

habilidade e realização. Esta metodologia ainda está engatinhando, em termos comparativos, pois as especificações necessárias para se criarem regras a partir das quais os computadores possam gerar itens de teste precisam ser consideravelmente mais detalhadas do que nos métodos tradicionais de geração de itens, e requerem um nível mais alto de fundamentação teórica. Não obstante, a geração computadorizada de itens já foi implementada no desenvolvimento de ferramentas para áreas – como avaliações de matemática e testagem de aptidões para aviação – nas quais (a) existem modelos cognitivos de desempenho, (b) os constructos a serem examinados podem ser representados em termos de uma sintaxe lógica e (c) a dificuldade dos itens pode ser medida por meio de referentes objetivos. Sem dúvida alguma, as técnicas de geração de itens vão continuar a ser ativamente pesquisadas pelos estudiosos, devido às muitas vantagens que apresentam em termos de eficiência e economia, bem como por seu potencial em aplicações pedagógicas (Irvine e Kyllonen, 2002).

Avaliação automatizada de questões dissertativas

Uma inovação significativa no esforço para padronizar a avaliação de questões dissertativas é o desenvolvimento de tecnologia informatizada para a *avaliação automatizada de dissertações (AAD)*. A tentativa de avaliar a prosa escrita por meio de programas de computador está em progresso nas últimas décadas e, como a geração computadorizada de itens, também foi facilitada por avanços na psicologia cognitiva e na ciência da computação. Embora ainda esteja em seus estágios iniciais, a AAD já demonstra grandes promessas, não apenas como meio de aumentar a fidedignidade e a validade dos escores, mas também como ferramenta educativa (Shermis e Burstein, 2003).

Teste a si mesmo

1. Os procedimentos envolvidos na análise de item dizem respeito primariamente aos _____ de testes.
 - (a) criadores
 - (b) usuários
 - (c) testandos
 - (d) administradores

2. A análise qualitativa de item tipicamente ocorre
 - (a) depois que um teste é padronizado
 - (b) ao mesmo tempo que um teste passa pela validação cruzada
 - (c) depois que o pool de itens é gerado
 - (d) um pouco antes do teste ser publicado

3. Os itens de escolha forçada pertencem à categoria dos
 - (a) itens de resposta construída
 - (b) itens de resposta selecionada
 - (c) nem a nem b
4. Qual das seguintes não é uma das vantagens dos itens de resposta selecionada em relação aos itens de resposta construída? Os itens de resposta selecionada
 - (a) são menos propensos a erros de avaliação
 - (b) fazem um uso mais eficiente do tempo de testagem
 - (c) são mais fáceis de quantificar
 - (d) são mais fáceis de preparar
5. Os índices de discriminação são estatísticas usadas primariamente para avaliar a _____ dos itens.
 - (a) validade
 - (b) justiça
 - (c) fidedignidade
 - (d) dificuldade
6. Quando se deseja produzir um teste que resulte na máxima diferenciação entre os testandos, deve-se buscar um nível médio de dificuldade (valor p) de
 - (a) 1,00
 - (b) 0,75
 - (c) 0,50
 - (d) 0,00
7. O r Pearson é o coeficiente de correlação mais usado para calcular índices da relação entre desempenho de item e posição no critério. Verdadeiro ou Falso?
8. Para um teste de pura velocidade, os índices habituais de dificuldade e discriminação de item são _____ para um teste de pura potência.
 - (a) menos apropriados do que
 - (b) mais apropriados do que
 - (c) tão apropriados quanto
10. A teoria da resposta ao item, até agora, tem sido aplicada menos extensamente na área da testagem
 - (a) de desempenho
 - (b) de aptidões
 - (c) de conhecimento
 - (d) de personalidade

Respostas: 1. a; 2. c; 3. b; 4. d; 5. a; 6. c; 7. Falso; 8. a; 9. c; 10. d.

FUNDAMENTOS EM USO DE TESTES

Antes de serem usados, os testes só podem ser avaliados do ponto de vista científico e técnico por pessoas que detenham os conhecimentos especializados necessários em desenvolvimento de testes, princípios psicométricos e aspectos do comportamento que os instrumentos buscam avaliar. Depois que a testagem é implementada – por meio dos processos de seleção, administração e avaliação – os resultados precisam ser avaliados, interpretados e comunicados, de forma apropriada à finalidade para a qual foram empregados, por profissionais com conhecimento do contexto em que a testagem acontece, dos aspectos técnicos e das questões psicológicas envolvidas em uma dada situação. No mínimo, avaliar as aplicações de testes psicológicos e educacionais envolve considerações pertinentes à habilidade com que o usuário emprega esses instrumentos e sua adequação aos testandos em que eles são usados. Em um sentido mais amplo, quando a testagem adquire significância prática na vida de indivíduos, o uso de testes também precisa ser avaliado à luz de valores sociais e prioridades políticas. Neste contexto mais amplo, a testagem pode se tornar altamente controversa, e os usuários devem se apoiar em seus códigos de ética profissional e pessoal para determinar se estão dispostos a emprestar seus conhecimentos a certos usos dos testes.

As práticas apropriadas de testagem são reguladas pelos princípios éticos e padrões promulgados pelas profissões que fazem uso de testes psicológicos e educativos (p. ex., American Counseling Association, 1995; AERA, APA, NCME, 1999; APA, 2002; National Association of School Psychologists, 2000). Nas últimas décadas, a preocupação crescente com a possibilidade do mau uso de testes levou essas profissões e as organizações envolvidas com testagem a se engajarem na preparação e divulgação de informações sobre a qualificação dos usuários. Um dos principais esforços nesta direção foi liderado pela APA e resultou na publicação de um conjunto de diretrizes claras que informam a todas as partes interessadas a respeito das qualificações que a APA considera importantes para o uso competente e responsável de testes (Turner, DeMers, Fox e Reed, 2001). O quadro Consulta Rápida 7.1 apresenta um breve resumo destas diretrizes. Além disso, os *Padrões de*

Qualificações para usuários de testes psicológicos

Conforme discutido no Capítulo 1, nenhum conjunto formal de credenciais, seja por formação, licenciamento ou certificação, pode garantir a competência no uso de um teste em particular em uma situação específica.

As qualificações dos usuários de testes psicológicos se baseiam em dois fatores principais:

1. Seu conhecimento e suas habilidades em
 - princípios psicométricos e estatística;
 - seleção de testes à luz de suas qualidades técnicas, da finalidade para a qual eles serão usados e das características dos examinandos;
 - procedimentos para administração e avaliação dos testes, bem como para interpretação, relato e salvaguarda dos resultados, e
 - todas as questões relevantes ao contexto e ao objetivo do uso do teste.
2. O grau em que os usuários tiveram experiência supervisionada adequada em todos os aspectos do conhecimento e das habilidades pertinentes ao uso pretendido de um teste específico.

Para mais informações, ver Turner, S.M., DeMers, S.T., Fox, H.R. e Reed, G.M. (2001). *APA's Guidelines for test user qualifications: An executive summary. American Psychologist, 56*, 1099-1113.

Testagem (AERA, APA, NCME, 1999) incluem um capítulo que descreve as responsabilidades dos usuários de testes, bem como capítulos separados dedicados a questões relacionadas à (a) justiça e imparcialidade na testagem e uso de testes, (b) testagem de indivíduos de origens linguísticas diversas, (c) testagem de indivíduos com deficiências, (d) testagem e avaliação psicológica, (e) testagem e avaliação educacional, (f) testagem para emprego e credenciamento e (g) testagem na avaliação de programas e políticas públicas.

Documentos que fornecem orientação adicional sobre o uso de testes como parte da tomada de decisões de importância crucial para estudantes (U.S. Department of Education, Office for Civil Rights, 2000) e princípios para a validação e o uso de procedimentos de seleção de pessoal (Society for Industrial and Organizational Psychology [SIOP], 2003) também estão disponíveis e devem ser consultados antes de se iniciar projetos nessas áreas. Uma característica particularmente útil do documento do U.S. Department of Education é que, além de resumir os princípios básicos das melhores práticas de testagem no campo da mensuração educacional, ele também discute os requisitos legais no nível federal que se aplicam ao uso não discriminatório de testes para finalidades de importância decisiva. O documento da SIOP, embora não pretenda interpretar estatutos, regulamentos ou jurisprudência relativos a decisões trabalhistas, oferece princípios para a aplicação e o uso de procedimentos de seleção que podem informar e orientar as partes responsáveis pela sua autorização e implementação na avaliação de adequação e propriedade desses procedimentos.

Nos últimos anos, as profissões mais fortemente associadas à testagem psicológica – que estão representadas na Comissão Conjunta sobre Práticas de Testagem (*Joint Committee on Testing Practices; JCTP*) – têm reconhecido cada vez mais que os testandos também precisam ser informados de seus direitos e responsabilidades

no processo de testagem. Esta conscientização culminou na publicação de um documento com o objetivo de fornecer tais informações (JCTP, 1998). O quadro Consulta Rápida 7.2 lista alguns dos direitos e responsabilidades mais importantes citados e descritos neste documento, que está disponível integralmente na seção de Testagem e Avaliação da página da APA na internet (<http://www.apa.org/science/ttrr.html>).

CONSULTA RÁPIDA 7.2
Direitos e responsabilidades dos testandos

O documento *Direitos e Responsabilidades dos Testandos: Diretrizes e expectativas (Rights and Responsibilities of Test Takers: Guidelines and Expectations; JCTP, 1998)* pode ser reproduzido e divulgado gratuitamente. Embora não tenha força de lei – e as leis estaduais e federais americanas se sobreponham aos direitos e responsabilidades nele expostos –, os profissionais envolvidos em processos de testagem têm a responsabilidade de garantir que os testandos estejam cientes das informações contidas no documento.

Alguns dos direitos dos testandos mais importantes incluem os seguintes:

- O direito de receber uma explicação antes da testagem sobre (a) o objetivo da mesma, (b) os testes a serem usados, (c) se os resultados serão relatados a outras pessoas e (d) o uso planejado para os resultados. Se tiverem uma deficiência ou dificuldade de compreender a linguagem do teste, os testandos têm o direito de perguntar e ser informados a respeito de possíveis adaptações.
- O direito de saber se um teste é opcional e de ser informado das conseqüências de não fazê-lo, não completá-lo inteiramente ou cancelar os escores.
- O direito de receber uma explicação sobre os resultados do teste dentro de um tempo razoável e em termos de compreensão comum.
- O direito de que os resultados do teste sejam mantidos em sigilo na medida garantida pela lei.

Algumas das responsabilidades dos testandos mais importantes incluem as seguintes:

- A responsabilidade de ler e/ou escutar seus direitos e suas responsabilidades.
- A responsabilidade de fazer perguntas antes da testagem sobre por que o teste está sendo aplicado, como será aplicado, o que ele terá que fazer e o que será feito com os resultados.

A aplicação de testes para fins inadequados ou de maneiras impróprias por usuários que carecem de treinamento e de qualificações necessárias invariavelmente resulta em mau uso dos testes. Este capítulo trata de algumas considerações essenciais que devem ser levadas em conta ao se usar testes psicológicos, incluindo questões relacionadas à seleção, administração e avaliação, bem como à interpretação e ao relato dos escores. Uma vez que grande parte das informações contidas nos capítulos anteriores é relevante para as questões fundamentais envolvidas no uso responsável de testes, os leitores talvez desejem rever partes deste volume quando tópicos já discutidos forem mencionados no presente capítulo. Uma vez que os usuários e os testandos estejam cientes do que constitui a prática correta da testagem, a possibilidade do mau uso dos testes diminui, e os potenciais benefícios inerentes ao seu uso têm maior probabilidade de serem realizados.

Não Esqueça

- Até mesmo o instrumento melhor desenvolvido e psicometricamente íntegro está sujeito ao uso impróprio.
- Segundo os *Padrões de Testagem* (AERA, APA, NCME, 1999, Capítulo 11), a responsabilidade pelo uso apropriado e pela interpretação correta dos escores de testes depende primariamente do usuário.
- O mau uso dos testes pode ocorrer em qualquer ponto do processo de testagem, começando com a seleção inadequada dos instrumentos, seja para a finalidade para qual serão aplicados, seja para os indivíduos para os quais serão administrados. Erros na administração ou avaliação e na interpretação ou relato dos resultados podem agravar o problema.
- Sempre que a possibilidade do uso de testes for contemplada, a melhor forma de prevenir seu uso impróprio é garantir desde o início que os indivíduos envolvidos em todas as etapas da testagem tenham as qualificações e a competência necessárias para desempenhar seus papéis no processo.

FUNDAMENTOS DA SELEÇÃO DE TESTES

Conforme discutido no Capítulo 1, os testes psicológicos são usados primariamente como auxiliares na tomada de decisões a respeito de pessoas em contextos educacionais, ocupacionais, clínicos, forenses e de aconselhamento vocacional. Além disso, os testes também são usados com frequência na pesquisa psicológica e recentemente começaram a ser aplicados no processo de psicoterapia para fins de desenvolvimento pessoal, aumento da autocompreensão ou ambos (Finn e Tonsager, 1997). As conseqüências do mau uso dos testes nestas várias aplicações diferem amplamente em termos de seu impacto potencial nos testandos individuais. Naturalmente, a quantidade de atenção dedicada a cada passo do processo de testagem –

Advertência

De todas as arenas nas quais os testes psicológicos são usados, talvez nenhuma seja mais controversa do que a das aplicações forenses. Devido à natureza litigiosa da maioria dos procedimentos legais, os profissionais que concordam em servir como testemunhas em julgamentos que envolvem evidências derivadas de testagens psicológicas podem esperar ter as bases de seus depoimentos questionadas em todas as oportunidades. É óbvio que, em situações como estas, os ditos especialistas que estejam malpreparados para defender seus testemunhos baseados em testes podem ser facilmente humilhados e constrangidos por advogados bem preparados para atacá-los. Uma forma instrutiva de se familiarizar com as potenciais armadilhas do mau uso de testes na arena legal, bem como em outros contextos, é a consulta aos vários volumes de *Coping with psychiatric and psychological testimony*, preparados por Jay Ziskin em colaboração com outros autores e voltados basicamente para advogados. A quinta e mais recente edição desta obra foi publicada em 1995 pela Law and Psychology Press, de Los Angeles, CA. Desde então, foi complementada por tomos lançados em 1997 e 2000. Dos três volumes da obra principal, o volume 2 – dedicado principalmente ao questionamento de testemunhos derivados de testes psicológicos de vários tipos – é o mais pertinente aos tópicos considerados neste capítulo.

da seleção do teste em diante – deve levar em conta e ser proporcional ao impacto que a testagem pode ter. Embora a maior parte do material a seguir também seja aplicável ao uso de testes para fins terapêuticos ou investigativos, o ponto de partida para a discussão da maioria dos tópicos deste capítulo é a premissa de que os testes são usados primariamente para a tomada de decisões a respeito de pessoas.

Usar ou não usar (testes), eis a (primeira) questão

Independentemente dos objetivos da testagem psicológica, a primeira questão a ser decidida é se a testagem é necessária. Para resolvê-la, os usuários de testes em potencial devem realizar uma análise de custo-benefício – semelhante à sugerida por Goldman (1971) – e considerar explicitamente as seguintes perguntas:

1. Que tipo de informações eu pretendo obter com a testagem?
2. Como estas informações serão usadas?
3. Quais destas informações já estão disponíveis em outras fontes?
4. Que outras ferramentas podem ser usadas para obter as informações que eu busco?
5. Quais as vantagens de usar testes em lugar de, ou juntamente com, outras fontes de informação?
6. Quais as desvantagens ou os custos, em termos de tempo, esforço e dinheiro, de usar testes em vez de, ou juntamente com, outras fontes de informação?

Se a fundamentação teórica para o uso de testes e a aplicação esperada de seus resultados não estão explícitas desde o início, os escores provavelmente não serão muito úteis, ou, pior ainda, tenderão a ser mal usados. Se as informações buscadas já estão disponíveis em outras fontes ou podem ser obtidas por outros meios, a testagem provavelmente será supérflua, a menos que seu objetivo seja confirmar o que já se sabe ou obter corroboração adicional. Por fim, se as vantagens ou benefícios a serem obtidos com o uso de testes não superarem seus custos ou desvantagens, incluindo qualquer dano potencial, a testagem obviamente não é aconselhável. O quadro Consulta Rápida 7.3 lista algumas das principais razões e circunstâncias nas quais a testagem pode ser desaconselhada.

Dois motivos importantes para usar testes

Como a testagem ocorre em tantos contextos diferentes e é usada para fins tão variados, é difícil discutir seu impacto em termos abstratos. No entanto, qualquer que seja o contexto em que os testes psicológicos são empregados – desde que estes tenham sido desenvolvidos cuidadosamente, e a documentação adequada de seu valor psicométrico para um determinado fim esteja disponível –, quando usados criteriosamente eles têm vantagens distintas em relação a outros métodos de coleta

Dez Motivos principais para *não* usar testes

Existem muitos motivos para que o uso de testes psicológicos seja desaconselhável e muitas situações em que devemos nos precaver ao utilizá-los, e a lista a seguir apresenta apenas os mais proeminentes. Com poucas exceções, os testes psicológicos não devem ser usados quando qualquer uma destas circunstâncias se verificar:

1. O objetivo da testagem é desconhecido ou pouco claro para o usuário do teste.
2. O usuário não está totalmente familiarizado com toda a documentação necessária para o teste, nem treinado quanto aos procedimentos relacionados.
3. O usuário não sabe para onde irão os resultados do teste ou como eles serão usados, ou não pode salvaguardar seu uso.
4. As informações buscadas com a testagem já estão disponíveis ou podem ser obtidas mais eficientemente por outras fontes.
5. O testando não está disposto ou não é capaz de cooperar com a testagem.
6. O testando pode sofrer algum dano devido ao processo de testagem.
7. O local e as condições da testagem são inadequados.
8. O formato ou os materiais do teste são inapropriados à luz da idade, sexo, origem cultural ou lingüística, deficiência ou qualquer outra condição do testando que possa invalidar os dados do teste.
9. As normas do teste são desatualizadas, inadequadas ou inaplicáveis ao testando.
10. A documentação sobre a fidedignidade ou a validade dos escores do teste é inadequada.

de informações a respeito de pessoas. As vantagens mais significativas que os testes psicológicos oferecem dizem respeito à sua eficiência e objetividade características.

- *Eficiência.* Muitas questões contempladas pela testagem psicológica podem ser respondidas por outros métodos, desde que os indivíduos que buscam as informações tenham o tempo e os recursos necessários para coletá-las. Por exemplo, a maioria das pessoas não requer o uso de testes para esboçar um perfil psicológico daqueles com quem elas tiveram um contato pessoal extenso e contínuo. Observando o comportamento dos indivíduos e interagindo com eles em uma variedade de situações por um período longo de tempo, podem ser coletados amplos dados com que se pode tirar conclusões a respeito de suas habilidades e atributos pessoais, e mesmo sobre sua tendência a se comportar no futuro. A testagem psicológica fornece àqueles que estão em posição de avaliar e tomar decisões a respeito de pessoas – mas que não têm a oportunidade de observações e interações prolongadas com elas – as ferramentas para coletar as informações necessárias de forma eficiente em termos de tempo e custo.
- *Objetividade.* Mesmo quando existem extensas oportunidades para observar e interagir com as pessoas, os dados que podemos coletar por meio de observações e interações informais podem ser de valor questionável ou limitado. Dados de observação obtidos de modo não-sistemático podem facilmente levar a julgamentos pouco precisos ou refletir a subjetividade

dos observadores, incluindo variações em *o que* eles observam e recordam e também em *como* observam, relatam e avaliam suas observações. É desnecessário dizer que esta subjetividade entra em jogo especialmente quando os observadores não são imparciais, como é o caso de quando são amigos, parentes ou colegas da pessoa observada. Mesmo quando os observadores são hábeis e capazes de manter o distanciamento em relação aos sujeitos, dados brutos coletados de observações e interações naturalistas não podem ser avaliados e interpretados apropriadamente sem algum referencial padronizado de comparação. Como foi discutido em capítulos anteriores, o sentido e o valor que podem ser derivados das amostras de comportamento obtidas por testes padronizados dependem quase inteiramente dos referenciais normativos, ou baseados no critério que estão disponíveis para comparações, e dos dados acumulados sobre fidedignidade e validade de escores de teste.

Utilidade dos testes

Sempre que o uso de testes para a tomada de decisões a respeito de pessoas é contemplado, a utilidade potencial de seu uso deve ser considerada e analisada. O conceito de *utilidade*, introduzido rapidamente no Capítulo 5 como uma consideração relacionada à validade, se refere a uma estimativa da desejabilidade subjetiva de um resultado ou evento. A utilidade é um aspecto do tópico muito mais amplo da teoria das decisões, uma abordagem muito usada na coleta e análise de informações, geralmente em forma matemática, para gerar estratégias racionais para a tomada de decisões (Bell, Raiffa e Tversky, 1988). Estimar o valor de decisões corretas *versus* o custo de decisões incorretas é invariavelmente uma questão complicada, em especial quando as decisões a serem tomadas afetam a vida de seres humanos de maneira significativa. A teoria das decisões se dedica ao desenvolvimento e à análise de possíveis estratégias para a tomada de decisões, usando as informações disponíveis para estimar sua utilidade através do cálculo dos custos e benefícios de vários resultados alternativos em termos quantitativos, geralmente monetários. Naturalmente, de modo a estimar a utilidade de um resultado, um ponto de vista subjetivo precisa ser adotado, e este se origina tipicamente dos valores e prioridades dos responsáveis pelas decisões.

Por exemplo, no campo da seleção para emprego, existem duas decisões possíveis (contratar e não contratar). Supondo que o critério de desempenho no trabalho pode ser dicotomizado nas categorias de sucesso e fracasso, estas decisões têm quatro resultados possíveis: (a) aceitações válidas, (b) rejeições válidas, (c) aceitações falsas e (d) rejeições falsas. As aceitações válidas são o mais desejável de todos os resultados, pois não apenas não representam riscos para o empregador nem para o candidato como, na verdade, oferecem benefícios a ambos. As aceitações falsas, por outro lado, representam alguns riscos para o empregador (p. ex., perda de receita ou indenizações pela incompetência do empregado, tempo e esforço desperdiçados em atividades de contratação e treinamento, aumento da rotatividade, etc.), mas geralmente não para o empregado. As rejeições válidas são em grande

parte vantajosas para o empregador, mas não para os candidatos rejeitados. As rejeições falsas não apresentam tipicamente riscos significativos para o empregador, a menos que as decisões sejam contestadas, mas podem prejudicar os candidatos que são rejeitados falsamente, talvez de maneira significativa.

Claramente, a utilidade dos resultados difere para as várias partes afetadas por uma decisão. Além disso, em potencial as conseqüências das decisões para todas as partes envolvidas nunca podem ser totalmente antecipadas, e muito menos quantificadas. Ainda assim, a noção de tentar explicitamente antecipar e estimar os prováveis riscos e benefícios inerentes ao uso de testes – em oposição a outras ferramentas – está no centro de seu uso responsável. Para tanto, a teoria das decisões fornece alguns conceitos básicos que podem ser usados ao se considerar se, e em que grau, os testes podem contribuir para melhores resultados ao aumentar o número de decisões corretas e minimizar o número de decisões incorretas. Esses conceitos são especialmente adequados para organizações que buscam determinar possíveis ganhos na precisão das decisões de seleção que podem ser alcançados implementando-se uma estratégia que inclua o uso de escores de testes (Kobrin, Camara e Milewski, 2002). Os seguintes são itens essenciais de informação necessários para se estimar tais ganhos:

- *Dados de validade*: Os escores de teste somente podem melhorar a tomada de decisões se tiverem validade demonstrável para avaliar a posição dos testandos ou predizer seu desempenho em um critério. Não havendo outras falhas, quanto mais alto o coeficiente de validade dos escores de teste, maior a precisão das estimativas de critério e predições (ver Capítulo 5).
- *Dados de taxa de base*: O conceito de *taxa de base* se refere ao ponto de partida do qual os ganhos potenciais em precisão podem ser calculados, ou seja, a probabilidade estabelecida de eventos em uma população antes da introdução de qualquer procedimento novo, como a testagem. Por exemplo, se a proporção de alunos que se formam em um programa de estudos é de 0,60 ou 60% daqueles que foram admitidos no programa, a taxa de base é 0,60. Em decisões de seleção de emprego, a taxa de base é a proporção de candidatos contratados para um determinado cargo que se mostram bem-sucedidos em seus empregos. Para fins de cálculo da *validade incremental*, ou melhora operada por um teste nas decisões de seleção, a taxa de base se refere à proporção de decisões corretas que são feitas sem o uso dos escores de teste. Não havendo outras falhas, os ganhos potenciais de precisão nas decisões de seleção são maiores quando as taxas de base são próximas a 0,50. As taxas de base próximas de extremos (p. ex., 0,10 ou 0,90) indicam que a seleção precisa é muito difícil ou muito fácil nas circunstâncias existentes. Nestes casos, usar escores de teste como base da seleção – mesmo que tenham um grau relativamente alto de validade – pode não aumentar a precisão das decisões, e até mesmo diminuí-la.
- *Razões de seleção*: Outra restrição à potencial contribuição dos escores de teste à melhoria das decisões de seleção tem origem na *razão de seleção*, que é a razão que resulta quando o número de vagas disponíveis é dividi-

do pelo número de candidatos a elas. Se existem 5 vagas disponíveis e 500 candidatos, a razão de seleção é $5 \div 500$, ou 1%, enquanto que, se somente 25 pessoas se candidatarem, a razão de seleção será $5 \div 25$, ou 20%. Naturalmente, razões de seleção menores permitem aos encarregados das decisões serem mais seletivos do que as razões maiores. Num exemplo extremo, se três vagas estiverem abertas e três pessoas se candidatarem a elas, a razão de seleção será de 100%, e o empregador não terá latitude de escolha se a organização precisar destes trabalhadores. Numa tal situação, a validade incremental que os escores de teste podem oferecer – independentemente de sua validade – é nula. Por outro lado, quando a razão de seleção é muito pequena, mesmo um teste cujos escores têm um grau apenas moderado de validade preditiva pode ajudar a aumentar a taxa de decisões precisas.

Como a discussão acima deixa claro, o grau de melhora na precisão das decisões de seleção que pode ser ganho com o uso de escores de teste no contexto educacional e profissional de emprego depende de uma combinação das taxas de base e razões de seleção em uma determinada situação, bem como da validade dos escores. Os efeitos interativos destas três variáveis na precisão da seleção têm sido apreciados há um bom tempo: já em 1939, Taylor e Russell publicavam um conjunto de tabelas que mostravam a proporção esperada de sucessos (aceitações válidas) que podem ser esperados para várias combinações de taxas de base, razões de seleção e coeficientes de validade. As tabelas Taylor-Russell, que são anteriores ao advento da teoria das decisões, fornecem informações básicas que podem ser usadas na avaliação da possibilidade de se usar um determinado teste em um certo contexto, mas não avaliam todos os resultados possíveis de uma decisão, nem contemplam todos os fatores pertinentes às decisões de seleção.

Desde a publicação das tabelas Taylor-Russell, os psicólogos organizacionais-industriais avançaram muito além delas e desenvolveram meios adicionais de avaliar os efeitos do uso de testes, incorporando conceitos e modelos da teoria das decisões. Os refinamentos incluem a estimativa dos efeitos do uso de testes em resultados outros que não o das aceitações válidas, como as rejeições falsas, e em resultados medidos por critérios contínuos ou graduados, que podem ser mais realistas para as aplicações de tomada de decisões do que a simples dicotomia sucesso-fracasso embutida nas tabelas Taylor-Russell. Outros aspectos da tomada de decisões, como o uso de múltiplos preditores em várias combinações e a avaliação de estratégias para seleção, classificação e colocação de pessoal em termos de sua utilidade desde vários pontos de vista (p. ex., aumento da produtividade, diminuição da rotatividade, minimização dos custos de treinamento, etc.) também foram investigados em relação ao uso de testes na administração de recursos humanos (Boudreau, 1991; Schmidt e Hunter, 1998). Para uma descrição detalhada de alguma das aplicações mais minuciosas de muitos conceitos da teoria das decisões ao desenvolvimento e avaliação de uma bateria de testes e outras ferramentas para seleção e classificação de pessoal, os leitores interessados podem consultar o relato de J.P. Campbell e Knapp (2001) sobre o Projeto A do Exército americano (ver Capítulo 5 para mais informações sobre este tópico).

Utilidade dos testes na tomada de decisões clínicas

Alguns conceitos da teoria das decisões podem ser aplicados para maximizar a eficiência da testagem e da avaliação em outras áreas além da seleção de pessoal. Na psicologia clínica, por exemplo, as *taxas de base* se referem à frequência com que condições patológicas, como a depressão, ocorrem em uma determinada população. Desde que as informações necessárias estejam disponíveis, o conhecimento das taxas de base pode ajudar a avaliar a possível utilidade do uso de testes para fins diagnósticos ou preditivos dentro de uma população. Como acontece nas decisões de seleção, a contribuição que os testes podem trazer para melhorar a precisão diagnóstica ou preditiva é maior quando as taxas de base são próximas de 0,50. Quando uma condição a ser diagnosticada é muito comum ou muito rara em uma população, isto é, se as taxas de base são altas ou baixas demais, a melhora na precisão diagnóstica que pode ser obtida com o uso de testes diagnósticos é mais baixa do que quando a condição ocorre com frequência moderada. O motivo para isso é o mesmo que aquele nas decisões de seleção, ou seja, as taxas de base têm um efeito limitante na precisão das decisões. Com respeito aos diagnósticos, se a taxa de base de uma condição em uma população é muito alta (p. ex., 0,80), a probabilidade de um *achado falso-positivo* – isto é, de se diagnosticar a condição quando ela não está presente – é baixa (0,20), mesmo se as decisões diagnósticas forem feitas de maneira aleatória. Da mesma forma, se a taxa de base for extremamente pequena, a probabilidade de um *achado falso-negativo* – isto é, de não se detectar a condição quando ela estiver presente – também é baixa. Nestas situações, o custo de usar escores de teste para identificar a presença de uma condição dentro de um grupo ou população inteira pode não ser justificado pelos ganhos relativamente modestos em precisão que os escores poderiam oferecer. Para uma excelente discussão das implicações das taxas de base na prática clínica, o leitor pode consultar o capítulo de Finn e Kamphui (1995) sobre esse tópico.

Deve-se observar que, no contexto clínico, assim como em outros campos da prática de avaliação, as decisões geralmente são tomadas de forma individual, usando-se dados de uma variedade de ferramentas, incluindo o conhecimento especializado e o julgamento do avaliador. Uma revisão completa do histórico médico, incluindo resultados de exames laboratoriais e achados das várias técnicas de neuroimagens disponíveis atualmente, para explorar ou excluir fatores fisiológicos (p. ex., uso ou abuso de drogas, transtornos neurológicos ou endócrinos, etc.) que possam estar causando ou contribuindo para sintomas psiquiátricos é uma parte indispensável das avaliações psiquiátricas e neuropsicológicas no contexto clínico. Por isso, a contribuição feita pelos resultados de testes psicológicos não pode ser avaliada isoladamente. Além disso, a potencial utilidade de decisões diagnósticas individuais baseadas somente em resultados de testes é mitigada pelas taxas de base baixas da maioria dos transtornos mentais na população geral (O'Leary e Norcross, 1998), pelos coeficientes de validade relativamente pequenos de muitos testes usados para esses fins, e pela natureza sempre em evolução da nosologia psiquiátrica e dos critérios diagnósticos, especialmente em relação aos transtornos de personalidade (Livesley, 2001). Além disso, questões contextuais ligadas às razões por que uma avaliação é realizada podem alterar significativamente a relativa

facilidade ou dificuldade de se obterem resultados precisos com a testagem ou outras ferramentas de avaliação. Em vista de todos estes fatores, não surpreende que a utilidade dos testes na tomada de decisões clínicas seja objeto de freqüentes debates no campo da psicologia. O quadro Consulta Rápida 7.4 lista algumas contribuições dignas de nota aos vários lados dessa discussão.

Outras ferramentas de avaliação

Existem, é claro, muitas vias que podem ser usadas em vez de ou em conjunção com os testes para se obterem informações para avaliar e tomar decisões a respeito de pessoas. Dados biográficos ou de estudos de caso, entrevistas, observações sistemáticas ou naturalistas, históricos acadêmicos e ocupacionais, bem como referências de professores e supervisores estão entre as ferramentas mais usadas na avaliação de indivíduos. Cada uma dessas fontes de informação, sozinhas ou combinadas, pode fornecer dados valiosos e contribuir para uma tomada de decisões bem-sucedida. De fato, várias dessas ferramentas de avaliação (p. ex., entrevistas, observações) podem ser padronizadas e avaliadas em termos de sua fidedignidade e validade, bem como em relação à sua utilidade na tomada de decisões. Quando a validade de tais ferramentas para um fim específico é defendida, seu uso se encaixa na esfera de ação e nas diretrizes dos *Padrões de Testagem* (AERA, APA, NCME, 1999, p.3-4). Se os dados produzidos por esses métodos forem demonstravelmente

CONSULTA RÁPIDA 7.4

Os testes psicológicos são úteis na tomada de decisões clínicas?

Esta questão é objeto de quase tantos debates quanto à questão de se a psicoterapia é eficaz no tratamento dos transtornos mentais. Poucos estudiosos respondem a qualquer uma delas com um sim ou não simples e definitivo. A maioria dos profissionais que usa testes no contexto clínico, forense ou de aconselhamento acredita piamente em seu valor, enquanto muitas vezes a oposto se aplica aos que não os empregam. Para uma amostra de opiniões sobre este tópico, os leitores podem consultar as seguintes fontes:

- Eisman, E.J., Dies, R.R., Finn, S.E., Eyde, L.D., Kay, G.G., Kubiszyn, T.W., Meyer, G.J. e Moreland, K. (2000). Problems and limitations in the use of psychological assessment in the contemporary health care delivery system. *Professional Psychology: Research and Practice*, 31, 131-140.
- Hummel, T.J. (1999). The usefulness of tests in clinical decisions. In J.W. Lichtenberg e R.K. Goodyear (Eds.), *Scientist-practitioner perspectives on test interpretation* (pp. 59-112). Boston: Allyn e Bacon.
- Kubiszyn, T.W., Meyer, G.J., Finn, S.E., Eyde, L.D., Kay, G.G., Moreland, K.L., Dies, R.R. e Eisman, E.J. (2000). Empirical support for psychological assessment in clinical health care settings. *Professional Psychology: Research and Practice*, 31, 119-130.
- Meyer, G.J., Finn, S.E., Eyde, L.D., Kay, G.G., Moreland, K.L., Dies, R.R., Eisman, E.J., Kubiszyn, T.W., & Reed, G.M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56, 128-165. (Ver a seção Comentários da edição de fevereiro de 2002 da *American Psychologist* para as respostas a Meyer et al.).

fidedignos e válidos, eles podem ser suficientes como base para a tomada de decisões, sejam sozinhos ou combinados.

Biodados

Informações sobre o histórico de vida, também conhecidas como *biodados*, podem ser obtidas por uma série de métodos, incluindo entrevistas, questionários e exames anteriores de registros de comportamentos, como históricos acadêmicos, registros policiais, etc. Como já foi dito, o exame do histórico médico é um aspecto indispensável de qualquer avaliação clínica de sintomas, que possam ter origem ou ser afetados por transtornos neurológicos ou metabólicos, ingestão de drogas ou outras possíveis condições físicas. Além disso, em quase qualquer tipo de avaliação, dados bem documentados a respeito do comportamento e das realizações passadas de uma pessoa estão entre as fontes mais válidas e confiáveis de informações avaliativas, porque são evidências factuais do que um indivíduo é capaz de fazer ou realizar. Por exemplo, com respeito à difícil questão de prever o comportamento violento, a abordagem atuarial da avaliação de risco fornece o ponto-âncora fundamental para predições. Esta abordagem – que pode ser complementada por uma avaliação do estado atual, bem como por informações clínicas e testagens – se baseia em um exame sistemático de dados de histórico que demonstraram relações empíricas com a periculosidade, como violência anterior, abuso de substâncias, estabilidade no emprego, transtorno mental e desajuste precoce (Borum, 1996; Monahan e Steadman, 2001). Um exemplo semelhante do valor dos dados de histórico pode ser obtido no campo do emprego, no qual geralmente se pressupõe que um provado histórico de sucesso em uma ocupação é um dos melhores determinantes possíveis do sucesso futuro. Assim, se as informações a respeito do desempenho progressivo de um empregado em potencial no trabalho são confiáveis e altamente favoráveis, a maioria dos empregadores não hesitaria em contratá-lo. Uma abordagem mais estruturada e formal à utilização de informações do histórico de vida de uma pessoa é por intermédio de dados biométricos derivados de inventários biográficos padronizados, que podem ser validados como preditores do desempenho futuro em uma variedade de contextos. Para uma explicação mais abrangente sobre o uso de informações biográficas em decisões de seleção, os leitores interessados devem consultar o *Biodata handbook*, organizado por Stokes, Mumford e Owens (1994).

Dados de entrevista

Dados de entrevistas podem fornecer uma abundância de informações em quase qualquer contexto de avaliação; eles propiciam uma oportunidade de observar o comportamento do entrevistado e coletar dados pertinentes de histórico de vida, bem como informações sobre as atitudes, opiniões e valores do indivíduo. Uma entrevista face a face é uma ferramenta muito flexível e, quando conduzida adequadamente, pode demonstrar um valor crítico na tomada de decisões a respeito

de pessoas. Em muitos contextos clínicos, forenses e profissionais de emprego, entrevistar os indivíduos que estão sendo avaliados, bem como aqueles que podem fornecer dados colaterais, é considerado um aspecto essencial do processo de avaliação (Hersen e Van Hasselt, 1998). No entanto, a fidedignidade e a validade dos dados de entrevista são altamente dependentes da habilidade e da objetividade do entrevistador ao coletar, registrar e interpretar as informações. Para combater os pontos fracos em potencial inerentes aos dados de entrevista, as práticas atuais na maioria dos campos que usam esta técnica enfatizam o treinamento intensivo dos entrevistadores ou as entrevistas estruturadas – que são na verdade instrumentos padronizados semelhantes aos testes em muitos aspectos –, ou ambos.

Observação

Outra fonte onipresente de dados de avaliação consiste em classificações, *checklists* e sistemas de codificação baseados em vários tipos de observação comportamental direta ou contatos relevantes prévios com a pessoa a ser avaliada. Assim como nas entrevistas, a fidedignidade e a validade das classificações e outros dados derivados de observações podem variar muito, dependendo do avaliador e do sistema usado para registrar as observações. Um treinamento que forneça padrões uniformes para que os observadores avaliem o desempenho pode melhorar a qualidade dos dados finais. O uso de escalas padronizadas para coletar dados de informantes como pais ou professores é um procedimento comum na avaliação de crianças e adolescentes (Kamphaus e Frick, 2002, Capítulos 7 e 8).

Fontes adicionais de informação

Muitas fontes alternativas de informação usadas na avaliação de indivíduos não podem ter sua fidedignidade e validades propriamente medidas. Por exemplo, o valor de cartas de recomendação e referências semelhantes depende do grau em que a pessoa que as fornece está disposta e é capaz de revelar informações pertinentes de forma conscienciosa e completa. Aquelles que recebem as referências quase nunca estão cientes dos fatores que podem interferir em sua confiabilidade. Da mesma forma, nas últimas décadas, a atribuição de notas altas em todos os níveis da escolarização desgastou o sentido de muitas credenciais acadêmicas, a ponto de as médias e mesmo alguns diplomas não poderem mais ser tomados por seu valor de face com respeito às competências que deveriam atestar.

Procurando e avaliando testes

Quando o uso de testes psicológicos para uma finalidade específica é contemplado, os usuários são confrontados com duas tarefas principais, quais sejam, (a) encontrar instrumentos disponíveis para os objetivos que têm em mente e (b) avaliá-los

desde esta perspectiva, bem como do ponto de vista dos testandos em potencial. Com relação à primeira tarefa, os leitores devem consultar a seção Fontes de Informações a Respeito de Testes, no Capítulo 1 deste livro, e, particularmente, o artigo sobre *FAQ/Finding Information About Psychological Tests* (Perguntas mais frequentes/Como encontrar informações a respeito de testes psicológicos; APA, 2003), criado e mantido pela equipe de Testagem e Avaliação da APA e disponível em sua página na internet. Conforme mencionado no Capítulo 1, este recurso é um excelente ponto de partida para a pessoa que busca informações sobre testes psicológicos publicados ou inéditos. O artigo lista e resume o conteúdo de todos os principais trabalhos de referência sobre testes de ambos os tipos e dá informações sobre como localizar testes e editoras e como comprar materiais, programas de computador e serviços de avaliação disponíveis e informações adicionais sobre o uso adequado dos testes.

Avaliar instrumentos que foram identificados como de possível uso para um objetivo específico é uma questão mais complexa. Com certeza, as revisões e a literatura sobre os testes em questão fornecem tipicamente informações básicas sobre o tipo de escores e a cobertura propiciados pelos testes, os objetivos e a população para os quais foram criados, como foram desenvolvidos, seus procedimentos de padronização, os tipos de itens que usam, possíveis fontes de viés de teste e achados relacionados à fidedignidade e à validade dos escores, bem como características práticas do conteúdo e delineamento que podem influenciar a facilidade de sua administração e avaliação. Estas informações podem ajudar o usuário a decidir se o instrumento parece ser apropriado para a finalidade e as pessoas para as quais se destina, ou se normas locais ou dados adicionais de fidedignidade e validade precisam ser coletados. Porém, mesmo quando um teste é considerado apropriado do ponto de vista psicométrico, o usuário se vê às voltas com muitas outras questões práticas que são pertinentes unicamente ao contexto e à situação específica do testando, em termos de qualificação e experiência que o uso de um

Não Esqueça

O artigo de Lawrence Rudner (1994) sobre "Perguntas a se fazer ao avaliar testes", disponível na Internet em <http://edres.org/pare>, fornece um bom ponto de partida para a avaliação de testes. A maior parte dos capítulos deste livro contém materiais relevantes à avaliação de testes. Um panorama breve de alguns aspectos-chave a se considerar no processo de seleção de instrumentos pode ser encontrado nos seguintes itens:

- Consulta Rápida 3.2: Informações necessárias para avaliar a aplicabilidade de uma amostra normativa
- Consulta Rápida 4.7: Considerações sobre fidedignidade na seleção de testes
- Consulta Rápida 5.8: Relações entre perguntas, decisões e predições que requerem validação relacionada ao critério
- Consulta Rápida 5.12: Estratégias de validação em relação à interpretação de escores de teste
- Tabela 5.1: Aspectos da validade de constructo e fontes de evidências relacionadas
- Consulta Rápida 6.6: O que torna as amostras de comportamento coletadas pelos itens de teste tão complexas?

teste requer, a limitações financeiras e temporais, etc. O quadro Consulta Rápida 7.5 lista algumas variáveis que precisam ser consideradas em relação aos indivíduos ou grupos específicos aos quais um teste será administrado.

CONSULTA RÁPIDA 7.5**Algumas variáveis a considerar ao selecionar testes para possíveis testandos**

- *Variáveis relacionadas à apresentação do teste:* A escolha de um meio específico de apresentação para os itens de teste, como lápis e papel versus administração computadorizada, ou apresentação oral versus visual, vai representar algumas vantagens ou desvantagens para os examinandos, dependendo de sua familiaridade com o meio, acuidade sensorial nas modalidades visual ou auditiva, habilidades motoras, etc. Mesmo quando um teste inclui alguns itens de prática para familiarizar os testandos com o meio escolhido, suas diferenças individuais provavelmente vão continuar a afetar o desempenho.
- *Variáveis relacionadas ao formato do teste:* Independentemente do conteúdo, os itens de resposta selecionada tendem a requerer habilidades mais receptivas do que os itens de resposta construída, que envolvem o uso de habilidades expressivas. Como os testandos diferem em termos destas habilidades, a escolha do formato dos itens também vai afetar seu desempenho. Outras variáveis, como o uso de limites de tempo ou administração individual versus grupal, também vão afetar os testandos diferencialmente, dependendo de suas origens culturais e experiências.
- *Variáveis relacionadas à linguagem dos itens de teste:* Sempre que a linguagem fizer parte de um teste, mas não de sua essência, as habilidades lingüísticas receptiva e expressiva dos testandos poderão afetar indevidamente os escores. Por isso, dependendo dos requisitos de um teste específico, os examinadores devem considerar e determinar se os testandos têm níveis suficientes de vocabulário e habilidade de leitura e escrita para compreender e tentar realizar as tarefas solicitadas. Para tanto, por exemplo, os manuais muitas vezes incluem informações sobre o nível de leitura necessário para se compreender os itens do teste.

Se o investimento de tempo, esforço e recursos financeiros necessários para implementar o uso de um teste em um dado contexto parecer se justificar pelas informações que os escores poderão fornecer, o passo seguinte para os usuários é efetivamente obter o teste e experimentá-lo em si mesmos ou em um ou mais indivíduos que estejam dispostos a se submeter a uma prova de administração. Para tanto, muitas editoras disponibilizam *conjuntos de prova* de seus instrumentos, que incluem os manuais do teste e amostras dos materiais necessários para administração e avaliação. Um estudo cuidadoso destes materiais, especialmente do manual, é considerado pelos *Padrões de Testagem* um pré-requisito para o uso criterioso de testes, porque eles proporcionam a documentação necessária para que os usuários avaliem o grau em que os resultados vão servir aos fins pretendidos. Além disso, os documentos fornecidos pelas editoras devem especificar as qualificações exigidas para administrar e interpretar os escores com precisão e, sempre que possível, alertar os leitores a respeito de possíveis usos impróprios de um teste (AERA, APA, NCME, 1999, Capítulo 5). Estes documentos podem incluir manuais técnicos, guias do usuário e outros materiais semelhantes que complementem o manual.

Não Esqueça

Um recurso único para a prática aplicada na leitura e interpretação dos materiais de manuais de testes é fornecido no livro de Ann Corwin Silverlake (1999), *Comprehending test manuals: A guide and workbook*, publicado em Los Angeles pela Pyczak.

ASPECTOS ESSENCIAIS DA ADMINISTRAÇÃO E AVALIAÇÃO DE TESTES

Ao contrário de muitos outros aspectos do uso de testes considerados no presente capítulo, os fundamentos da administração de testes podem ser resumidos facilmente em duas palavras: preparação adequada. A administração apropriada de testes psicológicos requer a preparação cuidadosa do ambiente da testagem, do testando e da pessoa que administra o teste.

Preparando o ambiente da testagem

Os princípios mais importantes a serem seguidos na preparação do ambiente no qual vai ocorrer uma testagem são antecipar e remover qualquer fonte potencial de distração. As salas devem ser adequadamente ventiladas e iluminadas, devem oferecer assentos e espaço apropriado para os testandos e devem estar livres de ruídos e outros estímulos (p. ex., alimentos ou bebidas) que possam perturbar a capacidade dos testandos de atentar para suas tarefas. Para evitar possíveis interrupções, é de praxe afixar um aviso – que muitas editoras de testes fornecem gratuitamente – na porta da sala de exame para alertar os passantes de que uma testagem está ocorrendo.

Além de simplesmente garantir um local adequado, a preparação do ambiente de testagem envolve a observação das instruções do manual do teste para sua administração, que buscam replicar as condições em que o teste foi padronizado o mais fielmente possível. Para testagens em grupo, estas instruções podem incluir o fornecimento dos materiais necessários ao teste (p. ex., lápis com borrachas) e uma disposição de cadeiras que impeça que os testandos conversem ou enxerguem as folhas de respostas uns dos outros, presença de supervisores, etc. Para testes individuais, medidas especiais devem ser tomadas para que o examinador possa apresentar os materiais na posição correta, registrar as respostas sem interferências e observar plenamente o comportamento do testando.

Como regra geral, a presença de qualquer pessoa além do examinador e do testando na sala onde ocorre a administração de um teste individual não deve ser permitida (ver, p. ex., National Academy of Neuropsychology, 1999). A presença de terceiros traz a possibilidade de distrair o testando, ou mesmo influenciar o processo de testagem, e introduz um elemento inconsistente com a administração padronizada e um risco adicional e desnecessário à segurança do teste. Pode haver circunstâncias especiais que requeiram a observação da administração de um teste

por outras pessoas – por exemplo, alunos que estejam recebendo treinamento formal. Idealmente, esta observação deve ser feita de uma sala adjacente à sala da testagem, por meio de um espelho unidirecional. Para testandos com dificuldade de se comunicar com o examinador, devido à pouca idade, origem lingüística ou deficiência, a presença de um dos pais ou de um intérprete pode ser necessária na sala de testagem. Nessas situações, bem como em quaisquer outros casos em que sejam feitas adaptações que possam influenciar a interpretação dos escores, o relatório dos resultados deve incluí-las.

Preparando o testando

Existem dois aspectos distintos da testagem relacionados à preparação do testando. O primeiro está em grande parte dentro da esfera de ação do examinador e diz respeito ao estabelecimento do *rapport*, bem como à orientação adequada do testando antes da administração do teste. O segundo aspecto, que não está sob o controle do examinador, é pertinente a todas as experiências de vida pregressas que os testandos podem ter tido e que podem afetar seu desempenho em um teste em particular.

Estabelecendo o *rapport*

No contexto da testagem, o termo *rapport* se refere à relação harmônica que deve existir entre testandos e examinadores. Para maximizar a fidedignidade e a validade dos resultados, uma atmosfera amigável precisa ser estabelecida desde o início da sessão de testagem, e o *rapport* idealmente deve variar de bom a excelente. A falta de *rapport* pode, é claro, ser atribuível a qualquer uma das partes na situação de testagem, e pode ter origem na inexperiência ou incompetência do examinador, uma disposição desfavorável por parte do testando, ou ambas. É desnecessário dizer que, se não há *rapport*, o desempenho no teste deverá ser afetado negativamente, até mesmo a ponto de os escores serem invalidados. Para construir o *rapport*, o examinador deve tentar despertar o interesse e a colaboração dos testandos no processo, para que eles reajam às tarefas propostas de maneira adequada, esforçando-se ao máximo nos testes de habilidade e respondendo aberta e honestamente aos instrumentos de avaliação da personalidade.

O processo de construção do *rapport* é mais extenso na testagem individual do que na grupal, porque aquela permite aos examinadores observar de perto e continuamente o comportamento do testando e fazer esforços para manter o *rapport* durante todo o processo da testagem. Não obstante, mesmo na testagem em grupo, o examinador deve tentar explicar os objetivos, procedimentos, etc. (ver o quadro Consulta Rápida 7.2), de forma amigável dentro dos limites impostos pelas instruções para a administração do teste fornecidos no manual, que devem ser seguidas para manter as condições uniformes.

Preparação para o teste do ponto de vista do testando

Não é preciso dizer que o objetivo da testagem e os aspectos contextuais da situação e das circunstâncias nas quais ela acontece têm uma influência significativa nas atitudes e na motivação dos testandos. Entre muitos outros fatores cognitivos e emocionais que afetam as predisposições dos testandos e o grau em que eles estão preparados para uma experiência de testagem, a ansiedade de teste e a sua sofisticação provavelmente são as mais discutidas e investigadas na literatura sobre a testagem psicológica. A seguir apresentamos um breve panorama de cada uma.

Ansiedade de teste. A perspectiva de ser avaliado tende a despertar um certo grau de apreensão na maioria dos testandos, especialmente quando os escores serão usados como base para a tomada de decisões que terão conseqüências importantes. Para alguns testandos esta apreensão é facilmente dissipada e elaborada – e pode até mesmo melhorar seu nível de desempenho aumentando sua estimulação fisiológica. Para outros, a ansiedade de teste se torna um fator debilitante ou mesmo incapacitante, que pode ter um efeito prejudicial significativo no desempenho no teste ou impedi-los de realizá-lo. Além disso, os motivos pelos quais os testandos experimentam ansiedade antes e durante um processo de testagem também podem variar muito. Algumas dessas razões podem estar relacionadas ao tipo de teste a ser realizado (p. ex., testes de velocidade, testes de matemática, etc.); outras, aos próprios testandos (p. ex., expectativas de fracasso devido a experiências anteriores), e ainda outras à função do contexto no qual a testagem acontece (p. ex., seleção para emprego) ou a uma combinação de variáveis. Embora os examinadores devam estar alertas para tentar reduzir o nível de ansiedade dos testandos, como parte do processo de construção do *rapport*, também existem medidas que os próprios testandos podem tomar para este fim. O quadro Consulta Rápida 7.6 lista recursos que podem ser úteis aos testandos e outros interessados nas abordagens atuais da avaliação e tratamento da ansiedade de teste.

Sofisticação de teste. Estritamente falando, a variável conhecida como *sofisticação de teste* (também denominada habilidade ou experiência de testagem) se refere ao grau de experiência ou prática dos testandos em se submeterem a testes. Como regra geral, na maioria dos tipos de teste de habilidade, ter tido a experiência de se submeter a um teste em particular tende a ser uma vantagem para o testando já que fornece prática e pode reduzir a ansiedade e aumentar a confiança. De fato, quando os indivíduos são retestados com a mesma forma ou uma forma alternativa de um teste de habilidade, seu segundo escore é quase invariavelmente mais alto do que o primeiro, um fenômeno conhecido como *efeito da aprendizagem*. Naturalmente, dada a enorme gama de tipos de testes e itens existentes, um testando pode ser muito sofisticado e experiente em um tipo de teste em uma determinada modalidade (p. ex., testes de realização de múltipla escolha de lápis e papel), mas nem um pouco familiarizado com testes de outros tipos em outros meios (p. ex., testes de habilidade geral administrados individualmente ou testes de desempenho administrados por computador). Da mesma forma, os testandos diferem muito em ter-

Fontes de informação sobre Teste de Ansiedade

Testandos que desejarem auxílio para lidar com a ansiedade de teste vão encontrar uma grande variedade de materiais disponíveis nas livrarias e na Internet. Os exemplos incluem:

- *Taking the anxiety out of taking tests: A step-by-step guide*, de S. Johnson. New York: Barnes & Noble Books, 2000.
- *No more test anxiety: Effective steps for taking tests and achieving better grades*, de E. Newman (disponível com CD de áudio). Los Angeles: Learning Skills Publications, 1996.
- The Test Anxiety Scale (Saranson, 1980), que oferece um modo rápido de medir o grau em que uma pessoa pode estar propensa a experimentar ansiedade de teste, está disponível gratuitamente no site da Learning Skills Publications (<http://www.learningskills.com/test.html>) e diversas outras páginas da Internet.
- Muitas páginas mantidas por centros universitários de aconselhamento podem ser acessadas pesquisando-se "ansiedade de teste" na Internet; estas páginas oferecem "dicas" de hábitos de estudo e outras informações para lidar com a ansiedade de teste.

Para aqueles que desejam se aprofundar nas teorias e pesquisas atuais sobre o tema, os seguintes trabalhos são recomendados:

- Sapp, M. (1999). *Test anxiety: Applied research, assessment, and treatment (2nd ed.)*. Latham, MD: University Press of America.
- Spielberger, C.D. e Vagg, P.R. (Eds.). (1995). *Test anxiety: Theory, assessment and treatment*. Washington, DC: Taylor & Francis.
- Zeidner, M. (1998). *Test anxiety: The state of the art*. New York: Plenum.

mos de seu espectro de experiências, que pode tê-los tornado mais ou menos preparados para uma experiência de testagem.

Dentro do processo de testagem em si, o modo tradicional de lidar com a variabilidade da sofisticação de teste tem sido fornecer instruções explícitas e itens de prática antes do teste como parte dos procedimentos padronizados de administração para garantir que os testandos sejam capazes de dominar adequadamente a mecânica das respostas. Embora estas sessões de orientação prática não possam de modo algum apagar as diferenças entre os testandos, elas podem pelo menos garantir que eles consigam manejar os procedimentos de testagem com competência. Além disso, sempre é bom que os examinadores indaguem a respeito das experiências anteriores dos testandos com o teste ou o tipo de teste que eles estão prestes a realizar e registrar esta informação para uso no momento da interpretação dos escores.

É claro que existem muitas outras vias que os testandos podem usar para adquirir habilidades e prática de testagem. Um dos métodos mais simples e mais eficientes para os potenciais candidatos a testes de admissão em faculdades e cursos de pós-graduação se familiarizarem com os itens e procedimentos destes testes é fazerem as amostras de itens fornecidas pelas editoras desses instrumentos (ver, p. ex., o quadro Consulta Rápida 6.2). Professores e orientadores também podem

ajudar a suprir seus alunos e clientes de informações e orientação sobre habilidades de testagem (Scruggs e Mastropieri, 1992). A importância que muitas pessoas dão à preparação para testes é sublinhada pelo fato de que esta se tornou um grande negócio, uma verdadeira indústria que oferece aos testandos uma multiplicidade de manuais, programas de computador, cursos e serviços tutoriais para uma variedade de programas de testagens. Se, e em que grau, os testandos conseguem obter ganhos significativos de escore por meio de vários métodos de treinamento e preparação para testes é uma questão muito discutida. Um relato interessante de algumas das questões e achados sobre este tópico é feito no capítulo sobre como o treinamento e as fraudes afetam o desempenho nos testes (*Gaming the Tests: How do Coaching and Cheating Affect Test Performance?*) do livro de Zwick (2002) *Fair game? The use of standardized admissions tests in higher education*.

Outro aspecto do debate sobre a preparação para testes diz respeito à distinção entre o treinamento intensivo voltado somente ou primariamente para a elevação dos escores, de um lado, e o ensino que contempla as metas mais amplas do currículo, de outro. Esta questão foi enfatizada pela prática bastante controversa da testagem obrigatória que muitos estados e localidades americanas estão instituindo com o objetivo de tornar as escolas públicas legalmente responsáveis pela aprendizagem dos alunos. Neste caso, como em muitos outros, considerações econômicas e políticas criam condições nas quais a finalidade mesma da testagem pode ser subvertida, e as responsabilidades das várias partes no processo educativo são desviadas dos usuários para os testes.

O problema da dissimulação do testando

Uma perspectiva inteiramente diferente das predisposições dos testandos se apresenta em situações nas quais o tipo de tomada de decisão para a qual um teste é usado promove a dissimulação. As tentativas da parte dos testandos de se apresentarem de maneira falsamente favorável ou desfavorável não são incomuns e podem ser conscientes ou não. Escalas de validade criadas para detectar vários tipos de tentativas de manipular impressões ou conjuntos de respostas, como postura defensiva, são embutidas em diversos inventários de personalidade (p. ex., o MMPI-II e o MCMI-III) e têm um longo histórico. Mais recentemente, instrumentos criados em especial para avaliar a possibilidade de dissimulação intencional em testes cognitivos administrados no contexto de avaliações forenses e neuropsicológicas – como o *Validity Indicator Profile* – foram acrescentados ao repertório de ferramentas disponíveis para este fim (R. Rogers, 1997). Sem dúvida alguma, obter a plena cooperação dos testandos no processo da testagem é uma questão crucial da qual dependem a fidedignidade e a precisão dos resultados obtidos (ver o quadro Consulta Rápida 7.3).

Para um levantamento das pesquisas sobre muitos tópicos adicionais que influenciam a perspectiva dos testandos sobre a testagem psicológica e educacional, os leitores podem consultar a obra organizada por Nevo e Jäger (1993) sobre este tema. Os estudos apresentados foram conduzidos por investigadores da Alemanha, Israel e Estados Unidos que procuravam coletar dados sobre opiniões, atitudes e

reações de examinandos a vários aspectos de suas experiências com a meta última de melhorar testes específicos e a testagem em geral.

A Preparação do examinador

Obtendo o consentimento informado

Segundo os *Princípios éticos dos psicólogos e código de conduta* (APA, 2002), antes da administração de um teste psicológico ou procedimento de avaliação, os psicólogos devem obter e documentar o *consentimento informado* dos testandos, oralmente ou por escrito. Para ser considerado *informado*, o consentimento obtido dos testandos deve ser precedido de uma explicação adequada da natureza e da finalidade da avaliação, bem como de informações a respeito dos limites do sigilo e de como a segurança dos resultados do teste será mantida. Outras questões práticas que podem afetar os testandos em uma situação específica (p. ex., valores cobrados, possibilidade recusar ou interromper a testagem, envolvimento de terceiros, etc.) também devem ser discutidas. Os testandos devem ter a oportunidade de fazer perguntas sobre o processo de testagem e receber respostas. O código de ética dos psicólogos (APA, 2002) lista algumas circunstâncias excepcionais em que o consentimento informado não é exigido, seja porque está implícito ou porque a avaliação é obrigatória. Medidas especiais relativas a estes casos, bem como a pessoas que não têm capacidade de consentir, também são descritas no código. Outras profissões envolvidas no uso da testagem psicológica e instrumentos de avaliação têm diretrizes semelhantes para a obtenção do consentimento informado (ver, p. ex., American Counseling Association, 1995).

A importância da preparação do examinador

O papel-chave da boa administração e avaliação dos testes na coleta de dados interpretáveis sempre deve ser enfatizado. Os procedimentos padronizados para a administração e a avaliação de um teste especificados em seu manual fornecem as bases que permitem a aplicação de um referencial normativo ou referenciado no critério para a interpretação dos escores. Com toda a certeza, desvios ou modificações dos procedimentos padronizados, às vezes, são inevitáveis ou necessários, como, por exemplo, quando a administração de um teste é atrapalhada por algum evento ou quando precisam ser feitas adaptações para testandos com deficiências. Nestes casos, os examinadores precisam documentar as modificações feitas. Além disso, se houver motivos para acreditar que o sentido dos escores pode ser afetado por perturbações ou modificações nos procedimentos padronizados, a natureza destas deve ser relatada aos responsáveis pela interpretação dos escores e pela tomada de decisões neles baseadas (AERA, APA, NCME, 1999).

No que diz respeito à preparação dos examinadores, a pessoa que administra um teste tem que estar totalmente familiarizada com os objetivos e procedimentos do instrumento, ser capaz de estabelecer *rappor*t e estar pronta para responder as

perguntas dos testandos ou lidar com qualquer emergência previsível que possa surgir durante a testagem. De modo geral, a administração de testes em grupo não requer qualquer treinamento adicional além do já mencionado. A testagem individual, por outro lado, inevitavelmente envolve muito mais preparação e geralmente requer experiência supervisionada. Por exemplo, quando as perguntas de um teste são apresentadas oralmente, elas devem ser formuladas literalmente. Para que a administração ocorra sem problemas e no ritmo correto e permita ao examinador a oportunidade de escutar atentamente e registrar as respostas (também literalmente), os examinadores precisam saber de cor não apenas os itens e sua seqüência, mas também uma série de regras adicionais para o *timing* da apresentação dos itens e pontos de início e parada, bem como para a avaliação das respostas à medida que são produzidas. Quando são usados materiais como peças de quebra-cabeças, blocos, figuras e outros objetos na testagem individual, eles devem ser apresentados e removidos da maneira exata determinada pelo manual. A menos que os examinadores memorizem e pratiquem todos os procedimentos do teste com antecedência até estarem totalmente proficientes, a administração e possivelmente também o desempenho do testando serão afetados de maneiras que podem facilmente pôr em risco a validade dos dados obtidos.

Administração de testes computadorizados

Uma forma certa de evitar a possibilidade de erros na administração de testes, bem como a de que o desempenho do testando seja influenciado indevidamente por variáveis relacionadas ao sexo, raça, idade, aparência e estilo interpessoal do testando e outras variáveis do tipo é por intermédio da administração de testes computadorizados. As vantagens deste meio de testagem com relação à relativa uniformidade na apresentação dos materiais de teste e à precisão com que o tempo das respostas podem ser controladas, registradas e avaliadas são evidentes. Na verdade, a tendência de substituir a testagem de lápis e papel pela administração de testes computadorizados está bem estabelecida, especialmente nos programas de testagem de larga escala, e deve continuar a se expandir porque a informatização se torna mais barata. Além disso, o desenvolvimento e o uso de testes – como o *Test of Variables of Attention* – que, por natureza, só podem existir em versões computadorizadas, também estão progredindo rapidamente. Por outro lado, a administração de testes individuais de habilidade geral, técnicas projetivas e muitos testes neuropsicológicos (desde que seja feita por um examinador treinado) tem algumas vantagens claras em termos dos dados qualitativos que podem ser coletados, especialmente no contexto de avaliações clínicas, neuropsicológicas e forenses. Além disso, antes de instituir uma mudança dos métodos tradicionais de administração para os métodos computadorizados, a comparabilidade dos resultados nas duas condições – que varia dependendo do tipo de instrumento em questão – precisa ser plenamente investigada (Mead e Drasgow, 1993). Para mais informações quanto à avaliação psicológica computadorizada, incluindo a questão da equivalência entre a administração de testes computadorizados e do tipo padrão, ver Butcher (2003).

Avaliação

Podemos recordar que as diferenças entre avaliadores e o tópico associado da fidedignidade do avaliador – discutidos no Capítulo 4 entre as possíveis fontes de erro de mensuração – foram considerados pertinentes apenas para a avaliação de respostas abertas, nas quais a subjetividade do avaliador desempenha algum papel. Quando isso acontece, os usuários de testes precisam determinar empiricamente que aqueles que vão avaliar as respostas – de testagens individuais ou grupais – têm treinamento suficiente para chegar a resultados virtualmente idênticos aos produzidos por um avaliador independente e experiente. Como regra geral, quando os coeficientes de fidedignidade entre avaliadores podem ser calculados, eles devem se aproximar de +1,00 e não ficar muito abaixo de 0,90. Quando os escores são expressos de alguma outra maneira, o objetivo deve ser chegar o mais perto possível de 100% de concordância nos escores determinados por avaliadores treinados independentes. Deve-se observar que a facilidade com que padrões tão altos de precisão conseguem ser alcançados quando a avaliação requer julgamentos subjetivos varia muito entre diferentes tipos de testes. Isso acontece porque as diretrizes para a avaliação de algumas respostas abertas (p. ex., respostas dissertativas em testes de realização) tendem a ser mais claras e fáceis de dominar do que outras (p. ex., respostas para técnicas projetivas como o Rorschach). Para uma discussão de questões relacionadas à fidedignidade de avaliador para o Sistema Rorschach (ver Acklin, McDowell, Verschell e Chan, 2000).

Em testes avaliados objetivamente – apenas contando-se as respostas em várias categorias e realizando-se os cálculos necessários para transformar os escores brutos em alguma outra forma numérica, como escores padrões de vários tipos – os erros de avaliação não devem ocorrer. Uma forma de evitá-los é pela administração e avaliação computadorizada. Outra é pelo uso de escâners óticos e programas de computador apropriados, embora isso demande um exame cuidadoso das folhas de resposta para se identificarem rasuras e outros problemas antes do escaneamento. Se a avaliação objetiva é feita à mão, especialmente com gabaritos sobrepostos às folhas de resposta, a possibilidade de erros de verificação tem que ser evitada ao máximo, enfatizando-se para os avaliadores a absoluta importância da precisão e instituindo-se procedimentos de verificação dupla de todos os cálculos e transformações de escore necessários.

A transformação dos escores brutos de testes administrados individualmente, como as escalas Wechsler, em escores padrões envolve tipicamente uma série de cálculos aritméticos, bem como a busca de equivalentes dos escores brutos em várias tabelas fornecidas pelos manuais. Para evitar erros resultantes de descuidos nestes procedimentos, as editoras também oferecem programas de computador que podem realizar todos os cálculos e transformações necessários depois que os escores brutos obtidos pelo examinador são (cuidadosamente) informados ao computador. Se os usuários de testes não puderem se valer deste tipo de ferramenta de auxílio à avaliação, as boas práticas de testagem requerem a verificação dupla da precisão de todos os cálculos e escores equivalentes obtidos com as tabelas.

FUNDAMENTOS EM INTERPRETAÇÃO E RELATO DE ESCORES DE TESTE

A maior parte do material precedente neste livro teve como objetivo transmitir aos leitores as complexidades envolvidas na aplicação e no uso adequado dos dados de testes psicológicos. A interpretação de escores e a comunicação das inferências obtidas com a testagem psicológica são os pontos culminantes do uso de testes, e também os componentes da testagem dos quais os testandos podem derivar os maiores benefícios ou os piores danos. Uma vez que todos os vários princípios envolvidos na interpretação e relato de resultados de testes não podem ser tratados em um único livro – e muito menos em uma única seção de um capítulo – vamos apresentar uma perspectiva geral destes aspectos da testagem e referências selecionadas sobre tópicos relacionados.

Uma perspectiva particular sobre a interpretação de testes

Os escores de testes psicológicos fornecem dados quantitativos mais ou menos fidedignos que *descrevem* de maneira concisa o comportamento produzido pelos indivíduos em resposta aos estímulos do teste. Quando são cuidadosamente selecionados, administrados e avaliados, os testes fornecem informações que podem ser usadas em uma variedade de formas, sendo a mais básica simplesmente localizar o desempenho do testando dentro das categorias descritivas ou normativas ou baseadas no critério, fornecidas pelo referencial empregado pelo teste (ver Capítulo 3). Se forem suficientemente fidedignos e apropriadamente interpretados, os dados de escores de um teste também podem auxiliar tanto a explicar a constituição psicológica dos indivíduos como a se tomar decisões a respeito deles com base em estimativas a respeito de suas características ou comportamentos futuros (ver Capítulo 5). Porém, para obter respostas defensáveis às perguntas complexas endereçadas aos psicólogos, professores e outros profissionais das áreas humanas, dados de múltiplas fontes muitas vezes são necessários, e o julgamento informado *sempre* deve ser exercido.

Fazer julgamentos informados a respeito de pessoas requer uma apreciação do valor e das limitações inerentes ao referencial, fidedignidade e validade dos escores de teste, bem como aos dados de todas as outras fontes que podem ser empregadas em um determinado caso. Além disso, requer conhecimento do contexto e de áreas específicas do comportamento humano relevantes ao assunto em questão. Tomar decisões a respeito de pessoas envolve invariavelmente julgamentos de valor por parte dos decisores e joga sobre eles uma responsabilidade ética pelas conseqüências das decisões tomadas. Infelizmente, na prática, as questões contextuais freqüentemente são ignoradas, os julgamentos de valor não são reconhecidos explicitamente e os escores de teste muitas vezes se tornam o principal ou mesmo o único fator determinante na tomada de decisões. Como resultado, por motivos de conveniência, o peso da responsabilidade por muitas decisões é transferido injustificavelmente dos usuários de testes e responsáveis pelas decisões para os testes em si. Uma forma de combater este problema é compreender e apreciar as

implicações da diferença entre testagem e avaliação psicológica, discutida no Capítulo 1. Esta diferença é semelhante à distinção entre conduzir exames médicos, por um lado, e integrar seus resultados com a anamnese e os sintomas atuais de um paciente para produzir um diagnóstico e um plano de tratamento, por outro (Handler e Meyer, 1998).

O panorama de vantagens e limitações da testagem e de outras ferramentas de avaliação, apresentado anteriormente neste capítulo, tentava transmitir a desejabilidade de integrar o máximo de evidências possíveis sempre que existe uma questão significativa que demanda a avaliação de indivíduos ou grupos. Com respeito à interpretação e ao uso de testes, a perspectiva em particular apresentada aqui é a seguinte:

1. Os testes psicológicos por vezes podem ser as ferramentas mais eficientes e objetivas disponíveis para a coleta de dados fidedignos e válidos a respeito de pessoas.
2. A testagem psicológica muitas vezes pode ser um componente valioso do processo de avaliação de indivíduos e grupos.
3. Os escores de testes psicológicos nunca devem ser a única fonte de informações sobre a qual basear decisões que afetam a vida de indivíduos.

O modo específico como a interpretação e o relato de escores de testes deve ser conduzido depende de dois fatores inter-relacionados: (a) a finalidade da testagem e (b) a parte em benefício de quem a testagem foi realizada. Em relação ao segundo fator, três possibilidades distintas determinam como os dados de testes são interpretados e comunicados.

- *Quando os psicólogos usam testes por sua própria iniciativa* (p. ex., como instrumentos de pesquisa), eles podem interpretar e relatar os dados agrupados dos participantes da pesquisa da maneira que acharem apropriada para os fins de sua investigação. Os requisitos legais e padrões éticos para pesquisas com participantes humanos regem as práticas neste tipo de testagem e incluem medidas para a obtenção do consentimento informado dos participantes e fornecimento de qualquer esclarecimento solicitado após a participação no estudo (APA, 2002). Com relação ao papel da testagem dentro do projeto de pesquisa em si, questões como a escolha do instrumento, o modo como os escores são relatados e o sentido atribuído a eles devem ser avaliadas à luz dos objetivos da investigação. A pesquisa e a publicação científicas são essencialmente iniciativas auto-reguladas, que usam o mecanismo da revisão por pares para avaliar os méritos substantivos e metodológicos de trabalhos propostos, em andamento ou concluídos. Por isso, quando pesquisas que incluem o uso de dados de testes são submetidas a revisões ou avaliadas para publicação, elas serão aceitas, rejeitadas, citadas ou ignoradas com base – entre outras coisas – no modo mais ou menos correto como estes dados foram empregados.
- *Quando os psicólogos usam testes em benefício de seus próprios clientes*, eles são exclusivamente responsáveis pela interpretação dos dados, integração

com outras fontes de informação e comunicação de seus achados aos clientes, de maneira útil e apropriada. Quer estas avaliações tenham como objetivo o diagnóstico, o planejamento do tratamento, o monitoramento do progresso ou a facilitação de mudanças, o cliente é o consumidor final das informações e o juiz último de seus benefícios. Na verdade, em muitas dessas situações, a interpretação dos escores, no que diz respeito às implicações destes, pode ser um esforço colaborativo entre o psicólogo ou terapeuta e o cliente (Fischer, 2000). O quadro Consulta Rápida 7.7 lista recursos que fornecem instruções explícitas para o uso de testes no contexto do aconselhamento e da avaliação clínica.

- *Quando os psicólogos usam testes a pedido de terceiros*, como uma organização ou outro profissional, estão agindo como consultores em um processo de avaliação iniciado por outros para seus próprios fins. Nestes casos, os limites da responsabilidade não são tão bem definidos quanto no exemplo anterior, porque o testando submetido à avaliação não é necessariamente o consumidor final dos dados do teste. Não obstante, como já foi mencionado, tanto do ponto de vista da ética quanto para a proteção de possíveis responsabilidades legais, os consultores devem ter clara finalidade do trabalho, não apenas em termos de quais informações o contratante está buscando, mas também de como elas serão usadas. Somente com este conhecimento os profissionais da avaliação poderão determinar se o uso de testes é justificado, como os escores devem ser relatados, que outras ferramentas podem ser necessárias para derivar as informações buscadas e se eles são capazes e estão dispostos a participar do processo de consultoria. A interpretação de resultados de testes envolve uma série de inferências feitas com base nos dados coletados a partir (a) das amostras de comportamento (respostas aos itens de testes), (b) da agregação destas amostras em um ou mais escores, (c) das evidências disponíveis da fidedignidade

CONSULTA RÁPIDA 7.7

Recursos sobre o uso de testes no aconselhamento e na prática clínica

Os artigos e livros que fornecem orientação e exemplos de interpretação e uso de testes psicológicos na prática clínica são muitos, e seu número está em constante crescimento. Os trabalhos incluídos nesta pequena lista apresentam apenas algumas das muitas formas possíveis de aplicar a testagem no contexto do aconselhamento e da clínica. Recursos para aplicações e populações mais especializadas estão incluídos no quadro Consulta Rápida 7.8.

- Beutler, L.E. e Groth-Marnat, G. (Eds.). (2003). *Integrative assessment of adult personality* (2nd ed.). New York: Guilford.
- Fischer, C.T. (1994). *Individualizing psychological assessment*. Hillsdale, NJ: Erlbaum. (Obra original publicada em 1985)
- Lowman, R.L. (1991). *The clinical practice of career assessment: Interests, abilities, and personality*. Washington, DC: American Psychological Association.
- Maruish, M.E. (Ed.). (2004). *The use of psychological testing for treatment planning and outcome assessment* (3rd ed., Vols. 1-3). Mahwah, NJ: Erlbaum.

dos escores obtidos, (d) da comparação dos escores com os referenciais normativos ou baseados no critério fornecidos pelo teste, (e) da avaliação dos escores à luz da qualidade dos dados de validação interna e externa disponíveis, (f) do contexto e da situação específica na qual a testagem acontece e (g) das características pessoais do indivíduo que está sendo avaliado. Quando todas estas fontes de evidências são vistas em conjunção e somadas às informações pertinentes coletadas por outros métodos, suas implicações e limitações com respeito a questões específicas de avaliação devem estar claras. Ocasionalmente (p. ex., quando dados de várias fontes fidedignas e presumivelmente válidas são mutuamente contraditórios ou quando há motivos para crer que algumas peças-chave das evidências – sejam derivadas dos testes ou de outras fontes – não são fidedignas), a natureza inconclusiva dos achados deve ser relatada. Nestes casos, as recomendações apropriadas incluem o encaminhamento a outro profissional ou uma coleta adicional de dados. O quadro Consulta Rápida 7.8 lista alguns dos muitos recursos atualmente disponíveis sobre a interpretação de testes e tópicos relacionados.

CONSULTA RÁPIDA 7.8**Mais informações sobre interpretação e avaliação de testes**

As fontes primárias de informação sobre o uso e a interpretação de testes e ferramentas de avaliação específicos são os manuais, livros de instruções e outras documentações de apoio fornecidas pelos autores e editoras dos testes. Além disso diretrizes sobre a interpretação de testes podem ser encontrados em um grande número de livros e publicações. Aqui estão alguns exemplos.

Para um panorama geral da interpretação de testes:

- Lichtenberg, J.W. e Goodyear, R.K. (Eds.). (1999). *Scientist-practitioner perspective on test interpretation*. Boston: Allyn & Bacon.

Para questões relacionadas à avaliação de crianças e adolescentes:

- Kamphaus, R.W. (2001). *Clinical assessment of child and adolescent intelligence* (2nd ed.). Boston: Allyn & Bacon.
- Kamphaus, R.W. e Frick, P.J. (2002). *Clinical assessment of child and adolescent personality and behavior* (2nd ed.). Boston: Allyn & Bacon.
- Sattler, J.M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego, CA: Author.
- Sattler, J.M. (2002). *Assessment of children: Behavioral and clinical applications* (4th ed.). San Diego, CA: Author.

Para perspectivas sobre a interpretação de testes para diversas populações:

- Sandoval, J., Frisby, C.L., Geisinger, K.F., Scheuneman, J.D., & Grenier, J.R. (1998). *Test interpretation and diversity: Achieving equity in assessment*. Washington, DC: American Psychological Association.
- Ekstrom, R.B., e Smith, D.K. (Eds.). (2002). *Assessing individuals with disabilities in educational, employment and counseling settings*. Washington, DC: American Psychological Association.

Comunicando resultados de testes e achados de avaliações

A diretriz mais básica a ser seguida ao se comunicar resultados de testes é fornecer as informações derivadas dos escores, incluindo suas limitações, em uma linguagem que o interlocutor possa compreender. No entanto, a maneira específica como os escores são relatados pode variar muito dependendo dos testes administrados, do contexto no qual a testagem acontece, dos objetivos desta e dos beneficiários pretendidos da informação. Portanto, o modo apropriado de relatar os resultados da testagem psicológica ou os achados de uma avaliação não pode ser condensado em um único conjunto de regras aplicáveis em todos os casos. Não obstante, questões pertinentes a alguns dos vários modos de comunicar resultados de testes são apresentadas nos parágrafos a seguir.

- Quando os resultados são comunicados diretamente aos testandos em boletins ou perfis de resultados produzidos por computador, a organização responsável pelo programa de testagem deve fornecer informações interpretativas adequadas. Por exemplo, o site da College Board na internet (<http://www.collegeboard.com/student/testing/sat/scores/understanding.html>) traz alguns materiais pertinentes à interpretação dos escores do SAT na seção “*Understanding Your Scores*” (Compreenda seus escores). Ferramentas de auxílio semelhantes estão disponíveis para a interpretação de escores derivados de outros programas de testagem em larga escala, como o ACT (em <http://www.act.org/aap/scores/under/html>)
- *Interpretações de escores derivados de programas de computador* aplicam regras de decisão baseadas na experiência clínica e no julgamento de especialistas, abordagens atuariais baseadas em associações estatísticas e correlações entre escores e critérios, ou ambas. Como e por quem estas interpretações computadorizadas de testes (*computer-based test interpretations*; CBTIs) são empregadas é objeto de contínuas discussões entre profissionais da testagem. Originalmente criadas como auxiliares para os clínicos na interpretação de inventários de personalidade e outros instrumentos diagnósticos, as CBTIs proliferaram rapidamente e foram aplicadas a diferentes tipos e usos de testes. Um dos problemas causados por esses avanços é que os serviços de CBTI com frequência são disponibilizados comercialmente para indivíduos que podem ter credenciais aparentemente apropriadas – como licenças para praticar psicologia ou diplomas médicos –, mas que não têm conhecimento suficiente sobre o teste em questão. Como consequência, os relatos gerados pela CBTI são vistos de maneira acrítica e usados como substitutos para a avaliação individualizada baseada em múltiplas fontes de dados e informada pelo julgamento de um profissional qualificado. Além disso, a natureza comercial regida por leis de propriedade intelectual das regras específicas usadas na geração destes relatórios muitas vezes impede a avaliação apropriada da validade de suas interpretações pelos usuários. Não obstante, os profissionais que se valem dos serviços de CBTI estão sujeitos às mesmas obrigações éticas com relação à competência que no uso de qualquer outra ferramenta de avaliação. Os

prestadores de serviços de CBTL, por sua vez, têm a obrigação ética de fornecer a seus usuários as informações pertinentes às fontes, bases de evidências e limitações das interpretações que produzem (APA, 2002; Moreland, 1991).

- *O meio tradicional de comunicar resultados de testes e achados de avaliações é o laudo psicológico por escrito.* Esta abordagem individualizada permite uma grande flexibilidade na adaptação dos relatórios aos objetivos da avaliação e às necessidades dos clientes e consumidores dos dados de testes psicológicos. Também ajuda o avaliador a organizar, esclarecer e sintetizar as informações de todas as fontes disponíveis e cria um registro dos achados de uma avaliação que pode ser consultado no futuro. Uma descrição abrangente das várias abordagens à elaboração dos laudos psicológicos e seus benefícios e problemas, bem como numerosos exemplos de como redigi-los são fornecidos por Norman Tallent (1993). Um guia mais condensado para a elaboração passo a passo de laudos, escrito por Raymond Ownby (1997), também merece a atenção dos leitores interessados neste tópico. O quadro Consulta Rápida 7.9 lista alguns exemplos do que fazer e do que evitar na interpretação de testes que são especialmente pertinentes à elaboração de laudos psicológicos.

CONSULTA RÁPIDA 7.9

O que fazer e o que evitar na interpretação de testes

Alguns exemplos do que não é interpretação de testes

- *Relatar escores numéricos:* Quer os escores sejam expressos como postos de percentil, QIs, escores de percentagem ou algum outro formato, simplesmente listá-los não basta para transmitir seu sentido e suas implicações.
- *Atribuir rótulos:* Classificar indivíduos em categorias diagnósticas (p. ex., Retardo Mental Leve ou Personalidade Borderline), ou tipologias, com base em seus escores, não é um substituto adequado para a interpretação que melhora a compreensão.
- *Formular achados em termos de generalidades triviais:* Muitas afirmações podem se aplicar igualmente bem a quase todos os seres humanos ou à maioria dos indivíduos em certas categorias (p. ex., pacientes psiquiátricos, crianças pequenas, etc.) devido às suas altas taxas de base nestas populações. Meehl (1956) propôs o termo "efeito Barnum" para caracterizar a natureza inútil deste tipo de descrição.

Como deve ser a interpretação de testes

- *No mínimo,* a interpretação de escores de teste para os consumidores de seus dados deve incluir uma explicação clara (a) do que o teste trata, (b) do sentido dos escores, (c) das limitações da precisão dos escores derivada do erro de mensuração, (d) de interpretações errôneas comuns de certos escores – como os QIs, por exemplo e (e) do modo como os resultados podem ser ou serão usados (AERA, APA, NCME, 1999, Capítulo 5).
- *Na melhor das hipóteses,* a interpretação de escores de testes agrega valor às amostras de comportamento coletadas com os testes, integrando-os com todos os outros dados disponíveis e usando o julgamento profissional informado para chegar a inferências úteis e ecologicamente válidas.

Garantindo a segurança dos dados de testes

A segurança dos dados de teste, quer eles consistam em registros, escores e laudos individualmente identificáveis ou nos materiais dos testes em si (p. ex., livretos, formulários, perguntas, chaves de respostas, manuais, etc), é uma responsabilidade primária dos usuários e instituições que controlam o acesso aos mesmos. Em certos casos, esta responsabilidade pode ser de difícil cumprimento porque (a) muitas exigências legais, requisitos institucionais, padrões profissionais e preocupações éticas diferentes e cambiantes regem a decisão de quando, como e para quem os dados de teste podem ser divulgados e (b) estas várias injunções podem diferir substancialmente, dependendo do contexto e da finalidade da testagem, e até mesmo entrar em conflito umas com as outras. Mais recentemente, por exemplo, o Estatuto da Privacidade da Lei da Portabilidade e Responsabilidade dos Seguros de Saúde (*Health Insurance Portability and Accountability Act [HIPAA] Privacy Rule*), que passou a vigorar nos Estados Unidos em 14 de abril de 2003, impôs exigências que estão levantando muitas questões e criando uma certa confusão entre os profissionais da testagem e da avaliação, bem como entre outros profissionais da saúde. Algumas respostas às dúvidas mais frequentes e mais informações sobre como a HIPAA afeta os psicólogos podem ser encontradas no *site* do *APA Insurance Trust* na internet (<http://www.apait.org/resources/hipaa/>).

Via de regra, as exigências legais que abordam a divulgação de dados de teste têm precedência sobre os estatutos e interesses de outras partes. No entanto, os usuários precisam estar cientes de seus deveres de (a) proteger o sigilo dos resultados e a segurança dos materiais de teste em contextos e situações específicas, (b) informar as partes envolvidas a respeito destes deveres, e (c) tentar resolver conflitos entre suas várias obrigações, de maneira consistente com os princípios éticos e padrões de sua profissão. Além dos *Padrões de Testagem* (AERA, APA, NCME, 1999) e dos *Princípios éticos dos psicólogos e código de conduta* (APA, 2002), os psicólogos organizados como classe publicaram uma série de documentos para auxiliar seus membros a identificarem e lidarem com questões relacionadas à segurança dos testes e seus dados e a prevenirem o mau uso dos mesmos. Alguns dos mais importantes entre estes documentos estão listados no quadro Consulta Rápida 7.10.

Um novo meio: A testagem na internet

O advento e a rápida expansão da internet nas duas últimas décadas estão revolucionando o campo da testagem psicológica, tanto quanto têm afetado quase todos os aspectos da sociedade contemporânea. Os leitores devem ter notado, espalhadas por todo este livro, um bom número de referências a páginas da internet que oferece informações sobre testes psicológicos e questões relativas à testagem. Além destas e das muitas outras fontes de informação sobre testes e testagem disponíveis *on-line* atualmente, a oferta de serviços de testagem psicológica na internet por uma variedade de provedores – alguns legítimos e outros nem tanto – também se verifica. Ainda assim, o impacto que a Internet deverá ter na testagem promete ser muito maior nas próximas décadas, devido basicamente à velocidade com que

Informações sobre como manter a segurança de testes e dados de testes

Além dos *Padrões de Testagem* (AERA, APA, NCME, 1999) e dos *Princípios éticos dos psicólogos e código de conduta* (APA, 2002), que tratam de questões ligadas à segurança de materiais de teste e seus dados, a APA orienta os usuários nestas questões em vários outros documentos. Os seguintes estão entre os mais pertinentes:

- American Psychological Association, Committee on Legal Issues. (1996). Strategies for private practitioners coping with subpoenas or compelled testimony for client records or test data. *Professional Psychology: Research and Practice*, 27, 245-251.
- American Psychological Association, Committee on Psychological Tests and Assessment. (1996). Statement on the disclosure of test data. *American Psychologist*, 51, 644-648.
- American Psychological Association, Committee on Psychological Tests and Assessment. (2003). *Statement on the use of secure psychological tests in the education of graduate and undergraduate psychology students*. Retirado em 19 de fevereiro de 2003 de <http://www.apa.org/science/securetests.html>
- Test security: protecting the integrity of tests. (1999). *American Psychologist*, 54, 1078.

os testes podem ser desenvolvidos, normatizados, publicados e revisados usando-se o potencial da rede mundial de computadores, bem como à eficiência e à economia com que os serviços de testagem podem ser oferecidos *on-line*.

Em todos os aspectos do uso da internet, seja para fins de comércio, entretenimento, informação ou comunicação, a natureza altamente democrática e acessível deste meio representa ao mesmo tempo imensas vantagens e perigos. Como o leitor deve ter captado, com relação à testagem, as possibilidades de usos prejudiciais apresentadas pela internet são abundantes, assim como seu potencial para avanços sem precedentes. Por isso, o exame profissional das dificuldades e benefícios inerentes às práticas de testagem psicológica atuais e previsíveis para o futuro na Internet já começaram (Buchanan, 2002). Um dos esforços mais meticulosos neste sentido até o momento é um relatório preparado pela Força-Tarefa da APA sobre Testagem Psicológica na Internet (Naglieri et al., 2004). Este relatório se concentra em uma variedade de aspectos relacionados à influência da internet na testagem e avaliação psicológica, incluindo questões psicométricas, éticas e legais, bem como considerações práticas quanto à segurança e acesso aos testes, entre outras coisas. Também contém exemplos que ilustram as aplicações atuais da testagem na internet e oferece vislumbres das possibilidades futuras criadas por este meio e recomendações para o futuro.

CONCLUSÃO

Considere as seguintes perguntas e problemas típicos:

- Meu filho tem transtorno de déficit de atenção?
- Como podemos selecionar os melhores candidatos para nosso departamento de polícia?

Pondo em Prática

Como testar sua capacidade de avaliar testes psicológicos

Uma série de páginas de Internet oferece levantamentos, questionários e outras ferramentas que afirmam ser testes psicológicos. Praticamente qualquer ferramenta de busca vai levá-lo a estas páginas se você digitar os termos "testes psicológicos" ou "testes de personalidade". Entre as muitas ofertas está uma variedade de instrumentos como "testes de QI", "testes de inteligência emocional" e "testes vocacionais", que podem ter aparência de testes psicológicos e apelo para leigos, mas pouca ou nenhuma base científica. Alguns destes ditos testes são oferecidos gratuitamente, outros envolvem o pagamento de uma taxa (geralmente) modesta de US\$ 5,00, US\$ 10,00 ou US\$ 15,00. Em muitos casos, os instrumentos gratuitos não incluem resultados completos (ou mesmo qualquer resultado), a menos que uma taxa seja paga primeiro. Muitas vezes os testandos não são informados disso antes de terem completado o teste.

Depois de ler com cuidado este capítulo, você deve estar ciente de que as editoras de testes psicológicos legítimos são extremamente criteriosas com a questão da segurança dos testes e não desejam ver seus produtos disseminados para o público em geral ou usados de maneira imprópria. Com raras exceções (p. ex., instrumentos em processo de desenvolvimento), não se encontram editoras ou autores sérios disponibilizando seus produtos livremente na Internet.

Uma forma de aplicar parte dos conhecimentos que você obteve lendo este livro é inspecionar alguns instrumentos oferecidos na Internet e avaliá-los com relação a (a) se eles satisfazem os critérios que os qualificam como testes psicológicos legítimos e (b) se as informações que eles fornecem têm algum valor.

- Este paciente de 65 anos está sofrendo de um transtorno incipiente de demência ou está deprimido devido à morte da esposa?
- Qual destes três candidatos igualmente experientes devo contratar para o cargo de diretor financeiro da minha empresa?
- Que faculdade devo escolher?
- Este estuprador pode ganhar liberdade condicional?
- Que tipo de intervenção terapêutica seria mais eficaz para este paciente?
- Este indivíduo está simulando invalidez para ganhar uma indenização ou está realmente inválido?

Embora todas essas perguntas e problemas de avaliação possam ser abordados com a ajuda de testes, nenhum pode ser respondido com um único escore ou mesmo com uma combinação deles. As determinações que precisam ser feitas nestas situações e outras semelhantes requerem dados de múltiplas fontes, bem como informações sobre os objetivos específicos das pessoas envolvidas, os contextos em que as perguntas são feitas e as conseqüências potenciais das decisões em cada caso.

A esta altura, deve estar claro que a aplicação de testes psicológicos é passível de erros a todo momento. No entanto, o mesmo se aplica a todas as iniciativas humanas. Uma vez que a pergunta essencial seja respondida afirmativamente (isto

é, se o uso de testes pode contribuir para uma tomada de decisões mais racional, equânime, benéfica e responsável em uma determinada situação), passa a ser necessário implementar o plano para um uso de testes o mais consciencioso possível do início ao fim. Se isto for feito, o papel dos testes, especialmente se comparado a outras ferramentas, poderá se mostrar bastante útil.

Teste a si mesmo

1. A primeira questão a considerar quando a testagem psicológica é contemplada é
 - (a) se a testagem é necessária
 - (b) que tipo de teste usar
 - (c) quais os custos associados à testagem
2. Qual das seguintes não é uma das vantagens primárias do uso de testes psicológicos?
 - (a) objetividade
 - (b) eficiência
 - (c) fácil disponibilidade
3. Não havendo outras falhas, o ganho potencial de precisão em uma decisão de seleção é maior quando a taxa de base é mais próxima de
 - (a) 1,00
 - (b) 0,75
 - (c) 0,50
 - (d) 0,00
4. Uma situação ideal, para fins de precisão na seleção de funcionários, envolveria uma taxa de base _____, uma razão de seleção _____ e um teste com grau de validade _____.
 - (a) moderada/alta/alto
 - (b) moderada/baixa/alto
 - (c) alta/baixa/moderado
 - (d) alta/alta/alto
5. Qual das seguintes ferramentas de avaliação têm maior probabilidade de fornecer informações válidas e fidedignas para uma avaliação individual?
 - (a) entrevista
 - (b) referências
 - (c) observação informal
 - (d) biodados
6. Como regra geral, a presença de terceiros, além do examinador e do testando, durante a administração de testes individuais, é
 - (a) desejável
 - (b) indesejável
 - (c) nem a nem b

7. Qual das seguintes não é uma das áreas na qual a administração de testes computadorizados oferece vantagens em relação à administração individual?
 - (a) uniformidade de procedimentos
 - (b) relação custo-benefício
 - (c) capacidade de precisão
 - (d) coleta de dados qualitativos

8. A perspectiva da testagem apresentada pela autora enfatiza o fato de que os testes psicológicos
 - (a) podem ser um componente valioso na maioria dos casos em que uma avaliação é necessária
 - (b) podem muitas vezes ser a única fonte de informações para basear decisões
 - (c) invariavelmente constituem as ferramentas mais eficientes disponíveis para avaliações

9. Ao comunicar resultados de testes aos consumidores de seus dados, as informações mais pertinentes a serem transmitidas são
 - (a) os escores numéricos obtidos pelos examinandos
 - (b) os rótulos ou diagnósticos derivados dos escores
 - (c) o sentido dos escores

10. As responsabilidades éticas e legais dos usuários de testes com respeito à interpretação apropriada dos resultados são explicitadas
 - (a) quando eles usam testes por iniciativa própria
 - (b) quando eles usam testes a pedido de terceiros
 - (c) quando eles usam testes em benefício de seus próprios clientes
 - (d) nunca

Respostas: 1. a; 2. c; 3. c; 4. b; 5. d; 6. b; 7. d; 8. a; 9. c; 10. d.