

ANÁLISE DA VALIDADE E FIDEDIGNIDADE DE UM TESTE DE CONTROLE DE VARIÁVEIS

ANALYSIS OF THE VALIDITY AND RELIABILITY OF A VARIABLE CONTROL TEST

Alessandro Damásio Trani Gomes¹

¹Universidade Federal de São João del-Rei/Departamento de Ciências Naturais,
alessandrogomes@ufjs.edu.br

Resumo

Este trabalho tem o objetivo de apresentar o processo de validação de um instrumento de pesquisa, desenvolvido para investigar a habilidade de alunos do Ensino Médio em controlar variáveis, aplicado a 185 alunos do 1º ano do Ensino Médio de uma escola pública federal. O teste passou por um processo de validação de conteúdo. Para se determinar a sua consistência interna, foram calculados os coeficientes alfa de Cronbach, beta de Silveira e Spearman-Brown. Por meio destes procedimentos, verificou-se a adequação do teste para avaliar a habilidade de estudantes em controlar variáveis, podendo ser utilizado como instrumento de pesquisa ou para o ensino de Física.

Palavras-chave: Validade, fidedignidade, controle de variáveis.

Abstract

This study presents the validation process of a research instrument developed to assess high school students' ability to control variables. Administered to 185 first-year students at a federal public school, the instrument underwent a content validation process. Internal consistency was assessed using Cronbach's alpha, Silveira's beta, and Spearman-Brown coefficients. The results confirmed the adequacy of the test for evaluating students' proficiency in controlling variables, and can be used as a research instrument or for teaching Physics.

Keywords: Validity, reliability; control of variable

Introdução

A ausência de discussões sobre a validade e fidedignidade dos instrumentos de pesquisa, bem como a validade dos resultados alcançados é apontada por Santos e Greca (2013) como um dos problemas da pesquisa na área de Educação em Ciências nas últimas décadas.

Ao analisar os dados originados a partir de um instrumento de pesquisa, um pesquisador precisa ter a garantia de sua qualidade. Esse instrumento só trará essa

garantia se possuir fidedignidade e validade suficientes. Segundo Vizzotto (2022), por validade, compreende-se a característica que atesta se o instrumento em questão mensura aquilo que se destina medir. Já a confiabilidade (fidedignidade) busca verificar a hipótese de que, se o mesmo instrumento fosse aplicado a um grupo semelhante em intervalos de tempo diferentes, este apresentaria desempenhos estatisticamente semelhantes.

O objetivo deste trabalho é apresentar o processo de validação de um instrumento de pesquisa, desenvolvido para investigar a habilidade de alunos do Ensino Médio em controlar variáveis. Este trabalho justifica-se, pois, em uma revisão dos últimos 30 anos de publicações sobre o ensino de Física, com enfoque no processo de medição científica na Educação Básica, Pigosso e Heidemann (2023) identificaram apenas 4 trabalhos que abordam explicitamente questões relacionadas a habilidades de controle de variáveis, sendo o mais recente de 2016.

A validade e a fidedignidade dos instrumentos de pesquisa

Segundo Moreira (2011), não adianta o pesquisador ser cuidadoso na escolha do delineamento da pesquisa se os instrumentos de coleta de dados forem inadequados. Deve-se então, observar questões relacionadas à fidedignidade (o instrumento mede corretamente o que se propõe a medir?) e à validade (o instrumento mede outra coisa diferente daquela que se propõe a medir?) para se garantir a qualidade dos dados disponíveis para análise.

Para Souza, Alexandre e Guirardello (2017, p. 650), a fidedignidade

é a capacidade em reproduzir um resultado de forma consistente no tempo e no espaço, ou a partir de observadores diferentes, indicando aspectos sobre coerência, precisão, estabilidade, equivalência e homogeneidade. Trata-se de um dos critérios principais de qualidade de um instrumento.

A determinação da fidedignidade é comumente realizada pela análise da consistência interna do teste. Pesquisas recomendam que, para o cálculo da homogeneidade e da consistência interna do teste, sejam calculados os valores dos coeficientes alfa de Cronbach e beta de Silveira (HUTZ, 2015; SILVEIRA, 1993). Pode-se também utilizar o coeficiente Spearman-Brown, quando o instrumento é dividido em duas partes equivalentes (duas metades ou itens pares e itens ímpares) de forma a separar os escores para cada respondente. Estes dois conjuntos de escores são, então, correlacionados (ARAÚJO; LABURÚ, 2009).

Para Moreira (2011), enquanto a fidedignidade é um pré-requisito para qualquer instrumento de pesquisa, a validade é a característica mais importante que ele deve possuir. Isto ocorre porque a validade está ligada ao relacionamento entre os dados obtidos e o propósito da pesquisa. Deste modo, validade é definida como o grau com que o instrumento realmente mede o que se propõe a medir. Neste trabalho, procuramos estabelecer a validade de conteúdo do instrumento. Esta técnica verifica “o grau em que cada elemento de um instrumento de medida é relevante e representativo de um específico constructo com um propósito particular de avaliação” (ALEXANDRE; COLUCI, 2011, p. 3063).

Aspectos metodológicos

A coleta de dados ocorreu em uma escola de Ensino Médio da rede federal de ensino de Belo Horizonte. Dos 191 estudantes matriculados no 1º ano, 185 responderam ao teste.

O instrumento utilizado é composto de uma série de dois esquemas experimentais que abordavam dois problemas diferentes, denominados de ‘Equilíbrio térmico’ e ‘Plano inclinado’. Cada problema consiste na avaliação de comparações entre dois desenhos distintos. Cada um dos dois desenhos ou esquemas apresenta um sistema com um conjunto particular de variáveis, que indica que variáveis da situação são alterados e quais permanecem iguais, sendo que o contraste entre cada par de situações representa um experimento.

O processo de validação de conteúdo foi realizado durante a construção do teste, por meio de pesquisa bibliográfica, buscando por instrumentos de pesquisa que também avaliassem a habilidade de reconhecer testes adequados e consistentes. O teste foi construído a partir de pesquisa anteriores (BORGES; GOMES, 2005; SILER; KLAHR, 2012).

O teste foi apresentado sob a forma de um livreto com 12 páginas, sendo uma página de instrução antes de cada situação-problema e dez contendo um esquema de um experimento em cada página (Figuras 1 e 2). Foram preparados dois tipos de livretos com a ordem de apresentação das duas situações-problema alternada. As situações-problema apresentavam estruturas similares de variáveis, apesar de tratarem de domínios teóricos bastante diferentes. Todas possuíam três variáveis independentes e categóricas que assumiam apenas dois valores. A linguagem utilizada para

descrever os problemas e as variáveis foi propositadamente coloquial. Durante a apresentação das situação-problema, optou-se pelo uso de proposições gerais, evitando-se terminologia técnica, para não sugerir hipóteses ou entendimentos específicos. Para cada uma, uma comparação experimental aleatória foi utilizada para uma explicação oral da tarefa. Em seguida, de acordo com os objetivos estabelecidos, os alunos avaliaram a adequação dos conjuntos de esquemas experimentais para testar o efeito de determinada variável sobre outra.

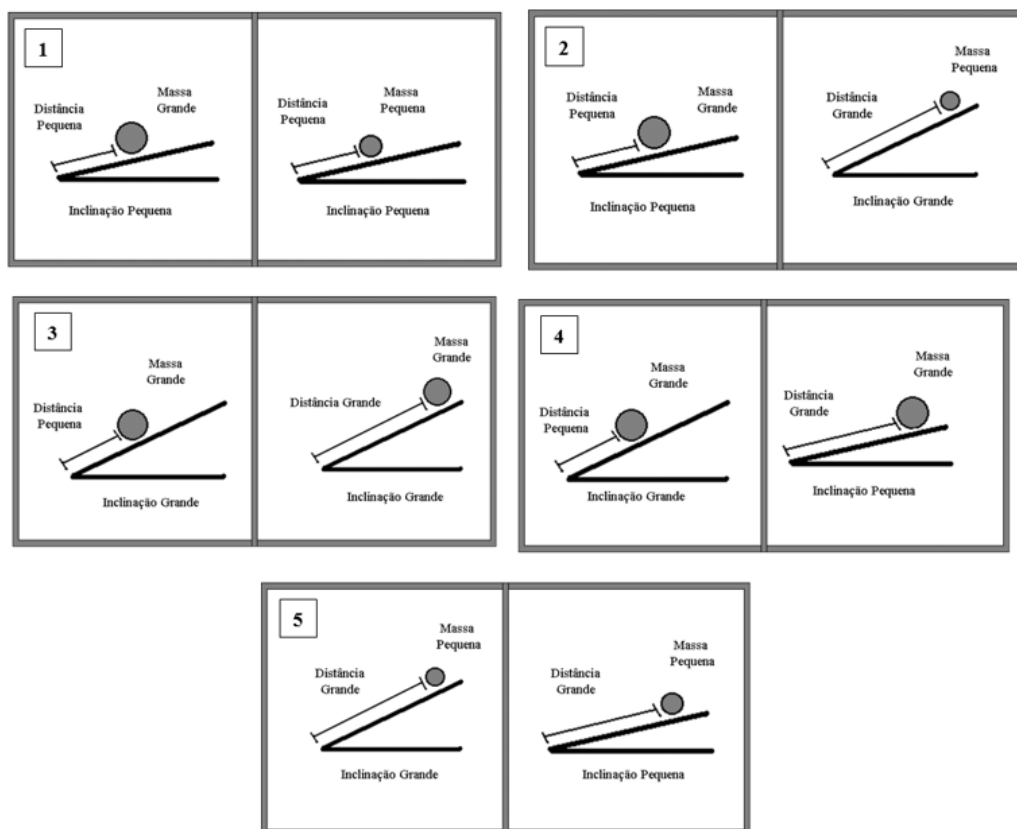
Figura 1 – Situação-problema do “Plano inclinado”

Suponha que um estudante, em uma atividade no laboratório de Física, tenha, sobre sua bancada, uma rampa cuja inclinação ele pode ajustar (duas inclinações possíveis, pequena ou grande), e que ele também dispõe de duas esferas de mesmo material, mas de massas diferentes, que podem ser soltas a partir de duas distâncias distintas com relação ao início da rampa (distância pequena ou grande).

Nas páginas a seguir, são exibidos 5 experimentos que o estudante realizou, com o objetivo **de determinar se a distância em que as esferas são soltas tem influência no tempo que as esferas gastam para percorrer a rampa.** Você deverá avaliar os experimentos que o estudante realizou, considerando a adequação entre o objetivo da atividade e o procedimento adotado. Para cada experimento, você deverá decidir se aquele experimento é bom ou ruim para alcançar o objetivo do estudante e justificar sua resposta.

Como exemplo, o experimento 1 (veja a figura) ilustra que o estudante colocou a rampa com uma inclinação pequena e soltou uma esfera de massa grande para rolar a partir de uma distância pequena do início da rampa. Quando a esfera chegou ao final da rampa, ele anotou o tempo gasto pela esfera para percorrer a distância. Em seguida, ele manteve a inclinação pequena da rampa e soltou uma esfera de massa pequena a partir de uma distância pequena e anotou o tempo. Com esse procedimento, o estudante é capaz de determinar, com certeza, se a distância em que as esferas são soltas tem influência no tempo necessário para ela descer a rampa?

Siga o mesmo raciocínio para os outros experimentos, justificando porquê você considera os experimentos bons ou ruins.



Fonte: Autoria própria.

Figura 2 – Situação-problema do “Equilíbrio térmico”

Suponha que um estudante, durante uma outra atividade no laboratório de Física, tenha, sobre sua bancada, 4 recipientes iguais (A, B, C e D), contendo a mesma quantidade de líquido:

-Recipiente A: Contém água quente;

-Recipiente B: Contém água fria;

-Recipiente C: Contém álcool quente;

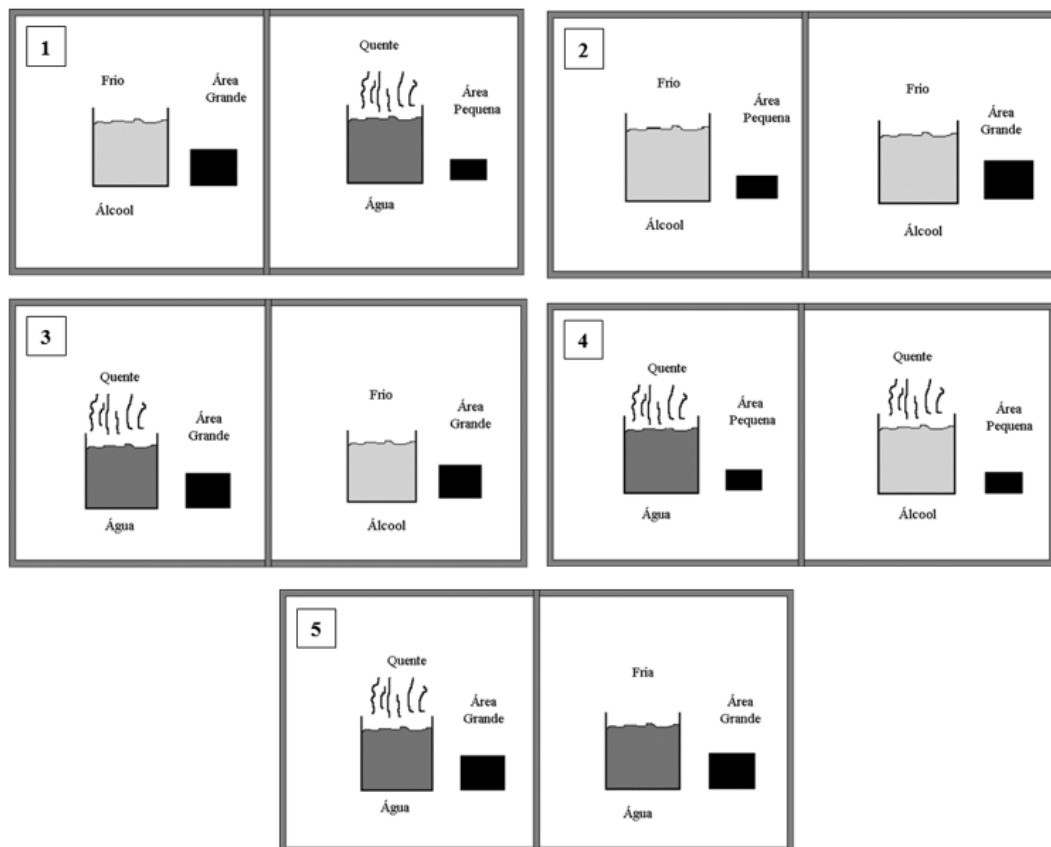
-Recipiente D: Contém álcool frio;

Além dos recipientes com os líquidos, o estudante dispõe de dois blocos de ferro, de mesma massa e com a mesma temperatura, porém de áreas superficiais diferentes, que serão colocados imersos dentro dos recipientes, formando um conjunto (bloco dentro do recipiente com líquido).

Nas páginas a seguir, são exibidos 5 experimentos que o estudante realizou, com o objetivo **de determinar se a temperatura do líquido tem influência sobre temperatura final do conjunto**. Você deverá avaliar os experimentos que o estudante realizou, considerando a adequação entre o objetivo da atividade e o procedimento adotado. Para cada experimento, você deverá decidir se aquele experimento é bom ou ruim para alcançar o objetivo do estudante e justificar sua resposta.

Como exemplo, o experimento 1 (veja a figura) ilustra que o estudante colocou o bloco com área grande dentro do recipiente contendo álcool frio. Esperou certo tempo e mediu a temperatura do conjunto. Depois, pegou o bloco com área pequena e o colocou dentro do recipiente contendo água quente e anotou a temperatura final do conjunto após certo tempo. Com esse procedimento, o estudante é capaz de determinar, com certeza, se a temperatura do líquido tem influência na temperatura final do conjunto?

Siga o mesmo raciocínio para os outros experimentos, justificando porquê você considera os experimentos bons ou ruins.



Fonte: Autoria própria.

Os estudantes trabalharam individualmente durante todo o teste, seguindo seu próprio ritmo. Para cada experimento, eles foram orientados a marcar se o experimento era

bom ou ruim, tendo em vista os objetivos estabelecidos. Os alunos demoraram, em média, 20 minutos para realizar o teste.

Resultados

Para a análise da validade e fidedignidade do teste, foram realizados os processos de correção e de atribuição de nota às respostas. As respostas aos 10 experimentos foram corrigidas de forma dicotômica, sendo atribuído o valor de um ponto caso o respondente acertasse e zero em caso de erro.

Na tabela 1 estão exibidos o índice de dificuldade, o índice de discriminação, o coeficiente de correlação bisserial e a correlação item-total dos 10 experimentos (itens). Os índices de dificuldade, calculados por meio da razão entre o número de sujeitos que acertaram o item e o número total de respondentes, variaram entre 25% e 80%. Seguindo os critérios encontrados em Pasquali (2017), pode-se dizer que o teste foi composto por dois itens fáceis, sete itens moderados e um item difícil.

Tabela 1 – Dados psicométricos dos itens do teste.

Experimentos	1	2	3	4	5	6	7	8	9	10
Índice de dificuldade	0,41	0,31	0,57	0,25	0,73	0,43	0,55	0,80	0,70	0,38
Índice de discriminação	0,65	0,73	0,90	0,64	0,77	0,79	0,74	0,67	0,75	0,67
Correlação bisserial	0,37	0,39	0,61	0,40	0,64	0,48	0,49	0,65	0,58	0,40
Correlação item-total	0,52	0,53	0,72	0,52	0,73	0,61	0,62	0,73	0,69	0,54

Fonte: Autoria própria.

O índice de discriminação mede, basicamente, o quanto cada item distingue sujeitos de desempenho superior e desempenho inferior. É desejável então que os testes tenham itens com elevados índices de discriminação, indicando que quem acertou o item, em média, obteve um bom desempenho no teste. O seu cálculo envolve a diferença entre a proporção de acertos entre os indivíduos de desempenho superior e inferior no teste (PASQUALI, 2017). O grupo de desempenho inferior foi composto por 28,1% dos alunos que obtiveram escores abaixo ou iguais a 3 (primeiro quartil). O grupo de desempenho superior foi formado pelos 25,4% dos alunos que obtiveram escores iguais ou superiores a 8 (terceiro quartil). Os índices de discriminação dos itens variaram entre 0,64 e 0,90, ficando todos acima de 0,40, valor mínimo adotado por Vilarinho (2015), para considerar o item adequado.

O coeficiente de correlação bisserial é, também, uma medida de associação entre o desempenho do indivíduo no item e no teste como todo e verifica se aqueles respondentes que obtiveram um desempenho satisfatório no teste, tenderam a acertar o item em questão. Os coeficientes variaram entre 0,37 e 0,65, ficando todos, novamente, acima do valor de 0,30, considerado por Vizzotto (2022) o mínimo para que se considerar um bom item.

Calculou-se os coeficientes alfa de Cronbach, beta de Silveira e Spearman-Brown, utilizando os valores das correlações item-total da tabela 1. Os valores obtidos, 0,82, 0,84 e 0,73 respectivamente, podem ser considerados adequados, o que atesta positivamente sobre a consistência interna e a fidedignidade do teste elaborado (SILVEIRA,1993).

Considerações finais

Conforme foi descrito, há várias maneiras de se determinar a validade e fidedignidade de um instrumento de pesquisa. O instrumento elaborado passou por um processo de validação de conteúdo, pela análise psicométrica dos seus itens e sua fidedignidade foi verificada pelos cálculos dos coeficientes alfa de Cronbach, beta de Silveira e Spearman-Brown.

O domínio de estratégias adequadas de controle de variáveis é um passo importante no desenvolvimento do indivíduo, uma vez que o seu domínio e aplicação propiciam um forte instrumento para a definição e execução dos procedimentos experimentais (ZIMMERMAN; KLAHR, 2018). Portanto, um instrumento de pesquisa validado para avaliar a habilidade dos alunos do Ensino Médio em controlar variáveis é importante tanto para o desenvolvimento de pesquisas na área de Educação em Ciências quanto para os professores de Física, ao oferecer aos professores um meio confiável de avaliar o progresso e adaptar seus métodos de ensino conforme necessário. Ao utilizar um instrumento validado, os professores podem identificar lacunas no entendimento dos alunos, desenvolver estratégias de ensino mais eficazes e acompanhar o crescimento individual de cada estudante ao longo do tempo.

Tentou-se, neste trabalho, contribuir para o debate sobre o rigor nas pesquisas da área de Educação em Ciências. Mais pesquisas que demonstrem processos diversos de validação de diferentes tipos de instrumentos de pesquisas, quantitativos ou qualitativos, são fundamentais e podem contribuir para conferir maior validade e

confiabilidade aos instrumentos que são utilizados para produzir os dados das pesquisas na área.

Agradecimentos

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo financiamento de pesquisa (APQ-00701-18).

Referências

- ALEXANDRE, N. M. C.; COLUCI, M. Z. O. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. **Ciencia & saúde coletiva**, v. 16, p. 3061-3068, 2011.
- ARAÚJO, N. R. S.; LABURÚ, C. E. Uma análise da validação e confiabilidade da escala de opiniões da seleção de experimentos de química (EOSEQ). **Ensaio: Pesquisa em Educação em Ciências**, v. 11, n. 2, p. 177-196, 2009.
- BORGES, A. T.; GOMES, A. D. T. Percepção de estudantes sobre desenhos de testes experimentais. **Caderno Brasileiro de Ensino de Física**, v. 22, n. 1, p. 71-94, 2005.
- HUTZ, C. S. et al. **Psicometria**. Porto Alegre: Artmed Editora, 2015.
- MOREIRA, M. A. **Metodologias de pesquisa em ensino**. São Paulo: Editora Livraria da Física, 2011, 234p.
- PASQUALI, L. **Psicometria: teoria dos testes na psicologia e na educação**. Petrópolis: Vozes, 2017.
- SANTOS, F. M. T.; GRECA, I. M. Metodologias de pesquisa no ensino de ciências na América Latina: como pesquisamos na década de 2000. **Ciência & Educação**, v. 19, n. 1, p. 15-33, 2013.
- SILER, S. A.; KLAHR, D. Detecting, classifying and remediating: Children's explicit and implicit misconceptions about experimental design. In PROCTOR, R. W. (Ed.), *Psychology of science: Implicit and explicit processes* Oxford University Press, 2012, p. 137–180.
- SILVEIRA, F. L. Validação de testes de lápis e papel. In: MOREIRA, M. A.; SILVEIRA, F. L. (Org.). **Instrumentos de pesquisa em ensino e aprendizagem**. Porto Alegre: EDIPUCRS, 1993.
- SOUZA, A. C.; ALEXANDRE, N. M. C.; GUIRARDELLO, E. B. Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. **Epidemiologia e serviços de saúde**, v. 26, p. 649-659, 2017.
- VILARINHO, A. P. L. Uma proposta de análise de desempenho dos estudantes e de valorização da primeira fase da OBMEP. Dissertação de Mestrado, Universidade de Brasília, Brasília (2015).
- ZIMMERMAN, C.; KLAHR, D. Development of Scientific Thinking. In WIXTED, J. T.; GHETTI, S. (Eds.) *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Volume 4 (Developmental and Social Psychology)*, Wiley, 2018, p. 223-248.