

TESTE CERTO-ERRADO: AUMENTO DO COEFICIENTE DE PRECISÃO

RENATO ALBERTO T. DI DIO

“No início do desenvolvimento de testes objetivos, o item certo-errado era muito adotado. Alguns dos primeiros testes publicados eram constituídos exclusivamente por itens desse tipo. Ultimamente, essa preferência tem decaído tanto que só raramente o encontramos em testes padronizados” (6,145).

Na mesma linha de pensamento diz Ebel: “True false items enjoyed considerable popularity in the early days of objective testing. Although they still appear in achievement tests constructed by teachers for their own use, they have almost disappeared from the tests constructed by specialists for use in wide-scale testing programs” (4,124).

No artigo “Writing the Test Item”, de autoria do mesmo Ebel, inserido em “Educational Measurement”, sob a coordenação de E. F. Lindquist, observa-se que “se forem incluídas num teste certo-errado, afirmações apenas aproximadamente verdadeiras criam um problema difícil para o examinando. Ele não só deve saber *até que ponto* a afirmação é verdadeira, mas precisa adivinhar também *que grau* de inverdade será tolerado pelo examinador” (5,208). Essa exigência de absoluta correção tende a limitar a aplicação e a validade de itens certo-errado.

Entretanto, “depois de um descrédito de muitos anos entre os especialistas em testes, alguns estudos levados a efeito na década de 60 mostraram-se de certa forma encorajadores em sua tentativa de restaurar a perda, ou pelo menos, a comprometida estima dos itens certo-errado. Dois aspectos parecem desempenhar um papel decisivo na baixa reputação dos itens certo-errado: (1) sua pretensa inferioridade em relação aos itens de múltipla-escolha; e (2) sua maior exposição ao elemento acaso” (2,9).

Uma série de tentativas nesse sentido foi representada pela ponderação dos itens, através da técnica que se convencionou chamar “peso-confiança” (*confidence-weighting*), adotada entre outros por Archer e Ebel (1 e 3).

Em resumo, a confiabilidade do teste certo-errado tem sido menor do que a do teste de múltipla-escolha principalmente por sua maior exposição ao acerto casual. Se essa desvantagem puder ser minimizada, se não mesmo abolida, o item certo-errado poderá recuperar a preferência de que gozava, tanto mais quanto é certo que a construção do item de múltipla

escolha exige dois requisitos nem sempre satisfeitos: a) inequívocidade quanto à melhor resposta; b) atratividade e, não obstante, incorreção das alternativas erradas.

O problema

A fim de diminuir o grau de exposição ao acerto casual, tanto maior quanto menor for o conhecimento da matéria por parte do aluno, que, à falta de informação, tenta a resposta ao acaso, imaginou o autor a construção de itens que exigem do examinando, além da indicação “certo” ou “errado”, a correção do item através da substituição de palavras-chave.

Melhor esclarecendo, quando o examinando decidir que o item é verdadeiro bastará que assinale a letra convencional, fazendo um círculo em torno de V (verdadeiro) ou C (certo). Quando se tratar de item havido por falso, o aluno deverá transformá-lo em verdadeiro, substituindo uma ou duas palavras sublinhadas na proposição inicial. Na substituição de palavras que tornam o item verdadeiro, são aceitas palavras sinônimas àquela esperada, como no caso da palavra “preciso”, que, em psicométrica, equivale a fidedigno ou confiável.

Plano de Pesquisa

A disciplina “Medidas Educacionais” constitui matéria obrigatória do 3.º ano do curso básico da Faculdade de Educação da Universidade de S. Paulo.

As turmas do diurno e do noturno foram subdivididas aleatoriamente em dois subgrupos: o grupo experimental — ao qual foi aplicado o teste “certo-errado” com correção por parte do aluno — e o grupo de controle — ao qual foi administrado o teste certo-errado convencional. Ambos os testes, de 40 itens, eram iguais. Apenas a maneira pela qual o aluno devia responder era diversa.

A pesquisa propôs-se a demonstrar que a exigência de transformação do item certo-errado em verdadeiro, através da substituição de palavras sublinhadas, elimina o fator casual e aumenta a fidedignidade.

Hipóteses

Hipótese nula: não há diferença significativa em fidedignidade entre o teste certo-errado convencional e o teste certo-errado construído de modo a obrigar o aluno a transformar em certo o item por ele considerado falso.

Hipótese de pesquisa: o teste certo-errado em que o examinando é obrigado a transformar o item errado em certo, através da substituição de uma ou duas palavras, é mais fidedigno do que o teste certo-errado convencional.

Amostra

O estudo incluiu a turma do diurno, constituída de 50 alunos (27 do grupo experimental e 23 do grupo de controle), e do noturno, formada de 30 alunos (19 do grupo experimental e 19 do grupo de controle).

Os coeficientes de precisão dos testes foram computados pela fórmula de Kuder Richardson.

Resultados

Os resultados obtidos constam da tabela abaixo:

	N	Diurno K-R	N	Noturno K-R
Grupo Experimental	27	0,89	19	0,80
Grupo de Controle	23	0,71	19	0,71
Total	50		38	

Conclusão

Em ambas as turmas apurou-se um substancial aumento do coeficiente de precisão: de 0,71 para 0,80. Isso quer dizer que na correlação do teste convencional havia 0,50 de variância comum, ao passo que, na correlação do teste experimental, a variância comum subiu para 0,64.

Aplicado o z de Fischer, aceito o limite de confiança de 0,95, teste unicaudal, a diferença apurada no grupo diurno é estatisticamente significativa.

Se a pesquisa for reaplicada com maior número de itens e maior número de estudantes, podem ser esperados resultados mais compensadores.

BIBLIOGRAFIA

1. Archer, N. Sidney, "Effects of Confidence Weighting by Fifth and Sixth Grade Students on Objective Test Scores", Súmula de Comunicação apresentada à sessão da American Educational Research Association sobre Medidas Educacionais, em Chicago, a 14-2-1963.
2. Di Dio, Renato A. T., Testes de Aproveitamento, Fidedignidade e Validade Interna em função de Formas, Pesos e Proposições, J. Ozon Editor, São Paulo, sem data.
3. Ebel, Robert L., "Some Effects of Credit for Confidence and Penalty for Error on True-False Test Scores", Súmula de Comunicação apresentada à sessão da American Educational Research Association sobre Desenvolvimento de Testes, em Chicago, a 25-2-1961.
4. Ebel, Robert L., Measuring Educational Achievement, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1965.
5. Lindquist, E. F. (coordenador), Educational Measurement, American Council of Education, Washington D. C., 1966.
6. Noll, Victor N., Introdução às Medidas Educacionais, Pioneira, Rio, 1961.

