



Ensaio Pesquisa em Educação em Ciências  
ISSN: 1415-2150  
ensaio@fae.ufmg.br  
Universidade Federal de Minas Gerais  
Brasil

Araújo, Nelci Reis Sales de; Laburú, Carlos Eduardo  
UMA ANÁLISE DA VALIDAÇÃO E CONFIABILIDADE DA ESCALA DE OPINIÕES DA SELEÇÃO DE  
EXPERIMENTOS DE QUÍMICA (EOSEQ)  
Ensaio Pesquisa em Educação em Ciências, vol. 11, núm. 2, diciembre, 2009  
Universidade Federal de Minas Gerais  
Minas Gerais, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=129512606002>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

## UMA ANÁLISE DA VALIDAÇÃO E CONFIABILIDADE DA ESCALA DE OPINIÕES DA SELEÇÃO DE EXPERIMENTOS DE QUÍMICA (EOSEQ)

### AN ANALYSIS OF THE VALIDATION AND RELIABILITY OF THE OPINION SCALE FOR CHEMISTRY EXPERIMENT SELECTION (EOSEQ)

Nelci Reis Sales de Araújo\*  
Carlos Eduardo Laburú\*\*

#### Resumo

O propósito desta investigação foi de explorar, interpretar e descrever o processo de validação e confiabilidade da escala de opiniões da seleção de experimentos, que proporciona medidas das convicções dos professores quando escolhem os experimentos. A escala do tipo *Likert* foi administrada a 113 participantes. Para a validação do conteúdo, utilizamos as técnicas do pré-teste e teste-piloto até a versão final que pelo nosso consenso com peritos da área, alcançamos bons resultados. As técnicas da forma paralela e *Spearman* verificaram a estabilidade no tempo, indicando resultados positivos de confiabilidade. O alfa de *Cronbach* item total de 0,80, 0,85 e 0,7 para geral, os professores e licenciandos, respectivamente, considerando a compreensão e variabilidade dos itens pertinentes ao serem interpretáveis.

**Palavras-chave:** Validade de Conteúdo. Confiabilidade. Instrumento de Pesquisa.

#### Abstract

The purpose of this investigation was to explore, interpret and describe the process of validation and the reliability of the opinion scale for select experiments which provides measures of the convictions of the teachers when they choose the experiments. The scale of type *Likert* was administered to 113 participants. To validate the content, we used the pretests and pilot test techniques until the final version that from our consensus with experts in the area, we obtained good results. The parallel form and Spearman techniques verified stability in time, indicating reliable positive results. The Cronbach alpha item totaled 0.80, 0.85 and 0.7 for general, the teachers and undergraduate teachers, respectively, considering comprehension and variability of the pertinent items to be interpreted.

**Keywords:** Content Validity. Reliability. Research Instrument.

---

\* Mestre em Ensino de Ciências e Educação Matemática, professora do Departamento de Química, Universidade Estadual de Londrina e da SEED-Pr, [nelquimica@yahoo.com.br](mailto:nelquimica@yahoo.com.br)

\*\* Departamento de Física, Universidade Estadual de Londrina, [laburu@uel.br](mailto:laburu@uel.br), apoio CNPq e Fundação Araucária

## 1. CONSIDERAÇÕES INICIAIS

As pesquisas de educação científica realizadas nos últimos anos, apontam a utilização de dados quantitativos, como recursos que contribuem para a investigação das preocupações na área de ensino de ciências. A exemplo, temos os trabalhos de Ginsberg e Panasuk (1998); Hudson e Skamp (2002); Hudson (2005) e Hazari *et al.* (2003) para a Física; Koul e Rubba (1999); Tuan *et al.* (2005); Caetano e Neto (2005) com professores de ciências; Laburú e Sales de Araújo (2006) para a Química, que administraram questionários estruturados do tipo Likert para investigar questões de ensino. Nesta última pesquisa, elaboramos o instrumento de pesquisa que denominamos de Escala de Opiniões da Seleção de Experimentos de Química (EOSEQ) para ilustrar as convicções dos professores a partir das opiniões envolvidas na escolha de experimentos ou equipamentos para o ensino de ciências.

O instrumento privilegia a natureza e os recursos de ensino de ciências das atividades experimentais, dadas pelas clássicas categorias Motivacional, Funcional, Instrucional e Epistemológica de Laburú (2005). Nesta perspectiva, temos o propósito de realizar a validação e confiabilidade, já que tais medidas refletem a qualidade do instrumento. O objetivo é explorar e interpretar as medidas dos itens que elaboramos a partir das categorias, para validar a série de dados coletados pelo instrumento, por meio de técnicas específicas de administração e análise, culminando na compreensão dos itens que englobam questões relacionadas às convicções dos professores de ciências. Também incluímos uma discussão a respeito da perspectiva didático-pedagógica adotada para a análise das atividades experimentais no ensino de ciências e os dados relativos às razões apresentadas pelos professores e licenciandos para a realização de atividades experimentais no ensino.

Na elaboração dos itens, utilizamos como referência as quatro categorias utilizadas em Laburú (2005) e Sales de Araújo (2006) para analisar os objetivos e finalidades dos laboratórios de ciências e que estendem às clássicas categorias dos trabalhos de Séré (2002) e que foram ampliadas, para que pudessem englobar os itens que descrevemos. Os autores fazem reflexões a respeito do laboratório didático no ensino de Física e Química, respectivamente, que há muito, estão sendo registradas,

destacando que, a partir do final dos anos setenta, a produção literária acerca deste assunto cresceu substancialmente nos mais variados pontos de vista, desde o Ensino Fundamental até o Ensino Médio. A pesquisa investigou os motivos dados pelos professores para a seleção de experimentos, procurando identificar e conhecer os padrões de discursos que eles mantêm. Os autores exploraram, identificaram e caracterizaram as 16 convicções da natureza das atividades e ensino de ciências, freqüentemente encontradas nas opiniões que mais parecem influenciar o pensamento dos professores para a escolha de experimentos. Na investigação de 2006, com objetivo de verificarmos a importância relativa desses itens, que elaboramos um instrumento de pesquisa estruturado e padronizado, fundamentado na escala desenvolvida por Rensis Likert (1932) de cinco pontos, sendo os escores: (5) concorda fortemente, (4) concorda, (3) indiferente, (2) discorda e (1) discorda fortemente. A escala realiza a mensuração de dados intangíveis, que por meio de testes estatísticos de correlação, pode-se determinar ou identificar o nível de relação entre eles. O enunciado construtor do instrumento “quando seleciono experimento de laboratório, penso...”, já deixa claro, de imediato, aos respondentes que os enunciados contêm os itens que estão baseados na seleção de experimentos, respondendo para cada item seu grau de concordância ou discordância, conforme seus motivos ou convicções de opção de experimentos.

O instrumento de pesquisa administrado proporcionou uma leitura de dados observáveis, mantendo vínculo com a realidade na qual foi planejado, conforme Freitas *et al.*, (1998a, b) e Straub (1989). Para os autores, é a validação e a confiabilidade do instrumento que permite a generalização dos resultados obtidos e consequentemente, a aplicação da mesma metodologia para amostras diferentes, mas com o mesmo perfil. A validade de conteúdo evidencia se o conteúdo de cada item do instrumento avalia os domínios a que se propõe. Para Straub (1989) a validação é o processo de revisão do universo de conteúdos, por meio de alguns pesquisadores da área, que avaliam versões do instrumento, até alcançarem um consenso e foi preconizado por Cronbach (1951).

O coeficiente de confiabilidade é um indicador da precisão das pontuações observadas dos sujeitos investigados, no sentido do grau de confiança e é estimado por procedimentos empíricos (BROWN, 2002; MUÑIZ *apud* GOMEZ, 2006). Os

procedimentos baseados na correlação entre duas distribuições de pontuações dos instrumentos de pesquisas são realizados por três técnicas. A técnica de forma equivalente (paralela ou alternativa), que conforme Garson (1999) mede a confiabilidade, pelas medidas idênticas e variações iguais dos erros; a técnica do teste e reteste que verifica a estabilidade, que é o tempo excedente da consistência, observando as flutuações das respostas dos pesquisados, como consequência do espaço de tempo, entre uma administração e outra (FREITAS *et al.*, 1998a; BROWN, 2001, 2002; DEL SIEGLE, 2006; MUÑIZ *apud* GOMEZ, 2006) e a técnica do teste-reteste com formas equivalentes que é realizada por meio da combinação desses dois procedimentos.

Os procedimentos que requerem uma só administração do instrumento de pesquisa, são realizados por duas técnicas baseadas nas correlações entre as diversas partes do teste: entre duas metades, entre todos os itens. A técnica baseada nas covariâncias entre os itens prediz o entendimento total ou parcial dos respondentes aos itens do instrumento, porque pode acontecer que, diante de uma afirmação, os sujeitos tenham diversas opiniões, ou ainda ocorrer que os itens tenham alta variabilidade nas respostas. Não porque estejam confusas, ou mesmo gerem diversas interpretações, mas porque os participantes têm várias opiniões para cada uma delas. Esse problema é chamado de consistência interna e um dos principais testes que medem esse parâmetro é o alfa de Cronbach (ANASTASI, 1977; BROWN, 2002; SANTOS, 1999). O coeficiente  $\alpha$  de Cronbach mede o grau de covariância de uma série de itens e varia de 0 a 1 e, quanto mais elevada a contagem, maior a confiabilidade da escala. Um valor de pelo menos 0,7 reflete uma fidedignidade aceitável, conforme Nunnaly (1978), embora alguns valores inferiores sejam aceitos na literatura (BROWN, 2002; SANTOS, 1999), como Garson (1999) que considera as correlações dos inter-itens superiores a 0,6.

A consistência interna também pode ser verificada pela técnica da correlação entre duas metades de um único teste. Nesse caso, os itens são considerados paralelos e as correlações entre as partes do instrumento são calculadas e os valores corrigidos, mediante a fórmula de *Spearman-Brown*, que estima uma correlação usando o coeficiente de correlação Produto-Momento de Pearson, uma medida estatística do grau de relacionamento entre as duas metades, ou seja, uma correção desse coeficiente

na fórmula (BROWN, 2002; LARSON; FARBER, 2004; FREITAS *et al.*, 1998a, b) e, seu valor, segundo Muñiz *apud* Gomez (2006) representa o grau em que o teste é homogêneo.

## 2. METODOLOGIA DE PESQUISA

O instrumento de coleta de dados do tipo escala Likert estabelece uma escala numérica, na qual a resposta favorável recebe o valor mais alto e a mais desfavorável recebe o valor mais baixo, caracterizando a escala ordinal ou por postos. A escala por postos, segundo Del Siegle (2006), ocorre quando os elementos na categoria de uma dada escala não são apenas diferentes dos elementos em outras categorias da mesma escala, mas também guardam certo tipo de relação com eles e pode ser verificada por correlações dos postos de *Spearman*.

Alguns programas *softwares* como o *Microsoft Excel, spreadsheet* do *Excel* da Del Siegle e o *STATISTICA Release 6* executam as correlações. Esse último contém as correlações tipicamente não-paramétricas, por exemplo, coeficiente de *Spearman* que é um coeficiente padronizado. Os dois últimos programas calculam as correlações da consistência interna ( $\alpha$  de *Cronbach*) e da confiabilidade *split-half* (coeficiente *Spearman-Brown*).

Esses programas foram utilizados nesta investigação que contou com uma amostra total de 113 participantes, sendo 67 professores de Química do Ensino Médio de escolas públicas de algumas cidades do Paraná e 46 licenciandos em formação da Universidade Estadual de Londrina no período de 2004 à 2006. Esse total foi distribuído de acordo com as etapas da investigação, pois alguns participaram de uma fase, mas não na seguinte. Isso ocorreu devido ao nosso acesso aos participantes, não porque eles não desejavam ou se negaram, a responder nossas perguntas, ao contrário, todos atenderam prontamente e expressaram suas opiniões. Para a análise das técnicas da administração única do instrumento de pesquisa realizamos com 58 professores e 32 licenciandos. A amostra foi considerada propícia à administração do instrumento de pesquisa, o suficiente para obtermos as informações desejadas, pois esses docentes realizam atividades experimentais, embora alguns executem mais que outros, mas não consideramos isso, como fator decisivo para que respondessem ao instrumento.

A preparação do instrumento de pesquisa passou por várias etapas, pois utilizamos múltiplos métodos que permitissem responder ao nosso propósito de validação e confiabilidade pela complementação de dados no fenômeno de nosso interesse, para uma maior compreensão da natureza das atividades e recursos de ensino de ciências para a seleção de experimentos. Para explicitar essa fenomenologia, administrarmos a investigação em duas fases descritas abaixo.

## **2.1 Fase I: Pré-teste e Teste-Piloto**

No pré-teste elaboramos 20 itens com base nos motivos que classificamos nas categorias Motivacional, Funcional, Instrucional e Epistemológica e administramos o questionário à uma amostra de 18 professores e 20 licenciandos, entre o final de 2004 e início de 2005. A intenção foi colocar o projeto de conteúdo dos itens a uma prova qualitativa dos respondentes. Durante essa etapa deixamos os participantes à vontade para responderem, conforme a compreensão dos enunciados, porém observamos que alguns deles não comprehendiam o conteúdo de algumas sentenças, o que possibilitou que reformulássemos alguns itens. Os 20 itens, alguns com as ressalvas, foram novamente colocados à prova qualitativa e quantitativa, perfazendo o teste-piloto, para a realização da técnica da forma equivalente, após a prova qualitativa de conteúdos dada pela validação dos avaliadores da área. Esta fase foi planejada para facilitar a revisão, direcionando para um instrumento formalmente validado, conforme os dizeres de Freitas *et al.* (1998a, b), que consideram o teste-piloto como uma experiência com o instrumento a uma pequena amostra do público de objetivo, no qual, podemos modificar algo antes de coletarmos os dados. O teste-piloto foi administrado a 51 participantes, sendo 33 professores e 18 licenciandos, a fim de localizar possíveis debilidades no questionário. Cada versão do instrumento refletiu melhorias sugeridas pela banca examinadora da área com nosso consenso, até o ponto dos itens serem continuamente reeditados para a próxima versão.

Para garantirmos a generalização dos resultados obtidos, utilizamos os preceitos de validade e confiabilidade, conforme sugeriram Straub (1989) e Freitas *et al.* (1998a, b), exploramos a validade de conteúdo e do construto dirigente, conforme preconizado por Cronbach (1951). Os itens foram submetidos a 4 avaliadores

independentes pertencentes à área, possibilitando tanto a revisão de 20 itens, previamente elaborados e aplicados por teste-piloto, como a re-elaboração de 4 itens que complementaram os que reformulamos, até o nosso consenso. A escala foi considerada por todos os juízes abrangente e o conteúdo representativo para mensuração das categorias.

De posse da versão final do instrumento de pesquisa, com cabeçalho indicativo do propósito de nossa investigação e da forma de respostas, solicitamos a 4 professores da área, que classificassem os 16 itens nas categorias. Para isso fornecemos uma síntese com as características das categorias. Como critério mínimo de aceitação, estabelecemos um valor igual ou superior a 80% de concordância entre os membros julgadores, como sugere Vianna (1982). Os resultados apontaram 90% de concordância, comprovando que os itens têm fidedignidade tanto em termos de conteúdo das categorias, quanto no resultado pela eficiência da sua administração. Os itens das categorias foram organizados neste instrumento, de forma que não aparecem na ordem de enumeração de uma mesma categoria. Cada item representa uma categoria diferente, o que possibilita aos participantes não terem seus pensamentos direcionados em uma seqüência lógica de uma determinada categoria. Portanto, procuramos dar aleatoriedade aos itens, conforme exibido abaixo. (inserir questionário)

## 2.2 Fase II: Validação Pelas Técnicas

As técnicas de validação do instrumento aconteceram durante os últimos meses de 2005 e final do primeiro semestre de 2006. O propósito era realizar a validação e confiabilidade do construto e foram utilizados métodos dissimilares, com intenção semelhante ao da triangulação na pesquisa em geral, dada pelos pesquisadores Cook e Campbell *apud* Straub (1989).

O teste-piloto aplicado anteriormente foi considerado razoavelmente paralelo para realizar a técnica em que obtemos o *coeficiente de equivalência*. Realizamos o procedimento do teste e reteste, mas como não tínhamos a intenção de aplicar o instrumento prévio e sim, a versão final, optamos então por também realizar o teste e reteste em forma equivalente e coeficiente de estabilidade da forma paralela pelo

cálculo da correlação de *Spearman* para encontrarmos o *coeficiente de estabilidade e equivalência*, conforme sugeridos por Freitas *et al.*, 1998a; Brown, 2001, 2002; Muñiz *apud* Gomez, 2006. Em ambas as técnicas, utilizamos a mesma amostra, porque os 51 participantes do teste-piloto, também responderam a versão final. A aplicação dos instrumentos foi realizada com um intervalo de 2 a 3 meses, conforme o tempo requerido para distribuição e recolhimento. Na técnica de teste e reteste pode ocorrer medidas de erros aleatórios. A principal fonte de erro que pode ameaçar a confiabilidade é a mudança na pontuação dos indivíduos, devido às flutuações que podem ocorrer nos mesmos, como resultado da passagem do tempo entre uma administração e outra (GOMEZ, 2006). O indivíduo pode recordar o conteúdo, um problema chamado de efeito de memória que o sujeito tem acerca das respostas que deu na primeira administração, portanto, ocasiona o aumento da confiabilidade (NUNNALLY, 1978). Ainda, segundo Anastasi (1977) o sujeito pode ter melhorado ou piorado notadamente em relação aos demais do grupo. Na técnica da forma paralela, a fonte de erros é a falta de equivalência entre as formas do instrumento e não é adequada para itens estáveis (GOMEZ, 2006).

A administração do instrumento final foi realizada uma única vez com 58 professores e 32 licenciandos. Mas, consideramos as estratégias que prevêem o grau de generalização das opiniões que, nesse caso, são dadas pelas consistências das respostas. As técnicas para verificarmos a consistência interna principalmente, pelo parâmetro do alfa de *Cronbach* foram utilizadas para obtermos um instrumento confiável. Determinarmos os coeficientes da consistência interna geral, para os professores e para os licenciandos. Também adotamos a estratégia das duas metades de um único teste. A subdivisão em duas metades foi feita, agrupando-se de um lado os itens pares e de outro os ímpares. Utilizamos o cálculo da correlação das metades, aplicando a fórmula de *Spearman-Brown*, conforme sugerido por Brown, 2002; Larson e Farber, 2004; Freitas *et al.*, 1998a,b; Muñiz *apud* Gomez, 2006.

### **3. RESULTADOS E DISCUSSÕES**

Os resultados obtidos utilizando os formulários equivalentes realizado com 51 pessoas, comparado com a versão final do instrumento, pelo uso do *Microsoft*

*Excel*, forneceram um coeficiente de correlação Produto-Momento de Pearson. As correlações foram calculadas para cada item antes e depois e são apresentadas na Tabela 1. Os cálculos das médias dos coeficientes inter-itens pelas correlações dos itens antes e depois de uma dada categoria, em relação aos itens antes e depois de outra categoria são apresentados na Tabela 2.

TABELA 1 - Coeficiente de Correlação Produto-Momento de Pearson Antes e Depois

Categoria Motivacional		Categoria Funcional		Categoria Instrucional		Categoria Epistemológica	
Item	r	Item	r	Item	r	Item	r
3	0,1522	1	0,1722	4	0,2627	2	0,0911
6	0,3004	5	0,3249	8	0,1149	7	0,3665
10	0,4361	9	0,2815	12	0,1651	11	0,0299
13	0,6467	14	0,3540	15	0,3809	16	0,4204

Fonte: Resultados da investigação. Elaborada pelos autores

TABELA 2 - Matriz de Correlação com Valores Antes e Depois para o Coeficiente de Correlação Produto-Momento de Pearson

C. antes C.depós	Categoria Motivacional antes	Categoria Funcional antes	Categoria Instrucional antes	Categoria Epistemológica antes
Categoria Motivacional depois	r 0,3839	r 0,1116	r 0,0805	r 0,0985
Categoria Funcional depois	r 0,1680	r 0,2832	r 0,0871	r 0,0255
Categoria Instrucional depois	r 0,2694	r 0,2374	r 0,2309	r 0,3196
Categoria Epistemológica depois	r 0,2187	r 0,2349	r 0,1371	r 0,2270

Fonte: Resultados da investigação. Elaborada pelos autores

Os resultados são positivos, no geral, obtivemos coeficientes em um alcance aceitável. Porém, nem todas as variáveis tiveram um coeficiente razoável, pois há alguns valores indicadores de relações entre itens com tendência moderada. Contudo, NunNaly (1978) afirma que seria pouco realístico esperar uma correlação de coeficientes extraordinariamente alta e, para Anastasi (1977), os coeficientes de 0,20 e acima podem ser significativos. Segundo Del Siegle (2006) se não houver nenhuma correlação, podemos dizer que uma variável não causou a outra, pois as correlações servem como indicações empíricas de relacionamentos possíveis entre variáveis e se um investigador, casualmente, descobre uma correlação, não significa que provou a existência de um relacionamento causal.

O coeficiente para a forma equivalente dos testes, verificado pelos itens antes e depois, forneceu um  $r$ -barra = 0,2812 e foi calculado no *Microsoft Excel*. O dado foi gerado pela correlação produto-momento de Pearson para cada item antes e depois, e dessas 16 correlações, calculamos a média ( $r$ -barra). Esse valor resultou no alfa de Cronbach igual a 0,86 quando usamos o  $N$  = 16, contudo, ao dobrarmos esse valor, como sugere Brown (2001, 2002) obtemos um  $\alpha$  = 0,93.

A média das correlações para cada categoria está de acordo com os resultados encontrados no trabalho de Laburú e Sales de Araújo (2005, 2006), que em pesquisa prévia investigaram as razões dos professores para a seleção de experimentos e baseados nos dados mais significativos para o ensino de ciências, elaboraram a escala de opiniões em questão. Na última pesquisa realizada com 90 indivíduos e a versão final do instrumento de pesquisa, encontraram que as categorias Instrucional, Funcional e Motivacional, com intensidades de 94% e 89% para as duas últimas, são as categorias priorizadas pelos professores, e que as categorias Motivacional e Instrucional, com 85% e 82%, respectivamente, são as categorias mais importantes para os licenciandos em formação.

Nas Tabelas 1 e 9 o item 13 “Naquele em que os estudantes vão gostar ou se entusiasmar com a observação do resultado do experimento” da categoria Motivacional apresentou a maior correlação. A variação indivíduos é pequena entre os escores, sugerindo que entenderam o enunciado do item e há relação com as suas opiniões. Ao contrário, o item 3 “Naquele que desperta a curiosidade dos estudantes pela novidade e os mantêm mobilizados durante a tarefa” apresentou uma maior variabilidade nos escores, indicando uma possível indecisão e ao mesmo tempo, minimizando o erro de efeito memória no índice de confiabilidade. O  $r$  moderado do item é explicado pela baixa correlação com o item 9 “Naquele que tem características simples e fácil de montar e trabalhar e que dá o resultado esperado e confiável” da categoria Funcional, indicando que a simplicidade, facilidade e reproduzibilidade de medidas do experimento escolhido, não influencia a opinião do docente de manter o fator curiosidade na motivação, como pode ser observado pelo  $r$  da Tabela 2.

A questão da segurança no laboratório é considerada como muito importante por 72% dos professores e 56% dos licenciandos, uma vez que escolhem experimentos que mantêm condições de serem realizados com segurança, conforme

encontramos em Laburú e Sales de Araújo (2006), reforçando a alta correlação. No tocante ao item 1 “No tempo de sua realização” há variabilidade entre os escores 3, 4 e 5, mostrando uma relação menor com as opiniões dos pesquisados.

A correlação do item “Naquele que ilustra, aplica (vivencia), descobre e reforça o conteúdo científico trabalhado em sala de aula” pode ser um indicativo de que há dúvidas para os participantes de que o experimento selecionado deva ilustrar, vivenciar ou, ainda, reforçar o conteúdo. Na pesquisa citada acima, o item mostrou as freqüências de 55% e 53% no escore 5 da escala para os professores e licenciandos, respectivamente. A correlação maior do item “Naquele em que o conteúdo faz conexões com a tecnologia ou cotidiano” indica uma maior relação com as opiniões de 53% dos licenciandos, já que observamos uma variabilidade das respostas dos professores. Contudo, 60% de docentes e 65% de discentes com respostas de escore 5, mantêm uma melhor relação com o item 4, justificando o *r* significativo.

A menor correlação do item “Naquele que confronta argumentos rivais, mitos e lendas e favorece os desejados pelas ciências” da categoria Epistemológica mostra uma alta variabilidade entre os escores 1 a 5 da escala. O resultado está de acordo com a pesquisa citada acima, mostrando que o item é considerado com escore 5 somente para 24% e 9,4% dos professores e discentes, respectivamente. A mesma pesquisa, somente com o instrumento final, apontou que a variação entre os escores 2, 3 e 4 é maior para os itens “Naquele que prova fatos para os estudantes acreditarem” e “Naquele que faz conexões entre as teorias científicas aprendidas pelos estudantes com o mundo”. Mas, como mantiveram as opiniões, as correlações são significativas.

Os resultados da Tabela 2 apontam uma maior correlação inter-itens para a categoria Motivacional, seguida da categoria Funcional e a mesma correlação para as categorias Instrucional e Epistemológica. A menor correlação entre as categorias Motivacional e Funcional indica que os itens da segunda não interferem na intenção de motivar dos professores ao escolher o experimento. Mas, é interessante notar que eles consideram importantes os itens da categoria Instrucional.

O coeficiente de confiabilidade do instrumento de medição em pauta pode ser considerado confiável. Todavia, é de nosso interesse observar que as inter-correlações entre as categorias foram menores que entre as categorias, indicando que elas são,

razoavelmente, diferentes entre si e cada uma delas contribui substancialmente, para a confiabilidade geral do instrumento de pesquisa.

A consistência interna ( $\alpha$  de Cronbach) e a confiabilidade *split-half* (coeficiente Spearman-Brown) para a técnica equivalente foram estimadas, ainda, com a utilização do *spreadsheet* do *Excel*, desenvolvido pela Del Siegle (2006), no qual encontramos o  $\alpha = 0,89$  para os 32 itens (16 antes e 16 depois), a Correlação *Split-Half (odd-even)* = 0,70 e também, o Spearman-Brown Prophecy igual a 0,80. Os mesmos resultados também foram obtidos no software *STATISTICA Release 6*, no qual gerou o resumo estatístico item-total exibido na Tabela 3 e o sumário da Confidabilidade *Split-Half* na Tabela 4. O valor da correlação entre as duas metades foi 0,7, com correções resultou no mesmo valor 0,8, fornecido para a predição de Spearman-Brown no programa da Del Siegle (2006).

TABELA 3 - Estatística Item-total da Técnica Paralela ou Equivalente

Número de itens nas escalas: 32/ Número de casos válidos: 51/ Número de casos com dados perdidos: 0/ Dados perdidos que foram deletados: casewise	
<b>RESUMO ESTATÍSTICO DA ESCALA</b>	
Média: 140,29411765	Soma: 7155,0000000
Desvio Padrão: 11,681256983	Variância: 136,45176471
Cronbach's alpha: ,890397934	Alfa Padronizado: ,891241315
Média da Correlação Inter-Itens: ,211277179	

Fonte: Resultados da Investigação. Dados Obtidos no Statistica Release 6

TABELA 4 - Consistência Interna (bipartição) da Técnica Paralela ou Equivalente

	Primeira Metade	Segunda Metade
Número de Itens	16	16
Cronbach's alpha	,839607008	,804071868
Correlação entre a 1 <sup>a</sup> e 2 <sup>a</sup> metade	,664104611	
Atenuação para a Correlação Correta	,808259903	
Confiabilidade de Split half	,798152480	
Confiabilidade de Guttman split-half	,797293880	

Fonte: Resultados da Investigação. Dados Obtidos no Statistica Release 6

Uma vez que os valores de fidelidade tendem a coincidir quando calculados, seja pelo método de Spearman-Brown (0,7982), seja pelo método de Guttman (0,7973), inferimos que a consistência interna desta escala é 0,80, conforme sugere Miguel (2004). Essa consistência interna encontrada é a mesma que obtemos quando realizamos a média (r-barra) das correlações inter-itens antes e depois das categorias e

aplicamos na fórmula. Os resultados desses procedimentos foram considerados adequados por nós, pois verificamos que o teste-piloto que elaboramos, previamente, estava consistente com o tema de nossa investigação e possibilitou a revisão final dos conteúdos dos itens. O estudo mostrou considerável estabilidade na produção dos itens. Isto dá sustento à validade de construto e sugere que estes itens representam características relativamente comuns das opiniões dos pesquisados.

A consistência interna e a confiabilidade *split-half* para a técnica teste e reteste foram estimadas com a utilização do software *STATISTICA Release 6*. Para o teste paralelo, encontramos um  $\alpha = 0,84$ , conforme exibido na Tabela 5. Os valores da Correlação *Split-Half (odd-even)* = 0,9023 e de *Guttman* = 0,8974, como apresentados, na Tabela 6, indicam que a consistência interna é 0,90. Para o teste atual, encontramos um  $\alpha = 0,80$ , de acordo com o resumo da Tabela 7, fornecido pelo programa. Os valores da Correlação *Split-Half (odd-even)* = 0,8003 e de *Guttman* = 0,7951 podem ser observados na Tabela 8.

**TABELA 5 - Estatística Item-total da Técnica Teste e Reteste (Teste Paralelo)**

Número de itens nas escalas: 16/ Número de casos válidos: 51/ Número de casos com dados perdidos: 0/ Dados perdidos que foram deletados: casewise	
<b>RESUMO ESTATÍSTICO DA ESCALA</b>	
Média: 71,137254902	Soma: 3628,0000000
Desvio Padrão: 6,593996081	Variância: 43,480784314
Cronbach's alpha: ,839607008	Alfa Padronizado: ,842116707
Média da Correlação Inter-Itens: ,258763401	

Fonte: Resultados da Investigação. Dados Obtidos no Statistica Release 6

**TABELA 6 - Consistência Interna (bipartição) da Técnica Teste e Reteste (Teste Paralelo)**

	Primeira Metade	Segunda Metade
Número de Itens	8	8
Cronbach's alpha	,642607279	,746086443
Correlação entre a 1 <sup>a</sup> e 2 <sup>a</sup> metade		,821983967
Atenuação para a Correlação Correta		--
Confiabilidade de Split half		,902295500
Confiabilidade de Guttman split-half		,897434972

Fonte: Resultados da Investigação. Dados Obtidos no Statistica Release 6

**TABELA 7 - Estatística Item-total da Técnica Teste e Reteste (Teste Atual)**

Número de itens nas escalas: 16/ Número de casos válidos: 51/ Número de casos com dados perdidos: 0/ Dados perdidos que foram deletados: casewise	
<b>RESUMO ESTATÍSTICO DA ESCALA</b>	

Média: 69,156862745	Soma: 3527,0000000
Desvio Padrão: 6,210869662	Variância: 38,574901961
Cronbach's alpha: ,804071868	Alfa Padronizado: ,801653001
Média da Correlação Inter-Itens: ,210138869	

Fonte: Resultados da Investigação. Dados Obtidos no Statistica Release 6

TABELA 8 - Consistência Interna (bipartição) da Técnica Teste e Reteste (Teste Atual)

	Primeira Metade	Segunda Metade
Número de Itens	8	8
Cronbach's alpha	,625052235	,713671583
Correlação entre a 1 <sup>a</sup> e 2 <sup>a</sup> metade	,667073483	
Atenuação para a Correlação Correta	,998771909	
Confiabilidade de Split half	,800292836	
Confiabilidade de Guttman split-half	,795112132	

Fonte: Resultados da Investigação. Dados Obtidos no Statistica Release 6

A consistência interna pelo cálculo do alfa de *Cronbach* também foi verificada pelo *software spreadsheet* do *Excel* da Del Siegle, que forneceu os mesmos valores para os alfas no procedimento de teste e reteste. A confiabilidade que verificamos pelos métodos de *split-half* e *Guttman* foi considerada satisfatória em todos os 16 itens do teste paralelo e do instrumento de pesquisa atual.

A Tabela 9 apresenta as médias das escalas que os professores e licenciandos (51 participantes) avaliaram antes, no teste paralelo, e depois, no instrumento de pesquisa, bem como o método da correlação de *Spearman* para cada item em ambas as etapas. O nível de significância escolhido para análise foi “alpha” = 5%, onde o valor p é menor do que 0,05 e os valores obtidos apresentam uma correlação significativa. Os valores observados indicam que há correlação significativa entre os itens. A correlação dos itens antes e depois, dá apoio à hipótese de que há uma consistência entre todos eles, consistência esta que reflete um aspecto duradouro da opinião mantenedora dos pesquisados para uma série de mensurações. Assim, entendemos que, os resultados da comparação do teste com o reteste nos permitiram assumir a estabilidade da escala no tempo.

TABELA 9 - Média das Escalas Atribuídas pelos Professores e Licenciandos aos Itens do Teste Paralelo Comparado com o Instrumento Atual pelo Cálculo da Correlação de Spearman

Itens	N1	N2	N3	N4	N5	Média	Spearman
1-paralelo	1	1	5	8	36	4,51	0,3177*
1-atual	0	1	2	15	33	4,57	
2-paralelo	1	0	2	21	27	4,43	0,1411
2-atual	0	1	1	21	28	4,49	
3-paralelo	0	1	6	18	26	4,35	0,3095*
3-atual	1	0	2	13	35	4,59	
4-paralelo	0	0	3	12	36	4,65	0,3239*
4-atual	0	1	2	15	33	4,57	
5-paralelo	2	1	5	12	31	4,35	0,3412*
5-atual	0	0	7	17	27	4,39	
6-paralelo	0	0	2	19	30	4,54	0,2161
6-atual	0	1	3	22	25	4,39	
7-paralelo	0	0	7	11	33	4,51	0,3630*
7-atual	2	2	16	18	13	3,74	
8-paralelo	0	0	2	10	39	4,72	0,2051
8-atual	1	0	2	20	28	4,45	
9-paralelo	0	1	11	23	16	4,06	0,3580*
9-atual	0	3	13	18	17	3,96	
10-paralelo	0	0	6	13	32	4,59	0,4305*
10-atual	0	1	3	12	35	4,59	
11-paralelo	0	3	8	16	24	4,20	0,0279
11-atual	2	2	17	20	10	3,67	
12-paralelo	0	0	3	17	31	4,55	0,2191
12-atual	1	1	7	20	22	4,20	
13-paralelo	0	0	10	17	24	4,27	0,6887*
13-atual	0	1	10	19	21	4,18	
14-paralelo	0	0	3	9	39	4,71	0,3672*
14-atual	0	0	4	14	33	4,57	
15-paralelo	1	2	3	15	30	4,39	0,4596*
15-atual	0	0	6	19	26	4,39	
16-paralelo	1	0	4	20	26	4,37	0,4299*
16-atual	0	0	3	24	24	4,41	

Fonte: Resultados da investigação. Elaborada pelos autores. Dados obtidos no Statistica Release 6.

\* Correlação de Spearman significativa  $p < 0,05$ .

Os resultados da consistência interna do instrumento de pesquisa foram obtidos a partir das correlações inter-itens, calculados no Microsoft Excel. Da matriz geral, o r-barra = 0,1886 que resultou em  $\alpha = 0,8$  é considerado de boa consistência, portanto, aceitável e interpretável, já que é uma boa escala pelos dizeres de Cronbach (1951, 1970) e Straub (1989). O r-barra para o grupo de professores é de 0,2633,

fornecendo um  $\alpha = 0,85$ . Para o grupo de licenciandos o r-barra é 0,1013, portanto, o  $\alpha = 0,64$ . Os mesmos resultados foram computados pela correlação do programa *spreadsheet* do *Excel*, no qual, encontramos para o grupo de discentes o  $\alpha = 0,66$ . Da mesma forma, também calculamos pelo software *STATISTICA Release 6*, no qual encontramos os mesmos resultados, conforme exibidos nas Tabelas 10, 11 e 12. Estas informações indicam que os itens mostraram-se bastante consistentes. Corroboramos com Brown (2001, 2002) que considera um bom valor o alfa de *Cronbach* = 0,8 para um número pequeno de itens em um instrumento do tipo Likert, pois o alfa é maior quanto mais itens contiver um questionário.

**TABELA 10 - Estatística Item-total do Instrumento de Pesquisa**

Número de itens nas escalas: 16/ Número de casos válidos: 90/ Número de casos com dados perdidos: 0/ Dados perdidos que foram deletados: casewise	
<b>RESUMO ESTATÍSTICO DA ESCALA</b>	
Média: 68,711111111	Soma: 6184,0000000
Desvio Padrão: 6,316101138	Variância: 39,893133583
Cronbach's alpha: ,795794006	Alfa Padronizado: ,788106432
Média da Correlação Inter-Itens: ,193962941	

Fonte: Resultados da Investigação. Dados Obtidos no Statistica Release 6

**TABELA 11 - Estatística Item-total do Instrumento de Pesquisa para os Professores**

Número de itens nas escalas: 16/ Número de casos válidos: 58/ Número de casos com dados perdidos: 0/ Dados perdidos que foram deletados: casewise	
<b>RESUMO ESTATÍSTICO DA ESCALA</b>	
Média: 69,862068966	Soma: 4052,0000000
Desvio Padrão: 6,239231497	Variância: 38,928009679
Cronbach's alpha: ,849398065	Alfa Padronizado: ,851135701
Média da Correlação Inter-Itens: ,271710893	

Fonte: Resultados da Investigação. Dados Obtidos no Statistica Release 6

**TABELA 12 - Estatística Item-total do Instrumento de Pesquisa para os Licenciandos**

Número de itens nas escalas: 16/ Número de casos válidos: 32/ Número de casos com dados perdidos: 0/ Dados perdidos que foram deletados: casewise	
<b>RESUMO ESTATÍSTICO DA ESCALA</b>	
Média: 66,625000000	Soma: 2132,0000000
Desvio Padrão: 5,998655763	Variância: 35,983870968
Cronbach's alpha: ,663977290	Alfa Padronizado: ,643325621
Média da Correlação Inter-Itens: ,107945582	

Fonte: Resultados da Investigação. Dados Obtidos no Statistica Release 6

O coeficiente de confiabilidade de coerência interna pela administração única do instrumento, também foi calculado pelo método das metades, lançando mão do coeficiente de correlação Produto-Momento de Pearson corrigido pela fórmula de *Spearman-Brown*. O coeficiente de correlação dos itens pares e itens ímpares é corrigido pela fórmula é 0,8136, demonstrando correlação bastante significativa, indicando que a escala apresenta considerável equivalência ou adequação de amostragem dos itens. Esse valor foi encontrado no *spreadsheet* do *Excel* de Siegle, no qual aplicamos a planilha com os itens já separados. Quando utilizamos a planilha sem separar os itens, a referida fórmula fornece o valor de 0,76, mostrando que nosso procedimento empírico de separação está correto.

Os dados com os itens separados e de item-total foram computados no *STATISTICA Release 6*. Lembramos que o alfa de *Cronbach* padronizado é 0,7881, em ambos os procedimentos que adotamos. A consistência interna pela bipartição, utilizando os métodos de *Spearman-Brown* e *Guttman Split-Half*, foi considerada por nós como positiva, pois, encontramos os valores de 0,7588 e 0,7505, respectivamente. Como os valores de fidelidade coincidem quando calculados pelas duas técnicas, inferimos que a consistência interna desta escala é 0,76, por ser o método de *Guttman* aquele que proporciona a estimativa mais conservadora quando, em casos como o presente, a evidência aponta para que ambas as partes da escala não têm o mesmo grau de fidelidade, conforme os dizeres de Miguel (2004). O mesmo valor foi encontrado pela *Spearman-Brown Prophecy* = 0,7588, do *spreadsheet* do *Excel* de Siegle, sem separar os itens. Como dissemos antes, encontramos 0,8136 com os itens separados, enquanto o *Split-Half (odd-even)* é 0,6858. Nesse caso, há diferenças entre os dois métodos, embora o alfa de *Cronbach* item-total, seja o mesmo = 0,8, nos dois casos, dos itens separados ou não. Contudo, para Del Siegle (2006), uma contagem total para as perguntas de número ímpar é correlacionada com uma contagem para as perguntas de número par, embora possa ser a primeira metade com a segunda metade, conforme seus escritos. E, de acordo com Brown (2001), o coeficiente *Spearman-Brown* da predição é usado ajustando as estimativas da confiabilidade do *split-half* para a confiabilidade do teste inteiro.

No mesmo sentido da técnica de *Guttman*, explicada por Miguel (2004), analisamos os dados do item-total obtidos pela aplicação da planilha total e não

fornecer o valor da atenuação para a correlação correta. Assim, inferimos para a consistência interna o valor de 0,78, pois os alfas das partes 1 e 2 são diferentes. Ambas as partes da escala têm o mesmo grau de confiabilidade, pois o alfa é 0,7 para cada parte das metades. Segundo os dizeres de Muñiz *apud* Gomez (2006), poderíamos separar de outra forma os 16 itens. Contudo, não tínhamos essa intenção, por considerarmos pertinentes os itens das categorias. Todavia, entendemos que, uma vez obtido um alfa consistente, consideramos a forma da bipartição propícia, já que a interpretação do alfa de *Cronbach* está relacionada à interpretação dada para as estimativas de confiabilidade baseadas no método *split-half*, como interpretamos acima. Isso porque o alfa é uma média (*r-barra*) de todos os coeficientes *split-half* do instrumento.

#### **4. CONSIDERAÇÕES FINAIS**

A investigação desenvolvida para o processo de validação e confiabilidade do instrumento de pesquisa realizada de acordo com as técnicas foi satisfatória. A exploração dos dados possibilitou identificarmos e descrevermos os valores que mostram a eficácia da escala que contempla a natureza e os recursos de ensino de ciências dadas pelas categorias Motivacional, Funcional, Instrucional e Epistemológica da seleção de experimentos, que consideramos pertinentes para o ensino de ciências.

O estudo exploratório foi extenso, porque queríamos assegurar os 20 itens que elaboramos e que pelo nosso consenso, com os avaliadores da área culminou em 16 itens, contudo esses 4 itens não foram eliminados e sim, acrescentados ao instrumento final. Dessa forma, não lançamos mão do método da análise de fator exploratória para selecionar os itens, tampouco, tínhamos a intenção de elaborar mais de 20 itens, já que essa técnica exige um número elevado de itens em análise para resultar no valor 16.

Os resultados alcançados apontaram um alto coeficiente de confiabilidade, que nos autorizam a dizer que o instrumento de pesquisa é considerado bom, ao ser interpretável. Entendemos que o alfa de *Cronbach* é uma ferramenta que pode ser utilizada para investigar a confiabilidade da consistência interna do instrumento de

pesquisa, bem como as técnicas que descrevemos são adequadas para a análise das correlações para o projeto de instrumentos de coleta de dados, ou ainda, que outros instrumentos, já validados, possam ser administrados para novos contextos.

A utilização do instrumento da escala de opiniões da seleção de experimentos de Química (EOSEQ) é propícia para avaliar a natureza das atividades e recursos de ensino de ciências das categorias da seleção de experimentos, até mesmo, a possibilidade da administração do instrumento para outras modalidades de levantamentos semelhantes dadas pelas categorias.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ANASTASI, A. **Testes Psicológicos**. Tradução de Dante Moreira Leite. 2. ed. São Paulo: EPU, 1977.
- BROWN, J. D. Can We Use Spearman-Brown Prophecy Formula to Defend Low Reliability? Shiken: **JALT Testing & Evaluation Sig Newsletter**, ISSN 1881-5537, v.4, n.3, p.7-9, Jan. 2001. Disponível em: <[www.jalt.org/test/bro\\_13.htm](http://www.jalt.org/test/bro_13.htm)>. Acesso em: 28 mar. 2006.
- \_\_\_\_\_. The Cronbach Alpha Reliability Estimate. Shiken: **JALT Testing & Evaluation Sig Newsletter**, ISSN 1881-5537, v. 6, n. 1, p. 16-18, Feb. 2002. Disponível em: <[www.jalt.org/test/bro\\_13.htm](http://www.jalt.org/test/bro_13.htm)>. Acesso em: 28 mar. 2006.
- CAETANO, H.; NETO, A. J. Natureza e Ensino da Ciência: Investigando as Concepções de Ciência dos Professores. **Enseñanza de las Ciencias**, Barcelona, número extra, VII congresso, 2005.
- CRONBACH, L. J. Coefficient alpha and the internal structure of the tests. **Psychometrika**, Springer New York, v.16, n.3, p.297-334, Sep. 1951.
- \_\_\_\_\_. **Essentials of psychological testing**. 3. ed. New York: Harper and Row, 1970.
- DEL SIEGLE, Ph. D. Escola de Educação da Universidade de Neag de Connecticut. Disponível em: <[www.gifted.uconn.edu/siegle/research/Instrument%20Reliability%20and%20Validity.htm](http://www.gifted.uconn.edu/siegle/research/Instrument%20Reliability%20and%20Validity.htm)>. Acesso em: 20 mar. 2006.
- FREITAS, H. M. R. *et al.* Quanti-Qualitative Instruments to Study the Decision-Making Process. **WP ISRC No. 020298**, Merrick School of Business, University of Baltimore, MD, EUA, Feb. 1998a, p.29.

\_\_\_\_\_ et al. The Design Process of a Cross-Cultural *Exploratory Quantitative - Qualitative Survey Research Project to Study the Decision-Making Process*. Association for Information Systems Americas Conference 1998. **Technology Research in Progress**, v.14, n.16, p.944-946, Aug. 1998b.

ROBERT, A. Y. **Enhancement of Reliability Analysis**: Application of Intraclass Correlations with SPSS/Windows v.8. Disponível em: <<http://www.nyu.edu/acf/socsci/Docs/intracls.html>>. Acesso em: 06 maio, 2006.

GINSBERG, E. S.; PANASUK, R. M. Survey of Physics Student Attitudes on Access to Problem Solutions. **Electronic Journal of Science Education**, University of Nevada, Reno, v.2, n.3, Mar. 1998.

GOMEZ, A. G. Procedimientos Empíricos para Estimar la Precisión de las Puntuaciones Proporcionadas por los Tests. In: MUÑIZ, J. **Teoría Clásica de los Tests**. Madrid: Pirámide, 1998, 392 p. Disponível em: <<http://www.ugr.es/~andreito/tema6.doc>>. Acesso em: 18 jun. 2007.

HAZARI, Z.; KEY, A. W.; PITRE, J. Interactive and Affective Behaviors of Teaching Assistants in a First Year Physics Laboratory. **Electronic Journal of Science Education**, University of Nevada, Reno, v.7, n.3, Mar. 2003.

HUDSON, P. Identifying Mentoring Practices for Developing Effective Primary Science Teaching. **International Journal of Science Education**, Routledge, v.27, n.14/18, p.1723-1739 (17), Nov. 2005.

\_\_\_\_\_.; SKAMP, K. Mentoring Preservice Teachers of Primary Science. **Electronic Journal of Science Education**, University of Nevada, Reno, v.7, n.1, Sep. 2002.

KOUL, R.; RUBBA, P. An Analysis of the Reliability and Validity of Personal Internet Teaching Efficacy Beliefs Scale. **Electronic Journal of Science Education**, University of Nevada, Reno, v.4, n.1 Sep. 1999.

LABURÚ, C. E. Seleção de Experimentos de Física no Ensino Médio: Uma investigação a partir da fala dos professores. **Investigações em Ensino de Ciências**, Porto Alegre, v.10, n.2, p.1-19, Ago. 2005.

LARSON, R.; FARBER, B. Tradução e revisão técnica Cyro de Carvalho Patarra. **Estatística Aplicada**. 2.ed. São Paulo: Prentice Hall, 2004.

LIKERT, R. Una Técnica para la Medicion de Atitudes. (A technique for the measurement of attitudes, Arquives of Psychology, n.140, p.1-50, 1932). In: WEINERMAN, C. H. **Escalas de Medicion en Ciencias Sociales**. Buenos Aires: Nueva Vision, p.201-260. 1976 .

MICROSOFT EXCEL **Pacote de Software**, 2005.

MIGUEL, J. P. **Análise de Fidelidade** – Consistência interna, FPCE-UC/2004. Disponível em: <[www1.fpce.uc.pt/ensino/psi/ae3/pratica/tarefa4s.pdf](http://www1.fpce.uc.pt/ensino/psi/ae3/pratica/tarefa4s.pdf)>. Acesso em: 2 maio, 2007.

NUNNALLY, J. **Psychometric Theory**. New York, NY: McGraw-Hill, 1978.

RELIABILITY CALCULATOR *spreadsheet* do Excel Criado por Del Siegle. Disponível em: <[dsiegle@uconn.edu](mailto:dsiegle@uconn.edu)>. Acesso em: 10 mar. 2006.

SALES DE ARAUJO, N. R. **Categorias para a Seleção de Experimentos de Química no Ensino Médio**: Um estudo comparativo das prioridades dos professores e licenciandos em formação. Dissertação (Mestrado em Ensino de Ciências e Educação Matemática) – Centro de Ciências Exatas, Universidade Estadual de Londrina, Londrina, 2006.

SÉRÉ, M-G. Towards Renewed Research Questions from the Outcomes of the European Project Labwork in Science Education, **Science Education**, Wiley, New York, v.86, n.5, p.624-644, Sep. 2002.

SANTOS, J. R. A. Alfa de Cronbach: Uma ferramenta para avaliar a confiabilidade das escalas. **Journal of Extension**. ISSN 1077-5315, EUA, v.37, n.2, Apr. 1999.

STATISTICA RELEASE 6. Statsoft, Inc. (2001). **STATISTICA (data analysis software system)**, version 6.<[www.statsoft.com](http://www.statsoft.com)>.

STRAUB, D. W. Validating Instrument in MIS Research. **MIS Quarterly**, University of Minnesota, v.13, n.2, p.147-169, Jun.1989.

TUAN, H. L; CHIN, C. C; SHIEH, S. H. The Development of a Questionnaire to Measure Students' Motivation Towards Science Learning. **International Journal of Science Education**, Routledge, v.27, n.6, p.639-654, May, 2005.

VIANNA, H. M. **Testes em Educação**. São Paulo: IBRASA, 1982.

## ANEXO - QUESTIONÁRIO

**Questionário de coleta de dados acerca da importância da escolha de experimentos de química para o trabalho de laboratório no ensino médio**  
Indique a escala ordinal (5 = concordo fortemente, 4 = concordo, 3 = indiferente, 2 = discordo e 1 = discordo fortemente).

**Quando seleciono experimento de laboratório, penso...**

Escala	5	4	3	2	1
No tempo de sua realização					
Naquele em que a observação é importante para a construção das teorias científicas					
Naquele que desperta a curiosidade dos estudantes pela novidade e os mantêm mobilizados durante a tarefa					

Naquele que permite uma metodologia que facilita os processos de ensino e aprendizagem			
Naquele em que os materiais/reagentes e equipamentos estão disponíveis no laboratório para sua realização, ou que ao menos os materiais são encontrados no dia-a-dia			
Naquele que é interessante, por tratar de fenômenos do cotidiano e sua tecnologia			
Naquele que prova fatos para os estudantes acreditarem			
Naquele que ilustra, aplica (vivencia), descobre e reforça o conteúdo científico trabalhado em sala de aula			
Naquele que tem características simples e fácil de montar e trabalhar e que dá o resultado esperado e confiável			
Naquele que motiva os estudantes para o conteúdo ministrado			
Naquele que confronta argumentos rivais, mitos e lendas e favorece os desejados pelas ciências			
Naquele em que a sua realização facilita a participação cooperativa dos estudantes e desenvolve habilidades e atitudes			
Naquele em que os estudantes vão gostar ou se entusiasmar com a observação do resultado do experimento			
Naquele que mantém condições de ser realizado com segurança			
Naquele em que o conteúdo faz conexões com a tecnologia ou cotidiano			
Naquele que faz conexões entre as teorias científicas aprendidas pelos estudantes com o mundo			

Data de recebimento: 20/06/07

Data de aprovação: 31/03/09

Data de versão final: 30/04/09