

Constituição de Testes de Desempenho para Avaliações Educacionais em Larga Escala. Luiz Pasquali (Universidade de Brasília).

A constituição de testes de personalidade, de habilidades, inventários vocacionais, de interesse, exames de vestibular, provas de sala de aula e avaliações de desempenho em larga escala requerem a manifestação de algum tipo de comportamento – uma ação observável e mensurável que meça algum atributo, traço ou característica. No desenvolvimento de testes para avaliações educacionais em larga escala, o rigor deve ser extremo, pois seus resultados indicarão sucessos e fragilidades do sistema avaliado. Nessa perspectiva, o teste cognitivo deve apresentar um conjunto de tarefas predefinidas, que o sujeito executará em uma situação sistematizada. Seu comportamento, nessa situação, será observado, descrito e julgado, utilizando-se números para descrevê-lo. As tarefas estarão fundamentadas em ações empíricas em que o traço latente deverá se manifestar.

Pretende-se discutir a constituição de testes de desempenho apresentando critérios para a correspondência entre a observação coletada e o objetivo da aplicação do teste, assegurando-se a sua validade, fidedignidade e objetividade. Para tanto, é substancial o seguimento de três passos: (a) avalia-se uma amostra representativa de comportamentos que envolvem o conhecimento dos diferentes tipos de habilidades requeridas para o bom desempenho de um determinado comportamento; (b) as amostras do comportamento devem ser obtidas sob condições padronizadas; (c) devem existir regras para a interpretação dos resultados, pois, elas são necessárias para assegurar que qualquer examinador que pretenda avaliar um mesmo conjunto de respostas que outro avaliador possa proceder à interpretação da mesma forma. Um teste se constituirá um bom instrumento para a finalidade a que se propõe se apresentar validade, fidedignidade e objetividade. A validade é o pressuposto mais importante, pois ela se refere à adequação, significação e utilização prática das inferências feitas a partir dos escores dos testes, sendo indicada pelo processo de acumulação de evidências que dão apoio a tais inferências. Os vários meios de acumular evidências de validade estão agrupados em categorias, sendo as mais comuns: (1) validade relacionada ao conteúdo; (2) validade relacionada ao critério; (3) validade relacionada ao construto. Em testes educacionais, a validade pode ser avaliada pela análise do conteúdo, em que se investiga se o conteúdo relacionado ao teste aplicado corresponde àquilo que foi preconizado no período de escolarização considerado na avaliação; pela escolha dos critérios que assegurem que o teste seja válido para a população avaliada; e pelo construto que se apoia na avaliação dos conceitos teóricos subjacentes ao desempenho. A fidedignidade pode ser avaliada por quatro meios distintos: pela estabilidade temporal, por conjuntos de itens equivalentes, pela consistência interna dos itens e pelas diferenças entre avaliadores. A objetividade do teste é a garantia de que os itens propostos irão cobrir objetivamente comportamentos passíveis de avaliação, permitindo tão somente uma resposta certa ou errada. Quando os testes são constituídos eficazmente, eles levam à reflexão dos pressupostos que os embasaram e podem desenvolver uma cultura acadêmica compartilhada e dedicada a assegurar e melhorar a qualidade da educação.