrika*, **16**: 297-334, 1951.
- _____. *Essentials of psychological testing*. New York, Harper, 1973.
- CRONBACH, L.J. e MEEHL, P.E. Construct validity in psychological tests. *Psychological Bulletin*, **52**: 281-302, 1955.
- GHISSELLI, E.E. *Theory of psychological measurement*. Bombay, TataMcGraw-Hill, 1964.
- GOLBERT, C. e SILVEIRA, F.L. Medida da audibilização em crianças em fase inicial da aprendizagem da leitura: construção e validação de um instrumento. *Educação e Seleção*, **17**: 99-113, 1988.
- GUILFORD, J.P. *Psychometric methods*. Bombay, Tata-McGraw-Hill, 1971.
- GUILFORD, J.P. e FRUCHTER, B. *Fundamental statistics in psychology and education*. New York, Mc-Graw-Hill, 1973.
- LORD, F.M. e NOVICK, M.R. *Statistical theories of mental test scores*. London, Addison-Wesley, 1968.
- MULAİK, A. *The foundations of factor analysis*. New York, McGraw-Hill, 1972.

NUNNALLY, J.C. *Psychometric theory*. New York, McGraw-Hill, 1967.

POPPER, K.R. *Conhecimento objetivo*. São Paulo, Edusp, 1975.

_____ *Conjecturas e refutações*. Brasília, Ed. Un. Brasília, 1982.

SHAW, M.E. e WRIGHT, J.M. *Scales for the measurement of attitudes*. New York, McGraw-Hill, 1967.

SILVEIRA, F.L. Relação entre duas estatísticas utilizadas em análise de consistência interna de testes psicométricos. *Ciência e Cultura*, **32**(2): 214-217, 1980.

_____ Relação entre os índices de discriminação de itens em testes psicométricos e duas outras estatísticas: variância do escore total e coeficiente de fidedignidade. *Ciência e Cultura*, **33**(2): 246-248, 1981a.

_____ Fidedignidade das medidas e diferenças entre grupos em psicologia e educação. *Ciência e Cultura*, **33**(5): 704-706, 1981b.

_____ Medida da atitude em relação à solução de problemas. *Revista Brasileira de Física*, **12**(3): 553-560, 1982.

_____ Coeficiente beta: estimativa do coeficiente de fidedignidade de uma variável compósita. *Educação e Seleção*, **11**:105-108, 1985.

_____ Três critérios para ponderação de itens em testes de aproveitamento escolar e seus efeitos sobre o coeficiente de fidedignidade. *Educação e Seleção*, **14**: 47-54, 1986.

_____ Avaliação do ensino: os enfoques objetivo e qualitativo. *Educação e Seleção*, **20**: 63-77, 1989.

SILVEIRA, F.L. e MOREIRA, M.A. Avaliação do desempenho do professor pelo aluno: evidências de validade de um instrumento. *Ciência e Cultura*, **36**(3): 466-472, 1984.

SILVEIRA, F.L., MOREIRA, M.A. e AXT, R. Validação de um teste para verificar se o aluno possui concepções científicas sobre corrente elétrica em circuitos simples. *Ciência e Cultura*, **41**(11): 1129-1133, 1989.

_____ Estrutura interna de testes de conhecimento em física: um exemplo em mecânica. *Enseñanza de las Ciencias*, **10**(2): 187-194, 1992.

SILVEIRA, F.L., MOREIRA, M.A. e NUNES, A.D. Avaliação do desempenho do professor pelo aluno: novas evidências de validade de um instrumento. *Ciência e Cultura*, **37**(8): 1237-1240, 1985.

VIANNA, H.M. *Testes em educação*. São Paulo, Ibrasa, 1973.

WHERRY, R.J. *Contributions to correlational analysis*. London, Academic Press, 1984.

Anexo I

ANÁLISE DE CONSISTÊNCIA INTERNA: PROCEDIMENTO DETALHADO

- 1^a Etapa: transformar a resposta de cada indivíduo em cada item em um escore.
- 2^a Etapa: calcular o escore total de cada indivíduo.
- 3^a Etapa: ordenar os indivíduos, em ordem decrescente, pelo escore total.
- 4^a Etapa: construir a matriz dos resultados (vide matriz em anexo).
- 5^a Etapa: determinar a frequência de cada escore em cada item (vide matriz em anexo).
- 6^a Etapa: calcular a média e a variância do escore total.
- a) $\Sigma T = 37 + 34 + 34 + \dots + 21 + 18 + 15 = 868$
 - b) $\Sigma T^2 = 37^2 + 34^2 + 34^2 + \dots + 21^2 + 18^2 + 15^2 = 24986$
 - c) Média: $\bar{T} = \Sigma T/N = 868/31 = 28,00$
 - d) Variância: $V_T = \Sigma T^2/N - (\bar{T})^2 = 24986/31 - (28,00)^2 = 22,00$
- 7^a Etapa: Calcular a média e a variância de cada item. Exemplificando com o item 1:
- a) $\Sigma I = (1 \times 2) + (2 \times 2) + (3 \times 9) + (4 \times 13) + (5 \times 5) = 110$
 - b) $\Sigma I^2 = (1^2 \times 2) + (2^2 \times 2) + (3^2 \times 9) + (4^2 \times 13) + (5^2 \times 5) = 424$
 - c) Média: $\bar{I} = \Sigma I/N = 110/31 = 3,548$
 - d) Variância: $V_I = 424/31 - (3,548)^2 = 1,089$

Matriz dos resultados

Item	1	2	3	4	5	6	7	8	Total
Indivíduo									
1. João	5	5	5	5	5	5	5	2	37
2. Teresa	5	4	4	5	3	5	5	3	34
3.	5	5	4	5	3	4	5	3	34
4.	4	4	5	4	4	5	5	2	33
5.	5	4	5	5	2	4	4	3	32
6.	4	5	4	4	3	5	5	2	32
7.	4	4	5	5	5	5	3	1	32
8.	4	5	5	4	4	4	5	1	32
9.	3	4	5	4	3	5	5	2	31
10.	4	3	4	5	2	5	5	3	31
11.	5	5	5	4	1	4	3	4	31
12.	4	4	3	4	5	4	4	3	31
13.	4	4	4	3	3	3	4	4	29
14.	4	4	4	4	2	4	3	3	28
15.	3	4	4	4	3	4	4	2	28
16.	3	4	5	3	2	4	5	2	28
17.	4	3	3	5	4	3	2	4	28
18.	4	4	4	3	3	5	4	1	28
19.	3	3	4	4	4	4	3	2	27
20.	4	4	3	3	2	3	4	4	27
21.	4	3	3	3	4	3	4	3	27
22.	3	4	3	3	4	4	4	2	27
23.	3	4	3	3	1	3	4	5	26
24.	3	3	3	4	3	4	3	3	26
25.	2	3	3	3	5	4	3	1	24
26.	3	2	3	3	4	3	4	2	24
27.	1	2	3	2	5	3	3	5	24
28.	4	3	3	3	4	1	1	4	23
29.	2	2	2	1	4	3	3	4	21
30.	3	3	1	2	3	2	2	2	18
31.	1	1	2	1	3	1	2	4	15
F(1)	2	1	1	2	2	2	1	4	
F(2)	2	3	2	2	5	1	3	10	
F(3)	9	8	11	10	10	8	8	8	
F(4)	13	14	9	10	9	12	10	7	
F(5)	5	5	8	7	5	8	9	2	

F - frequência

Item	1	2	3	4	5	6	7	8
Média	3,548	3,613	3,710	3,548	3,290	3,742	3,774	2,774
Variância	1,089	0,947	1,057	1,211	1,251	1,159	1,159	1,272

8ª Etapa: calcular o somatório das variâncias nos itens.

$$\Sigma V_I = 1,089 + 0,947 + \dots + 1,272 = 9,145$$

9ª Etapa: calcular o coeficiente de fidedignidade (coeficiente alfa de Cronbach).

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\Sigma V_I}{V_T}\right) = 8/7(1 - 9,145/22,00) = 0,668$$

As etapas 10 a 13 são utilizadas para se obter uma **estimativa** dos coeficientes de correlação item-total. É possível substituí-las pelo cálculo exato destes coeficientes; um coeficiente de correlação item-total é o coeficiente de correlação de Pearson entre os escores no item e o escore total; desta forma os cálculos envolvidos nas etapas 10 a 13 poderão ser substituídos pelos usuais cálculos dos coeficientes de correlação.

10ª Etapa: dividir os indivíduos em três grupos (superior, médio, inferior) através do escore total. Os grupos superior e inferior usualmente perfazem aproximadamente 25% dos indivíduos.

Os indivíduos 1 a 8 constituem o grupo superior e os indivíduos 25 a 31 o grupo inferior.

11ª Etapa: calcular a média de cada item no grupo superior e no grupo inferior. Exemplificando com o item 1:

a) Média do item 1 no grupo superior:

$$\bar{I}_S = (5 + 5 + 5 + 4 + 5 + 4 + 4 + 4)/8 = 4,500$$

b) Média do item 1 no grupo inferior:

$$\bar{I}_I = (2 + 3 + 1 + 4 + 2 + 3 + 1)/7 = 2,286$$

Item	1	2	3	4	5	6	7	8
Média GS	4,500	4,500	4,625	4,625	3,625	4,625	4,625	2,125
Média GI	2,206	2,286	2,429	2,143	4,000	2,429	2,571	3,143

12ª Etapa: calcular o índice de discriminação de cada item. Exemplificando com o item 1:

$$D_I = \bar{I}_S - \bar{I}_I = 4,500 - 2,286 = 2,214$$

Item	1	2	3	4	5	6	7	8
D_I	2,214	2,214	2,196	2,482	-0,375	2,196	2,054	-1,018

13ª Etapa: calcular o coeficiente de correlação item-total. Exemplificando o item 1:

$$r_{IT} = D_I / (K \times \sqrt{V_I}) = 2,214 / (2,57 \times \sqrt{1,089}) = 0,826$$

Observação: a constante que aparece na equação acima (K) depende da extensão dos grupos extremos; ela é 2,57 no caso atual. Para outras proporções dos grupos extremos consultar Silveira (1980). A constante K também pode ser estimada através da seguinte equação (Silveira, 1981a):

$$K = \Sigma D_I \sqrt{V_T}$$

Utilizando-se equação acima para se estimar K obtém-se:

$$K = (2,214 + 2,214 + \dots + 2,054 - 1,018) / \sqrt{22,00} = 2,55$$

Item	1	2	3	4	5	6	7	8
r_{IT}	0,826	0,885	0,831	0,878	-0,130	0,794	0,742	-0,351

Caso os coeficientes de correlação sejam calculados exatamente através da equação que define o coeficiente de correlação de Pearson, obtém-se os seguintes resultados:

Item	1	2	3	4	5	6	7	8
r_{IT}	0,779	0,813	0,823	0,856	-0,012	0,805	0,715	-0,287

14ª Etapa: verificar se existem itens com r_{IT} próximo de zero ou negativo. Em caso positivo, eliminar estes itens. No presente caso, os itens 5 e 8 devem ser eliminados.

15ª Etapa: recalculando o escore total de cada indivíduo eliminando os itens deficientes. No presente caso, tendo sido eliminados os itens 5 e 8, o escore total do 1º indivíduo passa a ser 30; o do 2º 28 e assim por diante.

16ª Etapa: repetir as etapas 6, 8 e 9 com o objetivo de encontrar o novo coeficiente de fidedignidade. Repetidas as etapas 6, 8 e 9 chega-se finalmente, neste exemplo, a um coeficiente de fidedignidade de 0,904.

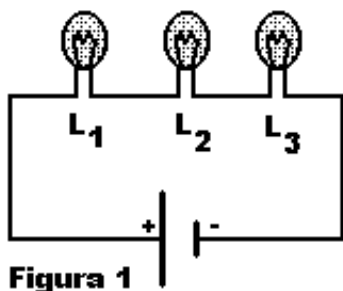
Anexo II

UM TESTE PARA VERIFICAR SE O ALUNO TEM CONCEPÇÕES CIENTÍFICAS SOBRE CORRENTE ELÉTRICA EM CIRCUITOS SIMPLES

Teste sobre corrente elétrica em circuitos simples*

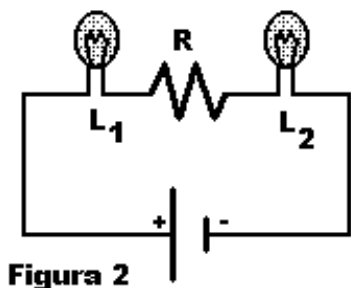
Em todas as questões deste teste admite-se que as lâmpadas sejam iguais. Os brilhos das lâmpadas crescem quando a intensidade da corrente elétrica aumenta. A bateria representada tem resistência elétrica desprezível.

1) No circuito da figura 1 pode-se afirmar que:



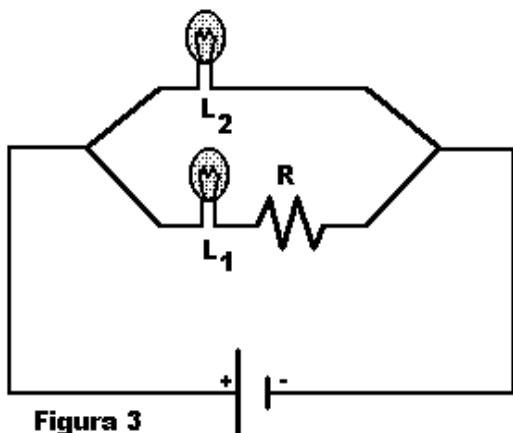
- a) L_1 brilha mais do que L_2 e esta mais do que L_3 .
- b) L_3 brilha mais do que L_2 e esta mais do que L_1 .
- c) as três lâmpadas têm o mesmo brilho.

2) No circuito da figura 2, R é um resistor. Neste circuito:



- a) L_1 e L_2 têm o mesmo brilho.
- b) L_1 brilha mais do que L_2 .
- c) L_2 brilha mais do que L_1 .

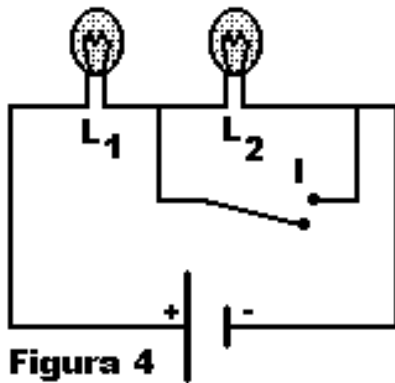
3) No circuito da figura 3, R é um resistor. Neste circuito:



- a) L_1 tem o mesmo brilho de L_2 .
- b) L_2 brilha mais do que L_1 .
- c) L_1 brilha mais do que L_2 .

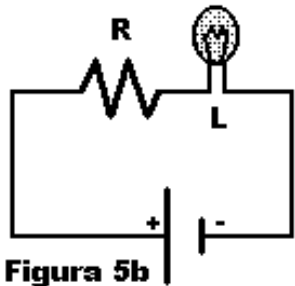
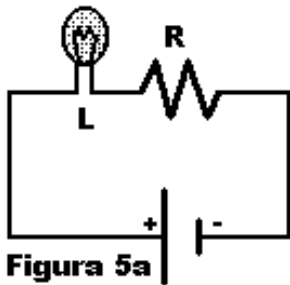
* - SILVEIRA, F. L., MOREIRA, M. A. e AXT, R. Validação de um teste para verificar se o aluno possui concepções científicas sobre corrente elétrica em circuitos simples. *Ciência e Cultura*, São Paulo, 41(11): 1129–1133, nov. 1989.

4) No circuito da Figura 4, I é um interruptor aberto. Ao fechá-lo:



- a) aumenta o brilho de L_1 .
- b) o brilho de L_1 permanece o mesmo.
- c) diminui o brilho de L_1 .

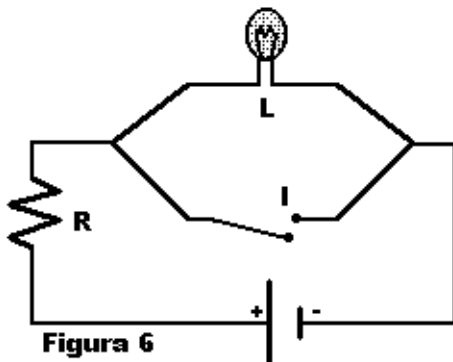
5) Nos circuitos 5a e 5b a lâmpada L, o resistor R e a bateria são exatamente os mesmos. Nestas situações:



- a) L brilha mais no circuito 5a.
- b) L brilha igual em ambos circuitos.
- c) L brilha mais no circuito 5b.

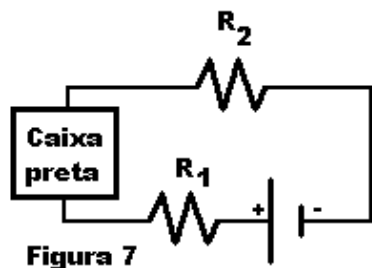
SILVEIRA, F. L., MOREIRA, M. A. e AXT, R. Validação de um teste para verificar se o aluno possui concepções científicas sobre corrente elétrica em circuitos simples. *Ciência e Cultura*, São Paulo, 41(11): 1129–1133, nov. 1989.

6) No circuito da figura 6, R é um resistor e I é um interruptor que está aberto. Ao fechar o interruptor:



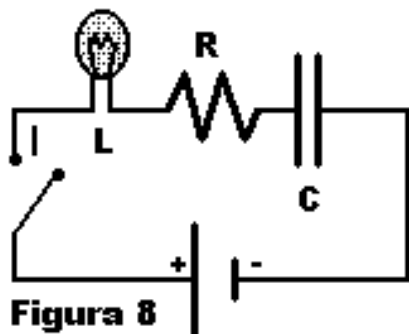
- a) L continua brilhando como antes.
- b) L deixa de brilhar.
- c) L diminui seu brilho mas não apaga.

7) No circuito da figura 7 R_1 e R_2 são dois resistores. A caixa preta pode conter resistores, baterias ou combinações de ambos. Para que a intensidade da corrente em R_1 fosse igual à intensidade da corrente em R_2 a caixa preta:



- a) deveria conter somente resistores.
- b) deveria conter no mínimo uma bateria.
- c) poderia conter qualquer associação de resistores e baterias.

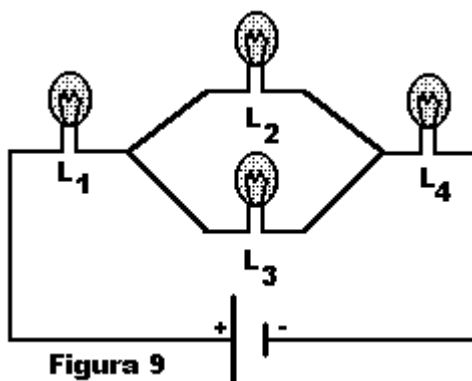
8) No circuito da figura 8, L é uma lâmpada, R um resistor, C um capacitor descarregado e I um interruptor aberto. Ao fechar o interruptor:



- a) L começa a brilhar e continua brilhando enquanto o interruptor estiver fechado.
- b) L não brilhará enquanto o capacitor não estiver carregado.
- c) L poderá brilhar durante parte do processo de carga do capacitor.

As questões 9 e 10 se referem ao circuito da figura 9.

SILVEIRA, F. L., MOREIRA, M. A. e AXT, R. Validação de um teste para verificar se o aluno possui concepções científicas sobre corrente elétrica em circuitos simples. *Ciência e Cultura*, São Paulo, 41(11): 1129–1133, nov. 1989.



9) No circuito da figura 9 o brilho de L_1 é :

- a) igual ao de L_4 .
- b) maior do que o de L_4 .
- c) menor do que o de L_4 .

10) No circuito da figura 9 o brilho de L_1 é :

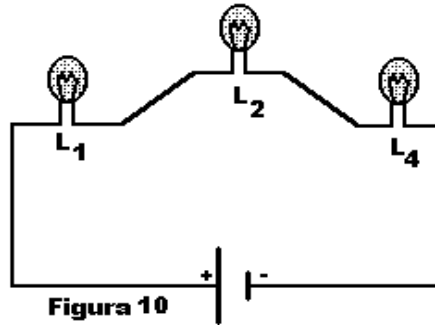
Eliminada na versão final do

- a) igual ao de L_2 .
- b) maior do que o de L_2 .
- c) menor do que o de L_2 .

11) No circuito da figura 9 o brilho de L_2 é:

- a) igual ao de L_4 . b) maior do que o de L_4 . c) menor do que o de L_4 .

O circuito da figura 9 foi modificado pois se tirou a lâmpada L_3 . O novo circuito é, então, o da figura 10.



SILVEIRA, F. L., MOREIRA, M. A. e AXT, R. Validação de um teste para verificar se o aluno possui concepções científicas sobre corrente elétrica em circuitos simples. *Ciência e Cultura*, São Paulo, 41(11): 1129–1133, nov. 1989.

12) Quando se compara o brilho de L_1 nos circuitos 9 e 10 ele é:

- a) maior no circuito 10. B) menor no circuito 10. C) o mesmo nos dois.

13) Quando se compara o brilho de L_2 nos circuitos 9 e 10 ele é:

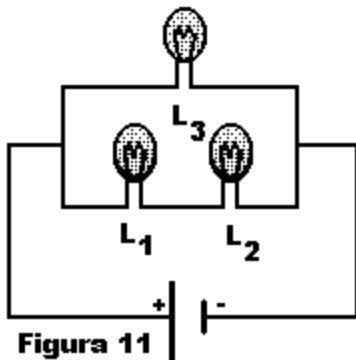
Eliminada na versão final do

- a) maior no circuito 10. B) menor no circuito 10. C) o mesmo nos dois.

14) quando se compara o brilho de L_4 nos circuitos 9 e 10 ele é:

- a) maior no circuito 10. B) menor no circuito 10. C) o mesmo nos dois.

15) No circuito da figura 11:



a) L_1 e L_2 têm o mesmo brilho que é menor do que o de L_3 .

b) L_1 brilha mais do que L_2 e do que L_3 .

c) L_1, L_2 e L_3 brilham igualmente.

16) No circuito da figura 12, quando o interruptor é aberto, as lâmpadas L_3 e L_4 deixam de brilhar, embora L_2 brilhe. O que acontece com as lâmpadas L_1 e L_5 ?

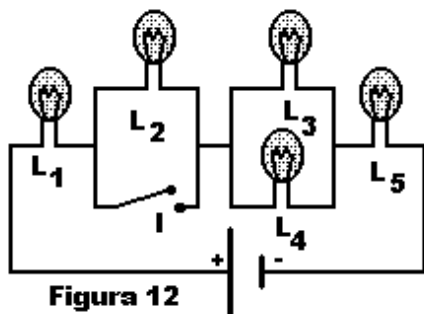


Figura 12

- a) nem L_1 , nem L_5 brilham.
- b) L_1 brilha e L_5 não brilha.
- c) L_1 e L_5 brilham.