

CAPÍTULO 1

Testagem e Avaliação Psicológica

Todos os campos de empreendimento humano usam medidas de alguma forma, e cada um tem seu próprio conjunto de instrumentos de mensuração e unidades de medida. Por exemplo, se você noivou recentemente ou está pensando em noivar, pode ter ouvido falar sobre uma unidade de medida chamada quilate. Se você comprou um computador, pode ter ouvido alguma coisa sobre uma unidade de medida chamada *byte*. Como aluno de medidas psicológicas, você necessita de familiaridade operacional com algumas das unidades de medida que costumam ser usadas em psicologia, bem como de conhecimento sobre alguns dos muitos instrumentos de mensuração empregados. Nas páginas a seguir, você obterá esse conhecimento e também familiaridade com a história da medida em psicologia e entendimento de sua base teórica.

Testagem e avaliação

As raízes da testagem e da avaliação psicológicas contemporâneas podem ser encontradas na França do início do século XX. Em 1905, Alfred Binet e um colega publicaram um teste concebido para ajudar a colocar as crianças das escolas de Paris em classes apropriadas. O teste de Binet teria consequências bem além do distrito escolar de Paris. No período de uma década, uma versão para a língua inglesa do teste de Binet foi preparada para ser usada em escolas nos Estados Unidos. Quando os Estados Unidos declararam guerra à Alemanha e entraram na Primeira Guerra Mundial, em 1917, os militares precisavam de uma maneira de avaliar rapidamente grandes números de recrutas em relação a problemas intelectuais e emocionais. A testagem psicológica forneceu essa metodologia. Durante a Segunda Guerra Mundial, os militares dependiam ainda mais dos testes psicológicos na avaliação de recrutas para o serviço. Depois da guerra, cada vez mais testes com o propósito de medir uma série crescente de variáveis psicológicas foram desenvolvidos e usados. Havia testes para medir não apenas a inteligência, mas também a personalidade, o funcionamento do cérebro, o desempenho no trabalho e muitos outros aspectos do funcionamento psicológico e social.

Definição de testagem e avaliação psicológica

A receptividade mundial ao teste de Binet no início do século XX produziu não apenas mais testes, mas mais desenvolvedores de testes, mais editores de testes, mais aplicadores de testes, e propiciou o surgimento do que, logicamente, se tornou conhecido como um empreendimento de testagem. “Testagem” era o termo usado para referir-se a tudo, da

administração de um teste (como em “Testagem em andamento”) à interpretação de seu escore (“A testagem indicou que...”).

Durante a Primeira Guerra Mundial, o termo “testagem” descrevia a avaliação em grupo de milhares de recrutas militares. Suspeitamos que foi então que o termo ganhou uma posição poderosa no vocabulário dos profissionais e das pessoas leigas. O uso de “testagem” para denotar tudo desde a administração até a interpretação do teste pode ser encontrado em manuais do pós-guerra (tais como Chapman, 1921; Hull, 1922; Spearman, 1927), bem como em vários escritos relacionados a testes por décadas subsequentes. Entretanto, na Segunda Guerra Mundial uma distinção semântica entre testagem e um termo mais inclusivo, “avaliação”, começou a surgir.

Ao longo dessa guerra, o U. S. Office of Etrategic Services (OSS) dos Estados Unidos, um antecessor da atual Central Intelligence Agency (CIA), usou uma variedade de procedimentos e instrumentos de medida – os testes psicológicos entre eles – na seleção de pessoal militar para posições altamente especializadas envolvendo espionagem, coleta de informações, etc. Por exemplo, um dos instrumentos empregados era uma técnica de entrevista muito desconfortável, de uma pessoa por um grupo, para avaliar quão bem os candidatos poderiam responder a um interrogatório ao estilo da Gestapo. Com uma luz cruelmente apontada para seus rostos, os entrevistados tinham que apelar para sua própria criatividade, capacidade de persuasão e outros recursos para se defender e explicar de forma satisfatória um determinado cenário para um grupo de entrevistadores cada vez mais hostis. Os candidatos poderiam ter de explicar por que estavam em um determinado edifício que era fora dos seus limites e fazendo algo que não estavam autorizados a fazer, como, por exemplo, procurando ou removendo arquivos confidenciais. Eles eram avaliados em uma série de variáveis, tal como sua capacidade de manter respostas não contraditórias. Hoje, esse método de avaliação, ou qualquer método de avaliação que tenha o potencial de causar dano às pessoas que estão sendo avaliadas, provavelmente suscitaria sérias preocupações éticas.

Conforme resumido em *Assessment of Men* (OSS Assessment Staff, 1948) e em outras fontes (Murray e MacKinnon, 1946), os dados da avaliação gerados eram submetidos a integração e estimativa criteriosas por pessoal altamente treinado do centro de avaliação. O modelo do OSS – usar uma variedade inovadora de instrumentos de avaliação junto com dados das avaliações realizadas por indivíduos bem-treinados – inspiraria mais tarde o que agora é referido como abordagem do centro de avaliação à avaliação de pessoal (Bray, 1982).

Os ambientes militar, clínico, educacional e comercial são apenas alguns dos muitos contextos que implicam a observação comportamental e a integração ativa por avaliadores de pontuações de testes e outros dados. Nessas situações, o termo *avaliação* pode ser preferível a *testagem*. Em comparação com a testagem, a avaliação reconhece que os testes são apenas um tipo de instrumento usado por avaliadores profissionais (junto com outros instrumentos, como a entrevista) e que o valor de um teste, ou de qualquer outro instrumento de avaliação, está intimamente ligado ao conhecimento, à habilidade e à experiência do avaliador.

A distinção semântica entre testagem psicológica e avaliação psicológica é nebulosa na conversação diária. É um pouco surpreendente que a distinção entre os dois termos ainda permaneça nebulosa em alguns manuais de “testagem psicológica” publicados. Contudo, a distinção é importante. É mais positivo para a sociedade em geral que exista uma definição clara desses dois termos e uma diferenciação entre eles, bem como de termos relacionados, como *aplicador de teste psicológico* e *avaliador psicológico*. Distinções claras entre esses termos também podem ajudar a evitar as guerras territoriais que agora fermentam entre profissionais da psicologia e membros de outras profissões que procuram usar vários testes psicológicos. Em muitos contextos de análise psicológica, conduzir uma avaliação re-

REFLITA...

Descreva uma situação na qual a testagem seja mais apropriada do que a avaliação. Em contrapartida, descreva uma situação na qual a avaliação seja mais adequada do que a testagem.

quer maior educação, treinamento e habilidade do que simplesmente administrar um teste.

Definimos **avaliação psicológica** como a coleta e a integração de dados relacionados à psicologia com a finalidade de fazer uma estimativa psicológica, que é realizada por meio de instrumentos como testes, entrevistas, estudos de caso, observação comportamental e aparatos e procedimentos de medida especialmente projetados. Definimos **testagem psicológica** como o processo de medir variáveis relacionadas à psicologia por meio de instrumentos ou procedimentos projetados para obter uma amostra do comportamento. Algumas das diferenças entre esses dois processos são apresentadas na Tabela 1.1.¹

Tabela 1.1 Testagem comparada a avaliação

Em comparação com o processo de administrar, pontuar e interpretar testes psicológicos (testagem psicológica), a avaliação psicológica pode ser concebida como um processo de resolução de problema que pode assumir muitas formas diferentes. Como a avaliação psicológica se processa depende de muitos fatores, especialmente a própria razão da avaliação ser feita. Diferentes instrumentos de avaliação – entre eles os testes psicológicos – poderiam ser mobilizados no processo de avaliação, dependendo dos objetivos, das pessoas e das circunstâncias particulares envolvidos, bem como de outras variáveis únicas à situação em particular.

Reconhecidamente, a linha divisória entre o que constitui testagem e o que constitui avaliação nem sempre é clara como gostaríamos que fosse. Entretanto, reconhecendo que existe tal ambiguidade, podemos trabalhar para apurar nossa definição e o uso desses termos. Parece útil perceber as diferenças entre testagem e avaliação em termos do objetivo, do processo e do desfecho de uma avaliação e também em termos do papel e da habilidade do avaliador. Lembre que, embora essas sejam distinções úteis a considerar, exceções sempre podem ser encontradas.

| Testagem | Avaliação |
|--|---|
| <i>Objetivo</i> Normalmente, obter alguma medida, em geral de natureza numérica, com relação a uma capacidade ou um atributo. | Normalmente, responder a uma questão de encaminhamento, resolver um problema ou tomar uma decisão por meio do uso de instrumentos de avaliação. |
| <i>Processo</i> A testagem pode ser de natureza individual ou grupal. Após a administração do teste, o testador normalmente soma “o número de respostas corretas ou o número de certos tipos de respostas [...] com pouca ou nenhuma consideração pelo como ou a mecânica de tal conteúdo” (Maloney e Ward, 1976, p. 39). | A avaliação costuma ser individualizada. Em comparação com a testagem, a avaliação em geral se concentra mais em como um indivíduo processa do que simplesmente nos resultados desse processamento. |
| <i>Papel do avaliador</i> O testador não é fundamental para o processo; em termos práticos, um testador pode ser substituído por outro sem afetar consideravelmente o resultado. | O avaliador é fundamental ao processo de seleção dos testes e/ou de outros instrumentos de avaliação, bem como para tirar conclusões a partir de toda a avaliação. |
| <i>Habilidade do avaliador</i> A testagem normalmente requer habilidades técnicas em termos de administrar e pontuar um teste, bem como para interpretar o seu resultado. | A avaliação normalmente requer uma seleção especializada dos instrumentos de avaliação, habilidade na avaliação e organização e integração criteriosas dos dados. |
| <i>Resultados</i> Normalmente, a testagem produz um escore ou uma série de escores no teste. | Normalmente, a avaliação envolve uma abordagem de resolução de problemas que mobiliza muitas fontes de dados visando esclarecer o motivo do encaminhamento. |

¹ Especialmente quando se discutem os princípios gerais relativos à criação de procedimentos de medida, bem como a criação, manipulação ou interpretação de dados gerados a partir desses procedimentos, a palavra *teste* (e termos relacionados, como *escore de teste*) pode ser usada no sentido mais amplo e mais genérico; ou seja, “teste” pode ser usado na forma abreviada para aplicar-se a quase todos os procedimentos que implicam mensuração (incluindo, por exemplo, medidas de desempenho situacional). Por isso, quando falamos de “desenvolvimento de teste” no Capítulo 8, muitos dos princípios estabelecidos se aplicarão ao desenvolvimento de outras medidas que não são, estritamente falando, “testes” (tal como as medidas de desempenho situacional, e outros instrumentos de avaliação). Dito isso, enfatizamos que existe uma distinção real e importante entre os termos *testagem psicológica* e *avaliação psicológica* e que se deve fazer o possível para não confundir o significado de ambos.

O processo de avaliação. Em geral, o processo de avaliação começa com um encaminhamento para avaliação a partir de uma fonte como um professor, psicólogo da escola, conselheiro, juiz, médico ou especialista de recursos humanos corporativo. Normalmente, uma ou mais questões de encaminhamento são colocadas para o avaliador sobre o avaliando. Alguns exemplos de questões de encaminhamento são: “Esta criança pode funcionar em uma sala de aula regular?”, “Este réu é competente para ser julgado?” e “O quanto se pode esperar que este empregado se saia bem se for promovido para uma posição executiva?”.

O avaliador pode se encontrar com o avaliando ou com outras pessoas antes da avaliação formal a fim de esclarecer aspectos da razão para o encaminhamento. O avaliador prepara-se para a avaliação selecionando os instrumentos a serem usados. Por exemplo, se a avaliação ocorrer em um contexto corporativo ou militar e a questão de encaminhamento disser respeito à capacidade de liderança do avaliando, o avaliador pode desejar empregar uma medida (ou duas) de liderança. Normalmente, a própria experiência pregressa, a educação e o treinamento do avaliador desempenham um papel fundamental nos testes específicos ou em outros instrumentos a serem empregados na avaliação. Às vezes, uma instituição na qual a avaliação está acontecendo prescreveu diretrizes relativas a quais instrumentos podem e quais não podem ser usados. Na maioria das situações, sobretudo aquelas que são relativamente novas para o avaliador, o processo de seleção do instrumento pode

REFLITA...

Que qualidades fazem um bom líder? Como essas qualidades poderiam ser medidas?

ser inspirado por alguma pesquisa em preparação para a avaliação. Por exemplo, na avaliação de liderança, o procedimento de seleção do instrumento poderia ser inspirado por publicações que tratem de estudos comportamentais de liderança (Derue et al., 2011), estudos psicológicos de líderes (Kouzes e Posner, 2007), questões culturais na liderança (Byrne e Bradley, 2007) ou quaisquer aspectos da liderança em que a avaliação será focalizada (p. ex., Carnevale et al., 2011; Elliott, 2011; Kalshoven et al., 2011).

Subsequente à seleção dos instrumentos ou procedimentos a serem empregados, a avaliação formal começará. Após ser realizada, o avaliador escreve um relatório dos achados visando responder à questão de encaminhamento. Sessões adicionais para *feedback* com o avaliando e/ou terceiras pessoas interessadas (tais como os pais do avaliando e o profissional que o encaminhou) também podem ser marcadas.

Avaliadores diferentes podem abordar a tarefa da avaliação de formas distintas. Alguns a abordam com participação mínima dos próprios avaliandos. Outros consideram o processo de avaliação mais uma colaboração entre o avaliador e o avaliando. Por exemplo, em uma abordagem à avaliação, referida (logicamente) como **avaliação psicológica colaborativa**, o avaliador e o avaliando podem trabalhar como “parceiros” do contato inicial ao *feedback* final (Finello, 2011; Fischer, 1978, 2004, 2006). Uma variedade da avaliação colaborativa inclui um elemento de terapia como parte do processo. Stephen Finn e colaboradores (Finn, 2003, 2011; Finn e Martin, 1997; Finn e Tonsager, 2002) descreveram uma abordagem colaborativa à avaliação chamada de **avaliação psicológica terapêutica**. Aqui, a autodescoberta terapêutica e novos entendimentos são encorajados durante todo o processo de avaliação. Saiba mais sobre essa abordagem, e “diga olá” ao próprio dr. Stephen Finn, na seção *Conheça um profissional da avaliação* deste capítulo. Essa seção foi concebida para ajudar os leitores a adquirir um conhecimento em primeira mão com o uso do teste “na vida real”. Em cada capítulo deste livro, você será apresentado a um diferente aplicador de teste e a uma perspectiva única sobre o processo de avaliação. A todos os profissionais da avaliação apresentados foi solicitado que escrevessem um ensaio compartilhando seus próprios pensamentos e experiências. Um trecho breve de cada ensaio é apresentado em cada capítulo, e a versão completa de cada ensaio está disponível (em inglês) em www.mhhe.com/cohentesting8.

Outra abordagem à avaliação que parece ter ganho impulso nos últimos anos, sobretudo em contextos educacionais, é denominada *avaliação dinâmica* (Poehner e van Com-

CONHEÇA UM PROFISSIONAL DA AVALIAÇÃO

Conheça o dr. Stephen Finn

Em avaliação terapêutica, usamos uma variedade de instrumentos psicológicos, incluindo testes de funcionamento cognitivo (p.ex., WAIS-IV) e de autorrelato de personalidade e sintomatologia (p.ex., o Rorschach). Nós os selecionamos com base em nossa sessão inicial com um cliente. Nesse encontro, ajudamos os clientes a formularem “questões de avaliação” pessoais que desejam ver respondidas, tais como “Por que é tão difícil para mim fazer contato visual?” ou “Por que nunca fui capaz de ter um relacionamento íntimo?”. Então selecionamos testes que ajudarão a responder às perguntas dos clientes, bem como àquelas que nos são feitas por quaisquer profissionais responsáveis pelo encaminhamento. Por exemplo, com as perguntas há pouco mencionadas, poderíamos propor a um cliente o MMPI-2-RF e o Rorschach, porque nossa experiência é que a combinação de um autorrelato e de um teste de personalidade baseado no desempenho é útil para nos ajudar a entender esses tipos de problemas. Na primeira sessão, também coletamos informação básica abrangente sobre as preocupações refletidas nas perguntas do cliente. Por exemplo, perguntamos sobre quando é mais difícil ou mais fácil para ele fazer contato visual, quando esse problema começou e o que ele já tentou para tratá-lo. Também questionamos sobre tentativas anteriores de ter relacionamentos íntimos.

Acreditamos que, na melhor das hipóteses, os testes psicológicos servem como “amplificadores de empatia” – ajudando a “nos colocarmos no lugar dos clientes” e a entender quebra-cabeças, dilemas ou pontos emperrados em suas vidas que não foram capazes de tratar de outras maneiras. Administramos os testes de uma forma padronizada no início de nossas avaliações e achamos que as informações fornecem hipóteses muito úteis sobre por que os clientes têm os problemas que têm. Com frequência, por meio desses testes, somos capazes de ajudar as pessoas a entenderem comportamentos confusos, mesmo autodestrutivos ou perturbadores, que outros profissionais da saúde mental não foram capazes de entender ou melhorar. E usamos os testes conscientemente para identificar tanto os pontos fortes quanto as dificuldades das pessoas.

Envolvemos os clientes como colaboradores e “coexperimentadores” durante as sessões de testagem. Por exemplo, com o cliente supramen-



Stephen Finn, PhD, fundador do Center for Therapeutic Assessment, Austin, Texas.

cionado, poderíamos discutir itens reais do MMPI-2 que sugerissem sensação de inutilidade e vergonha. Ou poderíamos pedir-lhe para pensar conosco sobre as seguintes respostas do Rorschach: “Um morcego que está voando com asas terrivelmente danificadas – não sei como ele continua a voar” e “Um cão sarnento – do tipo que ninguém tiraria do canil e levaria para casa”. Poderíamos mesmo pedir-lhe para experimentar fazer, na sessão, mais contato visual com o avaliador e prestar atenção ao que ele sente. Todas essas interações poderiam levar a discussões sobre o sentimento de inadequação e vergonha do cliente, como esse sentimento surgiu, e como isso tudo está relacionado às suas questões de avaliação sobre fazer contato visual e ter relacionamentos íntimos.

No final da avaliação, mostramos e conversamos com o cliente a respeito das pontuações reais do teste e discutimos os “próximos passos” que ele poderia dar para tratar os problemas existenciais que foram o foco da avaliação. Frequentemente uma avaliação terapêutica é uma boa porta de entrada para o tratamento psicológico...

Leia mais sobre o que o dr. Finn tinha a dizer – seu ensaio completo (em inglês) – em www.mhhe.com/cohentesting8.

pernelle, 2011). O termo *dinâmica* pode sugerir que uma abordagem psicodinâmica ou psicanalítica à avaliação esteja sendo aplicada. Entretanto, esse não é o caso. No presente contexto, o termo *dinâmica* é usado para descrever a natureza interativa e variável da avaliação. Em geral, **avaliação dinâmica** refere-se a uma abordagem interativa à avaliação psicológica que geralmente segue um modelo de (1) avaliação, (2) intervenção de algum tipo e (3) avaliação. A avaliação dinâmica costuma ser mais empregada em contextos educacionais, embora possa ser utilizada também em ambientes correccionais, corporativos, neuropsicológicos, clínico e em muitos outros.

A intervenção entre as avaliações, às vezes mesmo entre as questões individuais colocadas ou as tarefas dadas, poderiam assumir diferentes formas, dependendo do propósito da avaliação dinâmica (Haywood e Lidz, 2007). Por exemplo, um avaliador pode intervir no curso de uma avaliação das capacidades de um avaliando com *feedback* ou sugestões cada vez mais explícitas. O propósito da intervenção pode ser o fornecimento de assistência com o domínio da tarefa em questão. O progresso no domínio da mesma tarefa ou de tarefas semelhantes é então medido. Basicamente, a avaliação dinâmica proporciona um meio para avaliar como o avaliando processa ou se beneficia de algum tipo de intervenção (*feedback*, sugestões, instrução, terapia e assim por diante) durante o curso da avaliação. Em alguns contextos educacionais, a avaliação dinâmica pode ser vista como uma forma de medir não apenas a aprendizagem, mas o “potencial de aprendizagem” ou as habilidades de “aprender a aprender”. Os computadores são um instrumento usado para ajudar na obtenção dos objetivos da avaliação dinâmica (Wang, 2011). Existem outros...

Os instrumentos de avaliação psicológica

O teste

Um **teste** pode ser definido simplesmente como um dispositivo ou procedimento de medida. Quando é precedida de um modificador, a palavra *teste* se refere a um dispositivo ou procedimento que visa medir uma variável relacionada àquele modificador. Considere, por exemplo, o termo *exame (teste) médico*, que faz referência a um dispositivo ou procedimento para medir alguma variável relacionada à prática da medicina (incluindo uma ampla variedade de instrumentos e procedimentos, como raios X, exames de sangue e teste dos reflexos). De forma semelhante, o termo **teste psicológico** refere-se a um dispositivo ou procedimento visando medir variáveis relacionadas à psicologia (p. ex., inteligência, personalidade, aptidão, interesses, atitudes ou valores). Embora um exame médico possa envolver a análise de uma amostra de sangue, tecido ou similar, um teste psicológico quase sempre envolve a análise de uma amostra de comportamento. A amostra de comportamento poderia variar de respostas a um questionário escrito a respostas orais a perguntas relacionadas ao desempenho de alguma tarefa. A amostra de comportamento poderia ser extraída pelo estímulo do próprio teste ou poderia ser um comportamento ocorrendo de modo natural (observado pelo avaliador, em tempo real enquanto ocorre, ou gravado).

Os testes psicológicos e outros instrumentos de avaliação podem diferir com respeito a uma série de variáveis, como conteúdo, formato, procedimentos de administração, procedimentos de pontuação e interpretação e qualidade técnica. O *conteúdo* (assunto) naturalmente irá variar de acordo com o foco do teste em particular. Mas, mesmo dois testes psicológicos que se proponham a medir a mesma coisa – por exemplo, a personalidade – podem diferir bastante no conteúdo dos itens. Isso acontece porque dois desenvolvedores do teste poderiam ter visões inteiramente diferentes em relação ao que é importante na medida da “personalidade”; diferentes desenvolvedores de testes empregam definições diferentes de “personalidade”. Além disso, diferentes desenvolvedores chegam ao processo de desenvolvimento com diferentes orientações teóricas. Por exemplo, os itens em

um teste de personalidade de orientação psicanalítica podem ter pouca semelhança com os de outro com orientação comportamental, contudo ambos são testes de personalidade. Um teste de personalidade de orientação psicanalítica poderia ser selecionado por um avaliador de orientação psicanalítica, e um teste de personalidade de orientação existencial poderia ser escolhido por um avaliador dessa orientação.

O termo **formato** diz respeito à forma, ao plano, à estrutura, ao arranjo e ao leiaute dos itens do teste e, ainda, às considerações relacionadas, como limites de tempo. *Formato* também é usado em referência à forma como um teste é administrado: computadorizado, escrito ou alguma outra forma. Quando se faz menção específica a um teste computadorizado, o formato também pode envolver a forma do programa: compatível com PC ou Mac. O termo *formato* não se restringe aos testes. *Formato* também é usado para denotar a forma ou a estrutura de outros instrumentos e processos avaliativos, tais como as diretrizes para criar um portfólio de trabalho.

Os testes diferem em seus *procedimentos de administração*. Alguns, particularmente aqueles destinados à administração em uma base de um-para-um, podem requerer um administrador de teste ativo e bem informado. A aplicação pode envolver a demonstração de vários tipos de tarefas exigidas do avaliado, bem como a observação treinada de seu desempenho. De forma alternativa, alguns testes, sobretudo aqueles designados para administração a grupos, podem nem mesmo requerer que o administrador esteja presente enquanto os testandos completam independentemente as tarefas requeridas.

Os testes diferem em seus *procedimentos de pontuação e interpretação*. Para entender melhor como e por que, vamos definir o que seja *pontuação* e *pontuar*. Os aficionados por esportes estão familiarizados com esses termos. Para eles, esses termos referem-se ao número de pontos acumulados pelos competidores e ao processo de acumulação desses pontos. Em testagem e avaliação, podemos definir **pontuação** de modo formal como um código ou uma súmula, em geral, mas não necessariamente de natureza numérica, que reflete uma avaliação do desempenho em um teste, uma tarefa, uma entrevista ou alguma outra amostra de comportamento. **Pontuar** é o processo de atribuir esses códigos avaliativos ou súmulas ao desempenho em testes, tarefas, entrevistas ou outras amostras de comportamento. No mundo da avaliação psicológica, existem muitos tipos diferentes de pontuação. Algumas resultam da simples soma das respostas (tal como a soma de respostas dos tipos correta/ incorreta ou concordo/ discordo) e algumas são derivadas de procedimentos mais elaborados.

As próprias pontuações podem ser descritas e categorizadas de muitas formas diferentes. Por exemplo, um tipo de pontuação é o *ponto de corte*. Um **ponto de corte** (também referido apenas como *corte*) é um ponto de referência, geralmente numérico, derivado por julgamento e usado para dividir um conjunto de dados em duas ou mais classificações. Alguma medida será tomada ou alguma inferência será feita com base nessas classificações. Os pontos de corte nos testes, via de regra em combinação com outros dados, são usados nas escolas em muitos contextos. Por exemplo, eles podem ser usados para dar notas e para tomar decisões sobre a classe ou o programa ao qual as crianças serão encaminhadas. Eles também são utilizados por empregadores como auxiliares na tomada de decisão sobre contratação, colocação e promoção de pessoal. Os órgãos estaduais usam pontos de corte como auxiliares em decisões de licenças. Há provavelmente mais de uma dúzia de métodos diferentes que podem ser empregados para derivar formalmente os pontos de corte (Dwyer, 1996). Se você está curioso sobre quais são alguns desses diferentes métodos, fique atento; abordamos o assunto em um capítulo próximo.

Às vezes, nenhum método formal é usado para chegar a um ponto de corte. Alguns professores utilizam o método informal do “olhômetro” para declarar, por exemplo, que uma pontuação de 65 ou mais em um teste significa “passar” e uma pontuação de 64 ou

REFLITA...

Imagine que quisesse desenvolver um teste para um traço de personalidade que você chamou de “gótico”. Como definiria esse traço? Que tipos de itens incluiria no teste? Por que você incluiria esses tipos de itens?

abaixo significa “rodar”. Sejam derivadas formal ou informalmente, os pontos de corte costumam levar em consideração, pelo menos em alguma medida, os valores daqueles que os estabelecem. Considere, por exemplo, dois professores que ensinam a mesma matéria na mesma faculdade. Um poderia estabelecer uma nota de corte para passar bem mais alta (e mais difícil para os estudantes alcançarem) do que o outro professor. Também há um outro lado na equação humana em relação aos pontos de corte, que é raramente mencionado nos textos sobre mensuração. Esse fenômeno diz respeito às consequências emocionais de “não alcançar a nota” e de “passar raspando” (ver Fig. 1.1).

Os testes apresentam ampla diferença em termos de suas diretrizes para pontuação e interpretação. Alguns são autopontuados pelos próprios testandos, outros são pontuados por computador e outros requerem contagem por examinadores treinados. Alguns, como a maioria dos testes de inteligência, vêm com manuais que são explícitos não apenas sobre os critérios de pontuação, mas também sobre a natureza das interpretações que podem ser feitas das pontuações. Outros, como o Teste da Mancha de Tinta de Rorschach, são vendidos sem qualquer manual. O comprador (presumivelmente qualificado) compra os materiais de estímulo e então seleciona e usa um dos muitos roteiros disponíveis para administração, pontuação e interpretação.

Os testes diferem com relação a sua **solidez psicométrica** ou qualidade técnica. **Psicométrie** pode ser definida como a ciência da mensuração psicológica. Variantes desse



Figura 1.1 Emoção produzida pelos pontos de corte (limiares) da variável categórica.

Pessoas que mal alcançam o ponto de corte (limiar) da variável categórica podem se sentir melhor acerca de suas realizações do que aquelas que alcançam o ponto de corte (limiar) por uma margem substancial. Mas as que quase alcançaram o ponto de corte (limiar) podem se sentir pior do que aqueles que o perderam por uma margem substancial. Uma evidência consistente com essa visão foi apresentada na pesquisa com atletas olímpicos (Medvec et al., 1995; Medvec e Savitsky, 1997). Os medalhistas de bronze ficaram – um pouco paradoxalmente – mais felizes com o resultado do que os medalhistas de prata. Os que receberam o bronze podiam dizer para si mesmos “Pelo menos eu ganhei uma medalha” e ficar felizes com isso. Em contrapartida, os que ganharam a prata poderiam se sentir frustrados porque tentaram a medalha de ouro e deixaram de ganhá-la.

termo incluem o adjetivo *psicométrico* (que se refere à mensuração de natureza psicológica) e o substantivo **psicometrista** (referindo-se a um profissional que usa, analisa e interpreta dados de testes psicológicos). Falamos de solidez psicométrica de um teste quando nos referimos ao grau de consistência e precisão com que um teste psicológico mede o que se propõe a medir. Os profissionais da avaliação também falam da *utilidade* psicométrica de um determinado teste ou método de avaliação. Neste contexto, **utilidade** diz respeito ao valor prático que um teste ou outro instrumento de avaliação tem para um determinado propósito. Esses conceitos são abordados com mais detalhes em capítulos subsequentes. Agora, voltando a nossa discussão dos instrumentos de avaliação, apresentamos um instrumento bem conhecido que, como dizem, “dispensa apresentação”.

REFLITA...

Como um teste de inteligência poderia ter mais utilidade do que outro teste de inteligência no mesmo contexto escolar?

A entrevista

Na conversação diária, a palavra *entrevista* invoca imagens de conversa face a face. Mas a entrevista como um instrumento de avaliação psicológica em geral envolve mais do que conversa. Se for conduzida face a face, então é provável que o entrevistador esteja tomando nota não apenas do conteúdo do que é dito, mas também da forma como está sendo dito. Mais especificamente, estará anotando comportamento verbal e não verbal. O comportamento não verbal pode incluir a “linguagem corporal”, os movimentos e a expressão facial do entrevistado em resposta ao entrevistador, o grau de contato visual, a aparente disposição em cooperar e a reação geral às demandas da entrevista. O entrevistador também pode anotar a forma como o entrevistado está vestido. Aqui, variáveis como arrumado *versus* desleixado e adequado *versus* inadequado podem ser anotadas.

Devido a uma possível abundância de informação não verbal a ser obtida, idealmente as entrevistas são conduzidas face a face. Entretanto, esse tipo de contato nem sempre é possível, e as entrevistas podem ser conduzidas em outros formatos. Em uma entrevista conduzida por telefone, por exemplo, o entrevistador ainda pode ser capaz de obter informações além das respostas às perguntas sendo sensível a variáveis como mudanças no tom de voz do entrevistado ou o quanto determinadas perguntas precipitam longas pausas ou sinais de emoção na resposta. Naturalmente, as entrevistas não precisam envolver fala verbalizada, como quando são conduzidas na linguagem de sinais. Elas também podem ser conduzidas por vários meios eletrônicos, como seria o caso de entrevistas pela internet, por *e-mail*, e as conduzidas por meio de mensagens de texto. Em seu sentido mais amplo, então, podemos definir uma **entrevista** como um método de obter informação por intermédio da comunicação direta envolvendo troca recíproca.

REFLITA...

Que tipo de situação de entrevista você consideraria ideal para ser realizada inteiramente por meio de mensagem de texto?

As entrevistas diferem com relação a muitas variáveis, tais como propósito, duração e natureza. Elas podem ser usadas pelos psicólogos em várias áreas de especialidade para ajudar na realização de diagnóstico, tratamento, seleção e em outras decisões. Assim, por exemplo, psicólogos de escola podem usar uma entrevista para tomar uma decisão sobre a adequação de várias intervenções educacionais ou arranjos de classes. Um psicólogo nomeado pelo tribunal pode usá-la para auxiliar a orientar o tribunal a determinar se um réu estava em seu juízo perfeito no momento em que cometeu um crime. Um especialista em traumatismo craniano pode utilizar uma entrevista para ajudar no esclarecimento de questões relacionadas à extensão do dano ao cérebro causado pelo ferimento. Um psicólogo estudando o comportamento dos consumidores pode usá-la a fim de conhecer o mercado para vários produtos e serviços, bem como a melhor maneira de anunciar e promovê-los.

Uma entrevista pode ser usada para ajudar os profissionais de recursos humanos a fazerem recomendações mais informadas sobre contratação, demissão e promoção de

peçoal. Em alguns casos, é empregado o que é chamado de **entrevista em painel** (também denominada *banca de entrevista*). Aqui, mais de um entrevistador participa na avaliação. Uma suposta vantagem dessa técnica de avaliação de peçoal é que quaisquer vieses idiossincrásicos de um entrevistador isolado serão minimizados (Dipboye, 1992). Uma desvantagem da entrevista em painel diz respeito a sua utilidade; o custo de usar múltiplos entrevistadores pode não ser justificado (Dixon et al., 2002).

A popularidade da entrevista como um método de obter informação vai muito além da psicologia. Tente pensar em um dia em que você não esteve exposto a uma entrevista na televisão, no rádio ou na internet! Independentemente do meio pelo qual ela é conduzida, uma entrevista é uma coisa recíproca na medida em que o entrevistado reage ao entrevistador e vice-versa. A qualidade, se não a quantidade, de informação útil produzi-

da por uma entrevista depende em grande parte das habilidades do entrevistador. Os entrevistadores diferem em muitos aspectos: seu ritmo de entrevistas, seu *rappor*t com os entrevistados e sua capacidade de transmitir autenticidade, empatia e humor. Levando firmemente em conta essas diferenças, considere a Figura 1.2. Como os atributos de personalidade característicos dessas duas celebridades poderiam afetar as respostas dos entrevistados? Quais desses dois entrevistadores você acha que entrevista melhor? Por quê?

REFLITA...

Que tipos de habilidades o apresentador de um programa de entrevista deve ter para ser considerado um entrevistador eficiente? Essas habilidades diferem daquelas necessárias a um profissional no campo da avaliação psicológica? Nesse caso, como?

O portfólio

Estudantes e profissionais em muitos campos de atividade diferentes, variando de arte a arquitetura, mantêm arquivos de suas produções. Essas produções – sejam conservadas em papel, tela, filme, vídeo, áudio ou algum outro meio – constituem o que é chamado de um **portfólio**. Como amostras da capacidade e das realizações da pessoa, um portfólio pode ser usado como um instrumento de avaliação. Empregadores de artistas comerciais, por exemplo, tomarão decisões de contratação baseados, em parte, na impressão passada pelo portfólio de amostras de desenhos de um candidato. Como outro exemplo, considere os empregadores de pessoas com talento para programas de rádio ao vivo. Eles também tomarão decisões de contratação em parte com base em seus julgamentos de amostras de áudio de trabalhos anteriores do candidato.

O apelo da avaliação do portfólio como instrumento de apreciação se estende a muitos outros campos, incluindo a educação. Alguns têm afirmado, por exemplo, que a melhor avaliação das habilidades de escrita de um estudante pode ser realizada não pela administração de um teste, mas pedindo-lhe que reúna uma seleção de amostras de composições. Também no campo da educação, a avaliação do portfólio tem sido empregada como um instrumento na contratação de professores. O portfólio de um professor pode consistir em vários documentos, como planos de aula, escritos pu-

blicados e auxílios visuais desenvolvidos expressamente para o ensino de certas matérias. Todos esses materiais podem ser de extrema utilidade para aqueles que devem tomar decisões de contratação.

REFLITA...

Se você tivesse que preparar um portfólio representando “quem você é” em termos de sua carreira educacional, seus passatempos e seus valores, o que incluiria?

Dados da história de caso

Dados da história de caso refere-se a gravações, transcrições e outros relatos na forma escrita, pictórica ou outra que preserve informações de arquivo, os relatos oficiais e informais e outros dados e itens relevantes a respeito de um avaliando. Os dados da história de caso incluem arquivos ou trechos de arquivos mantidos em instituições e órgãos como escolas, hospitais, empregadores, instituições religiosas e órgãos da justiça criminal. Ou-



Figura 1.2 Sobre entrevistar e ser entrevistado.

Diferentes entrevistadores têm diferentes estilos de entrevista. Como você caracterizaria o estilo de entrevista de Jay Leno comparado ao de Nancy Grace?

tros exemplos de dados da história de caso são cartas e correspondências escritas, fotos e álbuns de família, recortes de jornais e revistas e vídeos, filmes e gravações caseiros. Amostras de trabalho, arte, rabiscos e relatos e figuras relacionadas a interesses e passatempos são outros exemplos.

A reunião de dados da história de caso, bem como de dados relacionados, em uma narrativa ilustrativa é referida por termos como *estudo de caso* ou *história de caso*. Podemos definir formalmente um **estudo de caso** (ou **história de caso**) como um relatório ou uma narrativa ilustrativa a respeito de uma pessoa ou um evento que foram compilados com base nos dados da história de caso. Um estudo de caso poderia, por exemplo, esclarecer como a personalidade de um indivíduo e um conjunto particular de condições ambientais se combinaram para produzir um líder mundial de sucesso. Um estudo de caso de um indivíduo que tentou assassinar uma figura política importante poderia esclarecer que tipos de indivíduos e condições levariam a tentativas semelhantes no futuro. O trabalho sobre um fenômeno psicológico social denominado *pensamento grupal* (*groupthink*) contém material de história de caso rico sobre tomadas de decisão coletivas que nem sempre resultaram nas melhores decisões (Janis, 1972; ver Fig. 1.3).

REFLITA...

Quais são os prós e os contras de usar dados de histórias de caso como instrumento de avaliação?

Observação comportamental

Se você deseja saber como alguém se comporta em uma determinada situação, observe seu comportamento nessa situação. Essa sabedoria “caseira” está por trás de pelo menos uma abordagem à avaliação. A **observação comportamental**, como é empregada por profissionais da avaliação, pode ser definida como a monitoração das ações dos outros ou de si mesmo por meios visuais e eletrônicos ao mesmo tempo em que se registram as infor-

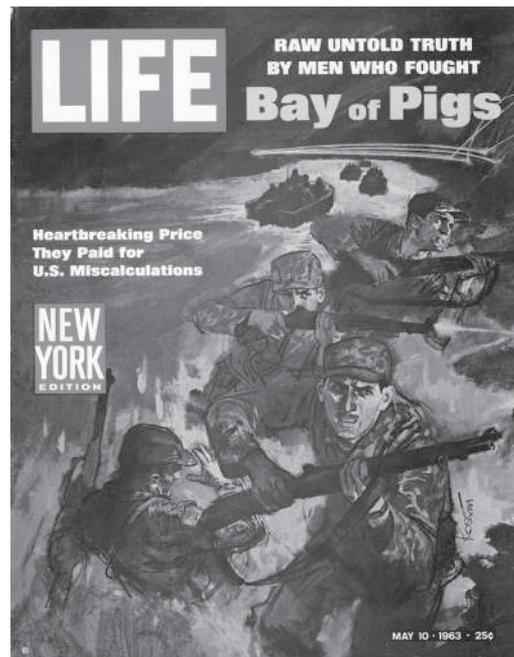


Figura 1.3 Estudo de caso do fenômeno de “pensamento grupal”.

De acordo com *Victims of Groupthink*, uma obra clássica do psicólogo social Irving Janis (1972), decisões tomadas por comitês ou grupos podem, em alguns casos, não ser tão sólidas quanto as tomadas por indivíduos. Esse fenômeno, conhecido como **pensamento grupal**, surge como resultado de forças variadas que podem levar os responsáveis pela decisão a chegar a um consenso (tal como a motivação para se comprometer e para minimizar visões conflitantes). Embora possam resultar em acordos, essas forças não resultam necessariamente no curso de ação ideal. Essa edição da revista *Life*, datada de 10 de maio de 1963, narrou uma invasão fracassada da Cuba de Fidel Castro na entrada sul daquela ilha, chamada de Baía dos Porcos. O “erro de cálculo” referido na capa da revista foi, de acordo com Janis, um estudo de caso de pensamento grupal.

mações quantitativas e/ou qualitativas relativas a essas ações. A observação comportamental é utilizada frequentemente como um auxílio diagnóstico em vários contextos, como unidades de pacientes internados, laboratórios de pesquisa comportamental e salas de aula. Ela pode ser usada para fins de seleção ou colocação em contextos corporativos ou organizacionais. Nesses casos, pode ser realizada como um auxílio para identificar as pessoas que melhor demonstrem as capacidades requeridas para uma determinada tarefa ou um trabalho. Às vezes, os pesquisadores se aventuram fora das fronteiras das clínicas, das salas de aula, dos locais de trabalho e dos laboratórios de pesquisa a fim de observar o comportamento de seres humanos em um ambiente natural – ou seja, o ambiente no qual normalmente se esperaria que o comportamento ocorresse. Essa variedade de observação comportamental é denominada **observação naturalista**. Assim, por exemplo, para estudar o comportamento socializante de crianças autistas com pares da mesma idade, uma equipe de pesquisa optou por ambientes naturais em vez de um ambiente laboratorial controlado (Bellini et al., 2007).

REFLITA...

Quais são os prós e os contras de usar a observação naturalista como instrumento de avaliação?

A observação comportamental como um auxílio para planejar a intervenção terapêutica provou ser extremamente útil em ambientes institucionais como escolas, hospitais, prisões e residências coletivas. Usando listas de comportamentos-alvo publicadas ou construídas por ela própria, a equipe pode observar em primeira mão o comportamento dos indivíduos e planejar intervenções de acordo. Em uma situação escolar, por exemplo, a observação natu-

ralista, no recreio, de uma criança de cultura diferente suspeita de ter problemas linguísticos poderia revelar que ela tem habilidades para falar a língua oficial, mas reluta – por razões de timidez, educação ou o que seja – em demonstrar essas habilidades para os adultos.

Na prática, a observação comportamental, e em especial a naturalista, tende a ser usada com mais frequência pelos pesquisadores em ambientes como salas de aula, clínicas, prisões e outros tipos de contextos nos quais os observadores tenham acesso fácil aos avaliandos. Para profissionais privados, em geral não é prático ou economicamente viável passar horas fora do consultório observando clientes enquanto eles seguem suas vidas diárias. Contudo, há alguns profissionais da saúde mental, como aqueles no campo da vida assistida, que encontram um grande valor na observação comportamental de pacientes fora de seu ambiente institucional. Para eles, pode ser necessário acompanhar um paciente fora dos muros da instituição para descobrir se ele é capaz de realizar atividades da vida diária de maneira independente. Nesse contexto, um instrumento de avaliação que se baseia muito na observação comportamental, como o Teste das Habilidades para Fazer Compras no Supermercado (Test of Grocery Shopping Skills) (ver Fig. 1.4), pode ser extremamente útil.

Testes de dramatização

Dramatização pode ser definida como a encenação de um papel improvisado ou parcialmente improvisado em uma situação simulada. Um **teste de dramatização** é um instru-



Figura 1.4 Verificação do preço (e julgamento) no corredor 5.

Criado primariamente para ser usado com pessoas com transtornos psiquiátricos, o Teste das Habilidades para Fazer Compras no Supermercado (Hamera e Brown, 2000) pode ser muito útil na avaliação de uma habilidade necessária para a vida independente.

mento de avaliação no qual os avaliandos são dirigidos a agir como se estivessem em uma determinada situação. Os indivíduos podem então ser avaliados com relação a seus pensamentos expressados, comportamentos, capacidades e outras variáveis.

REFLITA...

Quais são os prós e os contras da dramatização como um instrumento de avaliação? Em sua opinião, que tipo de problema apresentado seria ideal para a avaliação por dramatização?

A dramatização é útil para avaliar diversas habilidades. Assim, por exemplo, as habilidades de fazer compras no supermercado (Fig. 1.4) compreensivelmente poderiam ser avaliadas por meio da dramatização. Dependendo de como a tarefa é estabelecida, uma ida real ao supermercado poderia ou não ser necessária. Sem dúvida, a dramatização pode não ser tão útil quanto “a coisa real” em todas as situações. Contudo, é usada de forma bastante extensiva, sobretudo em condições nas quais seja muito demorado, muito

caro ou muito inconveniente avaliar em uma situação real. Assim, por exemplo, astronautas em treinamento podem precisar dramatizar muitas situações “como se” estivessem no espaço. Tais cenários “como se” para fins de treinamento resultam em uma economia verdadeiramente “astronômica”.

Indivíduos que estão sendo avaliados em um contexto corporativo, industrial, organizacional ou militar para capacidade gerencial ou de liderança podem ser colocados como rotina em situações de dramatização. Pode ser solicitado, por exemplo, que sejam mediadores em uma disputa hipotética entre os funcionários em um local de trabalho. O formato da dramatização pode variar de “cenários ao vivo”, com atores vivos, a simulações geradas no computador. As medidas de desfecho para esse tipo de avaliação poderiam incluir estimativas relacionadas a vários aspectos da capacidade do indivíduo de resolver o conflito, tal como eficácia da abordagem, qualidade da resolução e tempo até a resolução.

A dramatização como um instrumento de avaliação pode ser utilizada em vários contextos clínicos. Por exemplo, ela é empregada rotineiramente em muitas intervenções com viciados em drogas. Os clínicos podem tentar obter uma medida basal de abuso, fissura ou habilidades de enfrentamento administrando um teste de dramatização antes da intervenção terapêutica. O mesmo teste volta a ser administrado após o final do tratamento. A dramatização pode, portanto, ser usada como um instrumento de avaliação e também como uma medida de desfecho.

Os computadores como instrumentos

Já fizemos referência ao papel que os computadores desempenham na avaliação contemporânea no contexto de gerar simulações. Eles também podem ajudar na mensuração de variáveis que no passado eram muito difíceis de quantificar (ver Fig. 1.5). Mas talvez o papel mais óbvio como instrumento de avaliação seja na administração, na pontuação e na interpretação de teste.

Como administradores de testes, os computadores fazem mais do que substituir o “equipamento” que era tão amplamente utilizado no passado (um lápis número 2). Eles podem atuar como administradores de testes (*online* e *offline*) e como pontuadores de testes de alta eficiência. Em segundos, podem obter não apenas as pontuações do teste, mas padrões de pontuações. A pontuação pode ser feita no local (**processamento local**) ou conduzida em algum local central (**processamento central**). Se o processamento ocorrer em um local central, os dados relativos aos testes podem ser enviados e devolvidos dessa instalação central por meio de linhas telefônicas (**teleprocessamento**), correio ou mensageiro. Seja processado local ou centralmente, um relato do desempenho de um testando pode variar de uma mera listagem de uma pontuação ou pontuações (um **relatório de pontuação simples**) ao **relatório de pontuação estendido**, que inclui análises estatísticas do desempenho do testando. Um avanço dos relatórios de pontuação é o **relatório interpretativo**, que é diferenciado por nele estarem incluídas declarações interpretativas numéricas ou narrativas. Alguns relatórios interpretativos contêm relativamente pouca in-

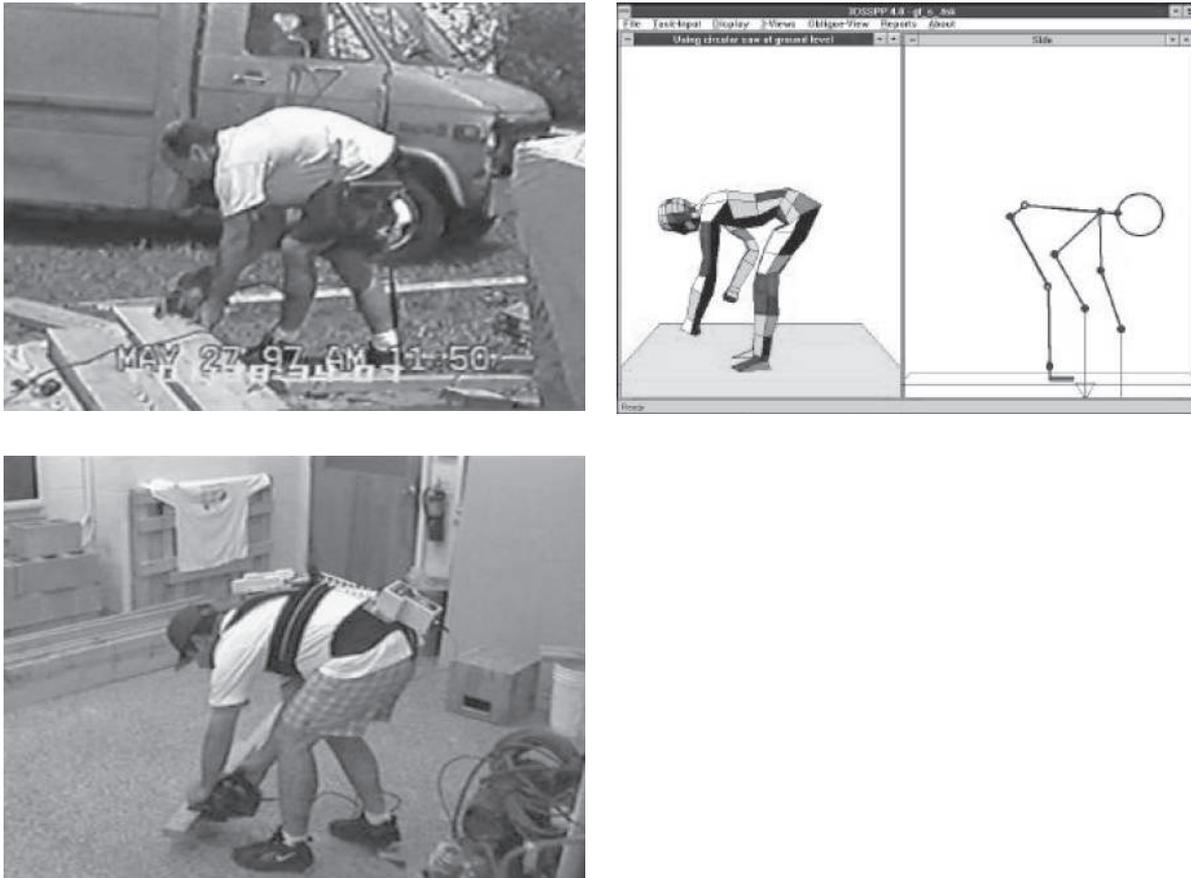


Figura 1.5 Um método de quantificar a tensão nas costas.

A aplicação inovadora da tecnologia dos computadores facilitou a mensuração de traços ou capacidades por técnicas que poderiam não ser medidas por métodos mais tradicionais. Uma dessas metodologias de avaliação combina as tecnologias de vídeo e computador para obter uma avaliação contínua da tensão nas costas (Mirka et al., 2000). Uma imagem (neste caso, de uma pessoa serrando) é capturada por uma câmera de vídeo e então convertida em uma representação mensurável por computador.

interpretação e apenas chamam atenção para certas pontuações altas, baixas ou incomuns que precisam ser focalizadas. No patamar superior desses relatórios está o que é às vezes referido como um **relatório consultivo**. Esse tipo de relatório, costuma ser escrito em linguagem apropriada para a comunicação entre profissionais da avaliação, pode fornecer uma opinião especializada relativa à análise dos dados. Um outro tipo de relatório de pontuação computadorizado visa integrar os dados de outras fontes, além do próprio teste, no relatório interpretativo. Esse **relatório integrativo** empregará dados coletados previamente (como registros de medicamentos ou dados de observação comportamental) no relatório do teste.

O acrônimo **CAPA** refere-se ao termo *avaliação psicológica assistida por computador* (*computer assisted psychological assessment*). A propósito, aqui a palavra *assistida* faz referência normalmente à assistência que o computador fornece ao aplicador do teste, não ao testando. Outro acrônimo com que você pode se deparar é **CAT**, este para *teste adaptado por computador* (*computer adaptive testing*). O *adaptado* nesse termo é uma alusão à capacidade do computador de adaptar o teste à capacidade do testando ou ao padrão de testagem. Assim, por exemplo, em um teste computadorizado de habi-

REFLITA...

Descreva um teste que seria ideal para ser administrado por computador. Então descreva outro que não seria ideal para ser administrado por computador.

lidades acadêmicas, o computador poderia ser programado para mudar de matemática para habilidades de português após três erros consecutivos nos itens de matemática.

A CAPA abriu um mundo de possibilidades para os desenvolvedores de testes, permitindo-lhes criar testes psicometricamente sólidos usando procedimentos e cálculos matemáticos tão complicados que levariam semanas ou meses para serem usados em uma época passada. Ela abriu um novo mundo também para os aplicadores de testes, permitindo a construção de testes sob medida com capacidades de pontuação e interpretação incorporadas antes impensáveis. Para muitos aplicadores, a CAPA foi um grande avanço em relação ao passado, quando tinham eles próprios de administrar os testes e talvez até colocar as respostas em alguma outra forma anterior à análise (p. ex., usando manualmente um gabarito de pontuação ou outro dispositivo). E mesmo após fazer tudo isso, começavam então as tarefas trabalhosas de pontuar e interpretar os dados resultantes. Entretanto, toda rosa tem seus espinhos; alguns dos prós e contras da CAPA são resumidos na Tabela 1.2.

Nos últimos anos, a perspectiva da aplicadores via internet parece ter ganhado a simpatia de clínicos, pesquisadores e outros aplicadores de testes (Bartram e Hambleton, 2006; Bersoff et al., 2012; Terzis e Economides, 2011). Estudos publicados empregaram a internet como um meio para a avaliação do risco para doença de Alzheimer (Brandt e

Tabela 1.2 CAPA: alguns prós e contras

| Prós | Contras |
|--|---|
| A CAPA economiza tempo do profissional na administração, pontuação e interpretação do teste. | Os profissionais ainda devem gastar tempo considerável lendo a documentação do programa e do equipamento e até livros auxiliares sobre o teste e sua interpretação. |
| A CAPA resulta em erros de pontuação mínimos resultantes de erro humano ou lapsos de atenção ou julgamento. | Com a CAPA, a possibilidade de erro de programa ou equipamento, de fontes difíceis de localizar, está sempre presente, como falhas do programa ou defeito de funcionamento do equipamento. |
| A CAPA garante a aplicação padronizada do teste a todos os testandos, com pouca ou nenhuma variação nos procedimentos de administração. | A CAPA deixa em desvantagem aqueles testandos que são incapazes de empregar estratégias familiares de realização de testes (pré-visualizar o teste, saltar perguntas, voltar à pergunta anterior, etc.). |
| A CAPA produz interpretação padronizada dos achados devido à eliminação de insegurança rastreável a diferentes pontos de vista no julgamento profissional. | A interpretação padronizada de achados da CAPA baseada em uma perspectiva unitária, estabelecida, pode não ser a ideal; a interpretação poderia aproveitar pontos de vista alternativos. |
| A capacidade dos computadores de combinar dados de acordo com regras é mais precisa do que a dos seres humanos. | Os computadores não têm a flexibilidade dos seres humanos para reconhecer a exceção a uma regra no contexto do "quadro completo". |
| Assistentes não profissionais podem ser usados no processo de administração do teste, e este normalmente pode ser administrado a grupos de testandos em um contexto. | O uso de não profissionais deixa pouca, ou nenhuma, oportunidade para o profissional observar o comportamento do avaliando durante o teste e notar quaisquer condições extrateste incomuns que possam ter afetado as respostas. |
| Grupos de profissionais, como a APA, desenvolvem diretrizes e padrões para uso dos produtos CAPA. | Não profissionais movidos pelo lucro também podem criar e distribuir testes com pouca consideração pelas diretrizes e padrões profissionais. |
| Os testes escritos podem ser convertidos para produtos CAPA com conseqüentes vantagens, tais como um tempo mais curto entre a administração e sua pontuação e interpretação. | O uso de testes escritos que foram convertidos para a administração por computador levanta dúvidas sobre a equivalência do teste original e sua forma convertida. |
| A segurança dos produtos CAPA pode ser mantida não apenas por meios tradicionais (tal como cabines de preenchimento trancadas) mas por produtos eletrônicos de alta tecnologia (tal como <i>firewalls</i>). | A segurança dos produtos CAPA pode ser violada por <i>hackers</i> , e a integridade dos dados pode ser alterada ou destruída por eventos inconvenientes, como a introdução de vírus de computador. |
| Os computadores podem adaptar automaticamente o conteúdo e a duração do teste com base nas respostas dos testandos. | Nem todos os testandos fazem o mesmo teste ou têm a mesma experiência na realização de testes. |

Rogerson, 2011), abuso de álcool e drogas (Sinadinovic et al., 2011) e, mesmo, adicção de internet (Young, 2011). Metodologias *online* foram usadas para avaliar interesses (Tracey, 2010), habilidades de resolução de problemas (Mayotte, 2010) e traços de personalidade (Migliore, 2011), entre outros atributos e condições.

A Comissão da APA sobre Testes e Avaliações Psicológicas foi convocada para considerar os prós e os contras da avaliação assistida por computador e avaliação usando a internet (Naglieri et al., 2004). Entre as vantagens citadas sobre os testes escritos estavam (1) os administradores do teste têm maior acesso a potenciais aplicadores devido ao alcance global da internet, (2) a pontuação e a interpretação dos dados do teste tendem a ser mais rápidas do que para os escritos, (3) os custos associados com a testagem via internet tendem a ser mais baixos do que os associados com os testes escritos e (4) a internet facilita a testagem de populações de outro modo isoladas, bem como de pessoas com deficiências para as quais chegar a um centro de testes poderia se revelar uma dificuldade. Poderíamos acrescentar que a testagem pela internet tende a ser “mais ecológica”, no sentido de que pode poupar papel, material de transporte, e assim por diante. Além disso, há provavelmente menos chance de erros de pontuação com os testes baseados na internet, comparados aos escritos.

Embora pareça ter muitas vantagens, a testagem via internet não está livre de possíveis armadilhas, problemas e questões. Uma questão básica tem a ver com o que Naglieri e colaboradores (2008) denominaram “integridade teste-cliente”. Em parte, isso se refere à verificação da identidade do testando quando um teste é administrado dessa maneira. Também diz respeito, em termos mais gerais, aos interesses às vezes variáveis do testando *versus* os do administrador do teste. Dependendo das condições da administração, os testandos podem ter acesso irrestrito a anotações, a outras fontes *online*, e a outros auxílios na realização do teste – isso apesar das diretrizes para a administração. Pelo menos com relação aos testes de desempenho, há alguma evidência de que a testagem via internet não supervisionada leva a “inflação da pontuação”, comparado com os testes aplicados de forma mais tradicional (Carstairs e Myers, 2009).

Um aspecto relativo à integridade teste-cliente tem a ver com o procedimento adequado para garantir que a segurança do teste administrado via internet não seja comprometida. O que impedirá que outros testandos visualizem cópias passadas – ou mesmo antecipadas – do teste?

Naglieri e colaboradores (2008) lembraram seus leitores da diferença entre testagem e avaliação e da importância de reconhecer que a testagem pela internet é apenas isso – *testagem*, não *avaliação*. Como tal, os aplicadores desses testes devem estar cientes de todas as possíveis limitações da fonte de pontuação do teste.

Outros instrumentos

A próxima vez que você tiver a oportunidade de assistir a um DVD, pare um pouco para considerar o papel que o vídeo pode desempenhar na avaliação. De fato, vídeos especialmente criados têm ampla utilização em contextos de treinamento e avaliação. Por exemplo, os funcionários de uma empresa podem ser solicitados a responder a uma variedade de incidentes de assédio sexual no local de trabalho apresentados em vídeo. Policiais podem ser questionados sobre como responderiam a vários tipos de emergências, que são apresentadas como encenações ou como gravações em vídeo de ocorrências reais. Os psicoterapeutas podem ser solicitados a responder com um diagnóstico e um plano de tratamento para cada um de diversos clientes apresentados a eles em videotape. A lista das possíveis aplicações do vídeo para a avaliação é infinita.

Além do vídeo, muitos outros itens comuns que você pode não associar prontamente com avaliação psicológica podem ser mobilizados apenas para esse propósito. Por exemplo, os psicólogos podem usar muitos dos instrumentos que costumam ser associa-

REFLITA...

Que cuidados os aplicadores de teste pela internet deveriam ter em mente em relação à fonte de dados de seus testes?

dos com a saúde médica, como termômetros para medir a temperatura corporal e manômetros para medir a pressão arterial. O equipamento de *biofeedback* é usado, às vezes, para obter medidas de reações corporais (tais como tensão muscular) a vários tipos de estímulos. Também há alguns instrumentos menos comuns, como o pletismógrafo peniano. Esse instrumento, concebido para medir a excitação sexual masculina, pode ser útil no diagnóstico e tratamento de predadores sexuais. A capacidade prejudicada de identificar odores é comum em muitos transtornos nos quais haja envolvimento do sistema nervoso central, e testes de olfato simples podem ser administrados para ajudar a determinar se

REFLITA...

Quando a avaliação por vídeo é uma abordagem melhor do que usar um teste escrito? Quais são as armadilhas, se houver, de usar o vídeo na avaliação?

tal prejuízo está presente. Em geral, não houve falta de inovação por parte dos psicólogos para criar instrumentos de medida ou adaptar instrumentos existentes, para uso na avaliação psicológica.

Até o momento, nossa introdução se concentrou em algumas definições básicas, bem como no exame de alguns dos “instrumentos envolvidos (na avaliação)”. Levantamos agora algumas questões fundamentais relativas a quem, o que, por que, como e onde da testagem e da avaliação.

Quem, o que, o porquê, o como e o onde?

Quem são as partes no empreendimento da avaliação? Em que tipos de contextos as avaliações são conduzidas? Por que a avaliação é conduzida? Como as avaliações são conduzidas? Onde buscar informações oficiais sobre os testes? Pense sobre a resposta para cada uma dessas perguntas importantes antes de continuar a leitura. Então compare suas próprias ideias com aquelas que se seguem.

Quem são as partes?

As partes no empreendimento da avaliação incluem os desenvolvedores e os editores dos testes, seus aplicadores e as pessoas que são avaliadas por meio deles. Além disso, podemos considerar a sociedade como um todo como uma parte no empreendimento da avaliação.

O desenvolvedor do teste. Os desenvolvedores e editores criam testes ou outros métodos de avaliação. A American Psychological Association (APA) estimou que mais de 20 mil novos testes psicológicos são desenvolvidos a cada ano. Entre eles estão alguns que foram criados para um estudo de pesquisa específico, alguns que foram criados na esperança de que seriam publicados e alguns que representam refinamentos ou modificações de testes existentes. Os criadores de testes trazem uma ampla série de conhecimentos e interesses para o processo de desenvolvimento de testes.²

Os desenvolvedores e editores de testes avaliam o impacto significativo que os resultados do teste podem ter sobre as vidas das pessoas. Conseqüentemente, inúmeras organizações profissionais publicaram padrões de comportamento ético que tratam de forma específica de aspectos do desenvolvimento e do uso responsável dos testes. Talvez o documento mais detalhado a tratar dessas questões seja o escrito em conjunto pela American Educational Research Association, a

NO BRASIL

Para informações sobre testes disponíveis, consultar o Sistema de Avaliação de Testes Psicológicos (SATEPSI) do Conselho Federal de Psicologia (CFP): <http://satepsi.cfp.org.br>.

² Para uma perspectiva fascinante na informação biográfica sobre uma amostragem de desenvolvedores de testes, navegue até a seção “Test Developers Profiles” (Perfil dos Desenvolvedores de Testes) de nossa página: www.mhhe.com/cohentesting8 (em inglês).

American Psychological Association e o National Council on Measurement in Education (NCME). Referido por muitos psicólogos simplesmente como “os Standards”, o livro *Standards for Education and Psychological Testing* abrange questões relativas à construção e à avaliação do teste, sua administração e uso, e suas aplicações especiais, tais como considerações especiais quando testando minorias linguísticas. Publicado inicialmente em 1954, revisões dos *Padrões* foram publicadas em 1966, 1974, 1985 e 1999. Esse livro é um trabalho de referência indispensável não apenas para os desenvolvedores de testes, mas também para seus aplicadores.

O aplicador do teste. Testes psicológicos e metodologias de avaliação são usados por uma ampla variedade de profissionais, incluindo clínicos, conselheiros, psicólogos escolares, pessoal de recursos humanos, psicólogos do consumidor, psicólogos experimentais e psicólogos sociais. Entretanto, a questão de quem está qualificado para aplicar os testes psicológicos tem gerado muita discussão.

Os *Standards* e outras diretrizes publicadas de organizações de profissionais especializados tiveram muito a dizer em termos de identificar quem é um aplicador de teste qualificado e quem deve ter acesso a (e permissão para comprar) testes psicológicos e instrumentos de avaliação psicológica relacionados. Contudo, existe controvérsia sobre quais profissionais com que tipo de treinamento devem ter acesso a quais testes. Membros de várias profissões, com pouco ou nenhum treinamento psicológico, têm buscado o direito de obter e usar testes psicológicos. Em muitos países, não existe uma regulamentação ética ou legal de uso de testes psicológicos (Leach e Oakland, 2007).

Portanto, quem são (ou devem ser) os aplicadores de testes? Terapeutas ocupacionais, por exemplo, devem ter permissão para administrar testes psicológicos? E quanto a empregadores e executivos de recursos humanos sem treinamento formal em psicologia? E quanto à questão de terceiras pessoas na sala quando a avaliação está acontecendo? Essa última questão é o assunto da discussão na seção *Em foco* deste capítulo.

Até agora, listamos uma série de questões controversas a respeito de *Quem?* que profissionais da avaliação especializados ainda discutem. Felizmente, há pelo menos uma questão a respeito de *Quem?* sobre a qual há bem pouca discussão: a relativa a quem é o testando ou avaliando.

O testando. Todos nós já fomos testados. Entretanto, nem todos abordamos os testes da mesma forma. No dia em que um teste deve ser administrado, os testandos podem diferir com respeito a inúmeras variáveis, incluindo estas:

- O grau de ansiedade que estão vivenciando e o quanto essa ansiedade poderia afetar significativamente os resultados de seu teste.
- O quanto entendem e concordam com a razão para a avaliação.
- Sua capacidade e disposição para cooperar com o examinador ou compreender as instruções escritas do teste.
- O grau de dor física ou sofrimento emocional que estão experimentando.
- O grau de desconforto físico provocado por não ter comido o suficiente, por ter comido demais ou por outras condições físicas.
- O quanto estão alertas e bem acordados, em oposição a desligados.
- O quanto estão predispostos a concordar ou discordar quando frente a afirmações de estímulo.
- A extensão do treinamento prévio que receberam.
- A importância que podem atribuir ao fato de se descreverem de forma positiva (ou negativa).

NO BRASIL

Apenas psicólogos registrados nos Conselhos Regionais de Psicologia (CRPs) podem adquirir e usar testes psicológicos.

REFLITA...

Além dos psicólogos, quem deve ter acesso permitido a, bem como o privilégio de usar, testes psicológicos?

EM FOCO

Os observadores devem ser parte do processo de avaliação?

O avaliador e o avaliado são duas partes em qualquer avaliação. A terceira parte pode ser um observador que esteja lá por inúmeras razões. A terceira pessoa pode ser um supervisor do avaliador, um amigo ou parente do avaliado, um representante da instituição na qual a avaliação está sendo conduzida, um tradutor, um advogado ou outra pessoa. Mas terceiras pessoas têm um lugar legítimo na sala durante uma avaliação por alguma razão? De acordo com Robert J. McCaffrey (2007), a resposta a essa pergunta é clara e inequívoca: “Não deve ser permitido que terceiros estejam presentes durante uma avaliação”.

McCaffrey e outros citam a pesquisa para apoiar a visão de que um processo de influência social ocorre por meio da simples presença de uma terceira pessoa (Gavett e McCaffrey, 2007; Vanderhoff et al., 2011; Yantz e McCaffrey, 2005, 2009). Essa influência social pode ser suficiente para afetar o desempenho do avaliado, de modo particular em tarefas envolvendo memória, atenção e outras funções cognitivas (Gavett et al., 2005). A terceira pessoa nem mesmo precisa estar fisicamente presente para afetar o desempenho de um avaliado – a observação por meio de uma câmera de vídeo, um espelho unidirecional ou um dispositivo semelhante também podem afetar seu desempenho (Constantinou et al., 2005).

O efeito de influência social que ocorre foi referido na literatura sobre testagem e avaliação como **facilitação social**, provavelmente porque a presença de terceiros era a princípio associada com aumentos no desempenho (Aiello e Douthitt, 2001). Entretanto, à luz de outras pesquisas sugerindo que uma plateia também possa ter o efeito de inibir o desempenho, um termo mais inclusivo – tal como *facilitação e inibição social* – poderia ser o mais correto (McCaffrey et al., 2005).

Os proponentes do acesso de terceiros à avaliação psicológica argumentam que ela é necessária para propósitos como treinamento clínico; não há substituto para a presença de um supervisor ali, na sala, para corrigir quaisquer erros de administração do teste que um avaliador em treinamento pudesse cometer durante o curso de uma avaliação. Outros argumentos a favor do acesso de terceiros podem citar a necessidade de tradutores ou de um advogado para garantir que os direitos de um avaliado sejam respeitados. Alguns estatutos estaduais preveem especificamente a



presença de terceiros observadores sob certas condições, embora a maioria dos estados ainda não tenha tratado essa questão por meio de legislação ou ação judicial (Duff e Fisher, 2005). Uma sondagem de uma amostra pequena ($n = 27$) de peritos forenses concluiu que a maioria ($n = 14$) era a favor de permitir a observação por terceiros; os restantes foram contra ou não tinham uma opinião formada sobre o assunto (Witt, 2003).

Os defensores da aplicação rigorosa de uma política que proíba a observação por terceiros durante a avaliação psicológica argumentam que alternativas a esse tipo de observação existem ou devem ser desenvolvidas. Por exemplo, em vez de permitir supervisores na sala durante uma avaliação, devem ser instituídos procedimentos de treinamento melhores – incluindo maior confiança na dramatização antes das administrações reais dos testes. McCaffrey (2005) alertou que certos dados de avaliação obtidos na presença de uma terceira pessoa podem ser considerados duvidosos em um tribunal e, portanto, são inadmissíveis. Ele alertou também que, sempre que houver uma terceira parte a uma avaliação, esse fato deve ser anotado claramente no respectivo relatório junto às possíveis consequências da presença da terceira parte.

- O quanto são, na falta de um termo melhor, “sortudos” e podem “superar as expectativas” em um teste de desempenho (ainda que possam não ter aprendido a matéria).

No sentido amplo no qual estamos usando o termo “testando”, qualquer um que seja objeto de uma avaliação ou apreciação pode ser um testando ou um avaliando. Por mais incrível que isso soe, significa que mesmo um indivíduo falecido pode ser considerado um avaliando. É verdade, esta é a exceção à regra, mas existe uma coisa chamada *autópsia psicológica*.

Uma **autópsia psicológica** pode ser definida como uma reconstrução do perfil psicológico de um indivíduo falecido com base em registros de arquivo, artefatos e entrevistas conduzidas previamente com o avaliando falecido ou com pessoas que o conheceram. Por exemplo, usando autópsias psicológicas, Townsend (2007) explorou a questão sobre se terroristas suicidas eram na verdade suicidas de um ponto de vista psicológico clássico. Ela concluiu que não eram. Outros pesquisadores forneceram avaliações psicológicas pós-morte fascinantes de pessoas de várias esferas da vida em muitas culturas diferentes (Bhatia et al., 2006; Chan et al., 2007; Dattilio, 2006; Fortune et al., 2007; Foster, 2011; Giner et al., 2007; Goldstein et al., 2008; Heller et al., 2007; McGirr et al., 2007; Owens et al., 2008; Palacio et al., 2007; Phillips et al., 2007; Pouliot e DeLeo, 2006; Sanchez, 2006; Thoresen et al., 2006; Vento et al., 2011; Zonda, 2006).

REFLITA...

Sobre que figura pública recentemente falecida você gostaria de ver uma autópsia psicológica? Por quê? Que resultados você poderia esperar?

A sociedade em geral

A singularidade dos indivíduos é um dos fatos característicos da vida mais fundamentais [...] Em todos os períodos da história da humanidade, os homens observaram e descreveram diferenças entre os indivíduos [...] Porém, os educadores, os políticos e os administradores sentiram, de alguma forma, a necessidade de organizar ou sistematizar a complexidade multifacetada das diferenças individuais. (Tyler, 1965, p. 3)

A necessidade social por “organizar” e “sistematizar” tem se manifestado historicamente em questões variadas como “Quem é feiticeira?”, “Quem é esquizofrênico?” e “Quem é qualificado?”. As perguntas específicas feitas mudaram com as preocupações sociais. Os métodos usados para determinar as respostas têm variado ao longo da história em razão de fatores como sofisticação intelectual e preocupação religiosa. Os proponentes da quiromancia, podoscopia, astrologia e frenologia, entre outras buscas, afirmaram que a melhor maneira de entender e prever o comportamento humano era pelo estudo das palmas das mãos, das plantas dos pés, das estrelas, de protuberâncias na cabeça, de folhas de chá, e assim por diante. Ao contrário dessas buscas, o empreendimento da avaliação tem raízes na ciência. Por meio de métodos sistemáticos e reproduzíveis que podem fornecer evidências convincentes, o empreendimento da avaliação responde ao que Tyler (1965, p. 3) descreve como demanda da sociedade por “alguma forma de organizar ou sistematizar a complexidade multifacetada das diferenças individuais”.

A sociedade em seu conjunto exerce de muitas formas sua influência como uma parte do empreendimento da avaliação. À medida que a sociedade evolui e surge a necessidade de medir diferentes variáveis psicológicas, os desenvolvedores respondem criando novos testes. Por intermédio de representantes eleitos para a legislatura, são promulgadas leis que governam os aspectos de desenvolvimento, administração e interpretação de testes. De modo similar, mediante decisões judiciais, a sociedade como um todo exerce sua influência sobre vários aspectos do empreendimento de testagem e avaliação.

Outras partes Além das quatro partes principais que focalizamos aqui, vamos considerar brevemente outras que podem participar de diversas formas no empreendimento de testagem e avaliação. Organizações, companhias e órgãos governamentais patrocinam o desenvolvimento de testes por várias razões, tais como certificar pessoas. As companhias

e os departamentos oferecem serviços de pontuação ou interpretação de testes. Em alguns casos, ambos são apenas extensões de editores de testes; em outros, podem ser independentes. Há pessoas cujas únicas responsabilidades são divulgação e venda de testes. Às vezes, essas pessoas são empregadas pelo editor do teste; às vezes, não. Há acadêmicos que revisam testes e avaliam sua solidez psicométrica. Todas essas pessoas, bem como muitas outras, são partes em maior ou menor grau do empreendimento da avaliação.

Tendo lhes apresentado a algumas das partes envolvidas no *Quem?* da testagem e da avaliação psicológica, vamos seguir em frente para tratar de algumas das questões de *O quê?* e *Por quê?*.

Em que tipos de contextos as avaliações são conduzidas, e por quê?

Contextos educacionais. Você provavelmente está familiarizado com os muitos tipos de testes aplicados na sala de aula. Como exigido por lei nos Estados Unidos, eles são administrados no início da vida escolar para ajudar a identificar crianças que possam ter necessidades especiais. Além dos testes de capacidade escolar, outro tipo em geral administrado é o **teste de realização**, que avalia o desempenho ou o grau de aprendizagem que ocorreu. Alguns dos testes de realização que você fez na escola foram idealizados por seu professor. Outros foram idealizados para uso mais difundido por educadores que trabalham com profissionais de mensuração. Nessa última categoria, acrônimos como SAT e GRE podem não ser estranhos.

Você sabe por experiência própria que um **diagnóstico** pode ser definido como uma descrição ou conclusão alcançadas com base em evidência ou opinião. Normalmente essa conclusão é alcançada por meio de um processo de diferenciação da natureza de alguma coisa e exclusão de conclusões alternativas. Da mesma forma, o termo **teste diagnóstico** se refere a um instrumento de avaliação usado para ajudar a restringir e identificar áreas de déficit a serem visadas para intervenção. Em contextos educacionais, testes diagnósticos de leitura, de matemática e de outras matérias acadêmicas podem ser administrados para avaliar a necessidade de intervenção educacional, bem como para estabelecer ou excluir a elegibilidade para programas de educação especial.

Os alunos recebem notas em seus boletins que não têm base em qualquer avaliação formal. Por exemplo, a nota no item “Trabalha e se dá bem com os outros” provavelmente seja baseada mais na *avaliação informal* da professora na sala de aula do que em pontuações em alguma medida de interação social publicada. Podemos definir **avaliação informal** como uma avaliação em geral não sistemática que leva à formação de uma opinião ou de uma atitude.

A avaliação informal não se limita, naturalmente, aos contextos educacionais; ela faz parte da vida diária. De fato, muitos dos instrumentos de avaliação que discutimos no contexto do ambiente educacional (tais como testes de realização, testes diagnósticos e avaliações informais) também são administrados em vários outros contextos. E alguns dos tipos de testes que discutimos no contexto dos ambientes descritos a seguir também são administrados em ambientes educacionais. Portanto, não esqueça de que os instrumentos de avaliação e as técnicas de mensuração que discutimos em um contexto podem muito bem ser usados em outros. Nosso objetivo nesse estágio inicial em nosso levantamento de campo é simplesmente apresentar uma amostra (não uma lista abrangente) dos tipos de testes usados em diferentes situações.

Ambientes clínicos Testes e muitos outros instrumentos de avaliação são amplamente usados em ambientes clínicos, como hospitais públicos, privados e militares, clínicas de internação e ambulatoriais, consultórios privados, escolas e outras instituições. Esses instrumentos são usados para ajudar a selecionar ou diagnosticar problemas de comportamento. Que tipos de situações poderiam induzir o emprego desses instrumentos? Aqui está uma pequena amostra.

REFLITA...

Que instrumentos de avaliação poderiam ser usados para abordar a capacidade de leitura de um estudante?

- Um cliente de psicoterapia particular deseja ser avaliado para ver se a avaliação pode fornecer alguns indícios não óbvios em relação a seu desajustamento.
- Uma psicóloga escolar avalia clinicamente uma criança que está experimentando dificuldades de aprendizagem para determinar que fatores são os principais responsáveis por elas.
- Um pesquisador de psicoterapia usa procedimentos de avaliação para definir se um determinado método de psicoterapia é eficaz para tratar um problema específico.
- Um psicólogo-consultor contratado por uma companhia de seguros é chamado para dar uma opinião quanto à realidade dos problemas psicológicos de um cliente; o cliente está realmente vivenciando tais problemas ou está apenas simulando?
- Um psicólogo indicado pelo tribunal é chamado para dar uma opinião acerca do grau de reabilitação de um prisioneiro violento condenado.

Os testes empregados em ambientes clínicos podem ser de inteligência, de personalidade, neuropsicológicos ou outros instrumentos especializados, dependendo da área de problema apresentada ou suspeitada. A marca registrada da testagem em ambientes clínicos é que o teste ou a técnica de mensuração são empregados com apenas um indivíduo de cada vez. A testagem grupal é usada principalmente para fins de triagem – ou seja, para identificar aqueles indivíduos que requerem uma nova avaliação diagnóstica.

REFLITA...

Que tipos de questões os psicólogos têm de considerar ao avaliar prisioneiros, comparando com a avaliação de executivos no local de trabalho?

Ambientes de orientação A avaliação em um contexto de orientação pode ocorrer em ambientes tão distintos quanto escolas, prisões e instituições governamentais ou privadas. Independentemente dos instrumentos específicos usados, o objetivo final de muitas dessas avaliações é a melhora do avaliando em termos de ajustamento, produtividade ou alguma variável relacionada. Medidas de habilidades sociais e acadêmicas e medidas de personalidade, interesse, atitudes e valores estão entre os muitos tipos de testes que um conselheiro poderia administrar a um cliente. As questões de encaminhamento a serem respondidas variam de “Como esta criança pode se concentrar melhor nas tarefas?” e “Para que carreira o cliente é mais adequado?” a “Que atividades são recomendadas para a aposentadoria?”. Tendo mencionado a aposentadoria, vamos apressar a introdução de outro tipo de ambiente no qual os testes psicológicos são muito usados.

Ambientes geriátricos Nos Estados Unidos, mais de 12 milhões de adultos estão atualmente na faixa etária de 75 a 84 anos; isso é cerca de 16 vezes mais pessoas nessa faixa etária do que havia em 1900. Quatro milhões de adultos no país têm hoje 85 anos ou são mais velhos, o que constitui um aumento de 33 vezes no número de pessoas dessa idade desde 1900. As pessoas nos Estados Unidos estão vivendo mais, e a população como um todo está envelhecendo.

Os norte-americanos mais velhos podem viver em casa, em residências especiais projetadas para a vida independente, em residências projetadas para vida assistida ou em instituições de cuidados de longo prazo, como hospitais e asilos. Onde quer que residam, os indivíduos mais velhos podem, em algum ponto, requerer avaliação psicológica para estimar o funcionamento cognitivo, psicológico, adaptativo ou outro. Uma questão em muitas dessas avaliações é o quanto os avaliandos estão desfrutando da melhor *qualidade de vida* possível. A definição de qualidade de vida tem variado, dependendo do ponto de vista, em diferentes estudos. Em algumas pesquisas, por exemplo, qualidade de vida é definida do ponto de vista de um observador; em outra, é definida da perspectiva dos próprios avaliandos e refere-se a um autorrelato sobre variáveis relacionadas ao estilo de vida. Qualquer que seja a definição, o que é normalmente avaliado nas estimativas de **qualidade de vida** são variáveis relativas a estresse percebido, solidão, fontes de satisfação, valores pessoais, qualidade das condições de vida e qualidade das amizades e de outros apoios especiais.

Falando de modo geral, de um ponto de vista clínico, a avaliação de adultos mais velhos tem mais probabilidade de incluir a triagem para declínio cognitivo e *demência* do que a de adultos mais jovens (Gallo e Bogner, 2006; Gallo e Wittink, 2006). **Demência** é uma

REFLITA...

Os testes são usados em ambientes geriátricos, de orientação e outros para ajudar a melhorar a qualidade de vida. Mas há alguns aspectos da qualidade de vida que um teste psicológico simplesmente não pode medir?

perda do funcionamento cognitivo (que pode afetar a memória, o pensamento, o raciocínio, a velocidade psicomotora, a atenção e as capacidades relacionadas, bem como a personalidade) que ocorre como resultado de dano a, ou perda, de células cerebrais. Talvez a mais conhecida das muitas formas de demência seja a doença de Alzheimer. O caminho até o diagnóstico pelo clínico é complicado pelo fato de que a depressão grave no idoso pode contribuir para o funcionamento cognitivo que imita demência, uma condição conhecida como **pseudodemência** (Madden et al., 1952). Também é verdade que a maioria dos indivíduos que sofrem de demência exibe sintomas depressivos (Strober e Arnett, 2009). Os clínicos contam com vários instrumentos de avaliação para fazer um diagnóstico de demência ou de pseudodemência.

Ambientes empresariais e militares Nas empresas, como nas Forças Armadas, vários instrumentos de avaliação são usados de diversas formas, talvez mais notavelmente na tomada de decisão sobre as carreiras do pessoal. Uma ampla variedade de testes de realização, aptidão, interesse, motivação e outros pode ser empregada na decisão de contratar, bem como nas decisões relativas a promoções, transferências, satisfação no trabalho e necessidade de treinamento adicional. Para um futuro controlador de tráfego aéreo, o desempenho bem-sucedido em um teste de atenção constante a detalhes pode ser um requisito de emprego. Visando à promoção para o posto de oficial nas Forças Armadas, o desempenho bem-sucedido em uma série de tarefas de liderança pode ser fundamental.

Outra aplicação dos testes psicológicos envolve a engenharia e o *design* de produtos e ambientes. Os psicólogos da engenharia empregam inúmeros testes existentes e especialmente concebidos na pesquisa, destinados a ajudar pessoas em casa, no local de trabalho e nas Forças Armadas. Produtos variando de computadores domésticos a mobília de escritório e painéis de controle de aviões beneficiam-se do trabalho desses esforços de pesquisa.

REFLITA...

Assuma o papel de um psicólogo de *marketing*. Que campanha publicitária você considera particularmente eficaz em termos de estimular o consumidor a “comprar” botões? Que campanha publicitária considera particularmente ineficaz nesse sentido? Por quê?

Usando testes, entrevistas e outros instrumentos de avaliação, psicólogos que se especializam em divulgação e venda de produtos estão empenhados em sentir o pulso do consumidor. Eles ajudam as corporações a prever a receptividade do público a um novo produto, a uma nova marca ou a uma nova propaganda ou campanha publicitária. Psicólogos que trabalham na área de *marketing* ajudam a “diagnosticar” o que está errado (ou certo) em relação a marcas, produtos e campanhas. Com base nessas avaliações, esses psicólogos poderiam fazer recomendações relativas a como novas marcas e novos produtos poderiam se tornar mais atraentes aos

consumidores e quando é hora de marcas e produtos mais antigos serem retirados ou revitalizados.

NO BRASIL

Não há um exame desse tipo para psicólogos. Basta a apresentação do diploma para o registro em um Conselho Regional de Psicologia (CRP).

Credenciamento governamental e organizacional Uma das muitas aplicações de determinadas medidas é no licenciamento governamental, na certificação ou no credenciamento geral de profissionais. Antes de serem legalmente habilitados a praticar a medicina, os médicos devem passar em um exame. Os graduados em direito não podem se apresentar ao público como advogados até terem passado no exame da ordem. Os psicólogos, da mesma forma, devem passar em um exame antes de adotar o título oficial de “psicólogo”.

Os membros de algumas profissões formaram organizações com requisitos para afiliação que vão além dos de licenciamento ou certificação. Por exemplo, médicos podem fazer treinamento especializado adicional e um exame de especialidade para obter a distinção de serem “certificados pelo conselho” em uma determinada área da medicina. Psicólogos que se especializam em certas áreas podem ser avaliados para receber um diploma do American Board of Professional Psychology (ABPP) reconhecendo a excelência na prática da psicologia. Outra organização, a American Board of Assessment Psychology (ABAP), concede seu diploma com base em um exame para aplicadores de testes, desenvolvedores de testes e outros que se distinguiram no campo da testagem e da avaliação.

Ambientes de pesquisa acadêmica Conduzir qualquer tipo de pesquisa em geral implica algum tipo de mensuração, e qualquer acadêmico que espere um dia publicar uma pesquisa deve ter, idealmente, um conhecimento sólido dos princípios de mensuração e dos instrumentos de avaliação. Para enfatizar esse simples fato da vida de pesquisa, considere a Tabela 1.3. Ela contém uma amostra (muito) aleatória de alguns estudos recém-publicados, com informação relativa às questões de pesquisa respondidas e aos instrumentos de avaliação usados para pesquisar a resposta àquela questão.

Outros contextos Muitos tipos diferentes de procedimentos de mensuração encontram aplicação em uma ampla variedade de contextos. Por exemplo, os tribunais baseiam-se em dados de testes psicológicos e testemunho pericial relacionado como uma fonte de informação para ajudar na resposta a perguntas importantes como “O réu é passível de ser submetido a julgamento?” e “Este réu sabia diferenciar o certo do errado no momento em que o ato criminoso foi cometido?”.

A mensuração pode desempenhar um papel importante na avaliação de programas, seja um programa governamental de larga escala ou um de financiamento privado de pequena escala. O programa está funcionando? Como ele pode ser melhorado? Os recursos estão sendo gastos nas áreas em que deveriam ser? Qual é a solidez da teoria na qual o programa é baseado? Esses são tipos de questões gerais a que os testes e os procedimentos de mensuração usados na avaliação do programa devem responder.

O uso dos instrumentos de avaliação pode ser observado na pesquisa e na prática em toda área de especialidade no campo da psicologia. Por exemplo, considere a **psicologia da saúde**, uma disciplina focalizada no entendimento do papel das variáveis psicológicas no início, no curso, no tratamento e na prevenção de doenças e deficiências (Cohen, 1994). Os psicólogos da saúde estão envolvidos no ensino, na pesquisa ou nas atividades de serviços diretos que visam promover a boa saúde. Entrevistas individuais, levantamentos e testes escritos são alguns dos instrumentos que podem ser empregados para ajudar a avaliar o estado atual com relação a alguma doença ou condição, calcular o progresso do tratamento e analisar o desfecho da intervenção. Uma linha geral de pesquisa na psicologia da saúde focaliza-se em aspectos da personalidade, do comportamento ou do estilo de vida relacionados com a saúde física. A metodologia empregada pode incluir relato sobre variáveis do respondente mensuráveis à medida que mudam em resposta a alguma intervenção, como educação, terapia, orientação, mudança na dieta ou mudança nos hábitos. Os instrumentos de mensuração podem ser usados para comparar um grupo de indivíduos de pesquisa de ocorrência natural a outro grupo (como fumantes comparados com não fumantes) a respeito de alguma outra variável relacionada à saúde (tal como longevidade). Muitas das questões levantadas na pesquisa relacionada à saúde têm consequências reais, de vida e morte. Todos esses questionamentos importantes, também os relativos a outras áreas da psicologia, requerem emprego de técnicas de avaliação sólidas.

REFLITA...

Qual é a questão de pesquisa que você acha que precisa ser tratada? Que instrumentos de avaliação seriam usados em um estudo para tratar essa questão?

Tabela 1.3 Instrumentos de avaliação usados em uma amostra (aleatória) de estudos de pesquisa

| A questão da pesquisa | Instrumento(s) de avaliação empregado(s) |
|---|--|
| Qual é o papel da inspiração no processo de escrita? (Thrash et al., 2010). | (1) Questionários especialmente idealizados, tal como um que comprove a ocorrência de ideias criativas durante a semana anterior, pontuado em uma escala de 1 (nunca) a 7 (muito frequentemente); (2) um ensaio de quatro páginas escrito pelos participantes do estudo; e (3) vários testes escritos, incluindo parte de um teste publicado chamado de Escala da Inspiração. |
| O quanto as medidas de relacionamento com os pares e as medidas comportamentais da infância predizem o comportamento de comunicação do adulto jovem nas redes sociais da internet? (Mikami et al., 2010) | Nomeações dos pares; (2) observação comportamental e avaliação da interação entre pares; (3) todos ou parte de testes publicados; e (4) número de amigos no <i>Twitter</i> ou no <i>Facebook</i> dos participantes classificado por número de amigos e postagens de amigos, bem como informações relacionadas, tais como fotos de si mesmo (classificadas por adequação) e descrição escrita de si mesmo (classificada por agressão, hostilidade ou negatividade percebidas). |
| Os pombos são mais inteligentes do que os matemáticos? (Herbranson e Schroeder, 2010) | Para os pombos, os dispositivos de mensuração usados foram três câmaras operantes contendo um alimentador com três chaves para serem abertas com o bico, por meio do qual os pássaros podiam ter acesso ao alimento. Cada uma dessas unidades era ligada a um computador que controlava os eventos experimentais, registrava os dados e calculava as estatísticas. Para os seres humanos, o dispositivo de mensuração era um monitor de painel plano de 40 cm equipado com um quadro de toque infravermelho que permitia aos participantes responderem a itens de estímulo tocando a tela. Esse dispositivo era ligado a um computador que controlava os eventos experimentais, registrava os dados e calculava as estatísticas. |
| O que podemos aprender sobre o risco de acidentes de motoristas adolescentes iniciantes observando seu comportamento na direção? (Lee, Simons-Morton et al., 2011) | Análise dos dados do sistema eletrônico de aquisição de dados instalado nos veículos de motoristas de 16 anos recentemente habilitados. O sistema consistia em câmeras, sensores e um computador com disco rígido removível. |
| Como a avaliação dos portfólios de professores pode contribuir para a educação docente eficiente? (Fox et al., 2011) | Avaliação qualitativa de dados do portfólio do professor. |
| O que pode ser aprendido a partir da observação das deliberações de um júri simulado com relação à apresentação de certos tipos de evidência? (Finkelstein e Bastounis, 2010) | Observação comportamental das respostas de júris simulados para manipulações experimentais. |
| Como as crianças que rodaram na primeira série se comparam a seus colegas que passaram para a segunda série com relação aos desfechos psicossociais de longo e de curto prazos? (Wu et al., 2010) | (1) Dados da história de caso obtidos da escola, incluindo registros de realização acadêmica; (2) vários testes escritos publicados e especialmente produzidos; (3) avaliação do comportamento social pelos pares. |
| Em pacientes com esclerose múltipla, qual é a relação entre sua fadiga cognitiva autorrelatada e seu desempenho em uma medida de tempo de resposta? (Bruce et al., 2010) | Uma série de testes, administrados oralmente, por computador ou por escrito, incluindo tudo ou parte de (1) Avaliação Computadorizada do Viés de Resposta, (2) o Teste de Atenção Diária e (3) a Escala de Impacto da Fadiga. |
| Um nível de capacidade hipnótica prediz uma resposta a sugestões visando diminuir a experiência de dor? (Milling et al., 2010) | (1) Escala de Responsividade a Sugestão da Carleton University (uma medida da capacidade hipnótica); (2) um aparelho usado para induzir dor mensurável por pressão do dedo; e (3) a medida da intensidade da dor por meio de uma escala de 11 pontos. |
| Comparados com caucasianos não hispânicos, os caucasianos hispânicos relatam mais sintomas de transtorno de estresse pós-traumático (TEPT)? (Marshall et al., 2009) | (1) Entrevista para verificar a etnia dos participantes; e (2) Lista de Verificação de Transtorno de Estresse Pós-traumático (usada para medir a gravidade do sintoma de TEPT). |

(Continua)

Tabela 1.3 Instrumentos de avaliação usados em uma amostra (aleatória) de estudos de pesquisa (Continuação)

| A questão da pesquisa | Instrumento(s) de avaliação empregado(s) |
|--|---|
| A terapia cognitivo-comportamental de grupo para cessação do tabagismo é eficaz com fumantes afro-americanos? (Webb et al., 2010) | (1) Questionário da história de tabagismo; (2) medida de prontidão para parar de fumar; (3) Teste para Dependência de Nicotina de Fagerstrom; e (4) instrumento para medir o hálito de monóxido de carbono. |
| Como as últimas declarações de condenados no corredor da morte podem ser categorizadas? (Gerhart et al., 2010) | Estudo de caso de registros públicos pelo estado do Texas. |
| Qual o valor de um procedimento de dramatização virtual para avaliar as respostas de mulheres a ameaça sexual? (Jouriles et al., 2011) | Respostas observadas de mulheres na dramatização face a face ou dramatização virtual, envolvendo uma situação sexualmente ameaçadora. |
| Como podemos entender melhor os aspectos emocionais do relacionamento treinador-atleta? (Huguet e Philippe, 2011) | Análise de estudo de caso de quatro entrevistas com um atleta, incluindo análise de relacionamentos passados com treinadores. |
| O potencial para violência de um grupo ideológico pode ser avaliado pelo estudo da página do grupo na internet? (Angie et al., 2011) | Avaliação do material postado nas caixas de mensagens com o objetivo de estimar processos particulares à filiação do grupo ideológico. |
| Como podemos entender melhor o estresse da guerra e seu papel na síndrome de estresse pós-traumático em soldados? (Lee, Goudarzi et al, 2011) | Sistema de avaliação pela internet para registrar e avaliar vários aspectos da experiência de estresse de guerra dos soldados. |

Como as avaliações são conduzidas?

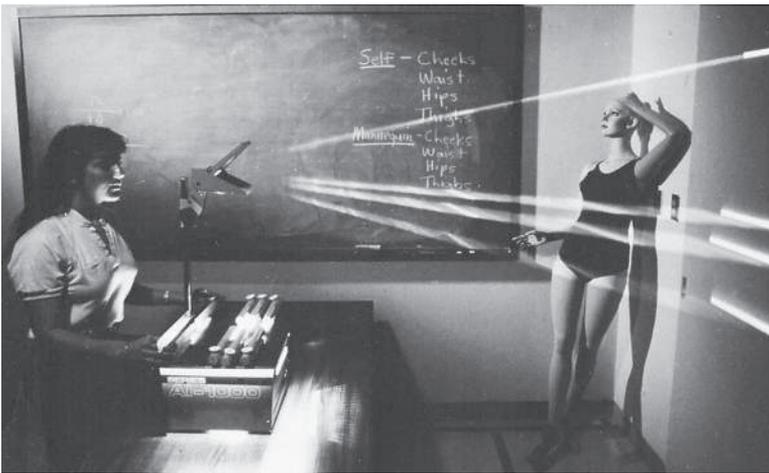
Se existe a necessidade de medir uma determinada variável, uma forma de medi-la será criada. Conforme a Figura 1.6 apenas começa a ilustrar, as formas pelas quais as mensurações podem ser feitas são limitadas apenas pela imaginação. Lembre que essa figura ilustra apenas uma pequena amostra dos muitos métodos usados na testagem e na avaliação psicológicas. As fotos não visam ilustrar os tipos mais típicos de procedimentos de avaliação. Antes, seu propósito é chamar atenção para a ampla variedade de instrumentos de mensuração que foram criados para usos variados.

Os aplicadores de testes responsáveis têm obrigações antes, durante e após um teste ou qualquer procedimento de mensuração ser administrado. Para fins de ilustração, considere a administração de um teste escrito. Antes do teste, as diretrizes éticas prescrevem que, quando os aplicadores têm reservas a respeito dos testes, devem selecionar e usar apenas o teste ou os testes que são mais apropriados para o indivíduo que está sendo testado. Antes de ser administrado, um teste deve ser armazenado com razoável garantia de que seu conteúdo específico não será dado a conhecer ao testando antecipadamente. Outra obrigação do aplicador antes da administração do teste é assegurar que uma pessoa preparada e treinada de maneira adequada o administre com correção.

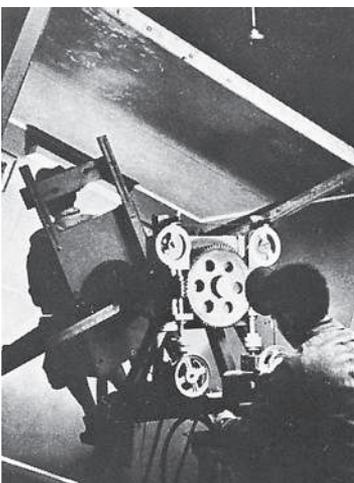
O administrador do teste (ou examinador) deve estar familiarizado com os materiais e procedimentos relacionados com ele e deve ter no local da aplicação todos os materiais necessários para administrá-lo de forma apropriada. Esses materiais poderiam incluir um cronômetro, um suprimento de lápis e um número suficiente de *protocolos* do teste. A propósito, na conversação diária não relacionada a testes, *protocolo* se refere à etiqueta diplomática. Um uso menos comum da palavra é um sinônimo para a primeira cópia ou o rascunho de um tratado ou outro documento oficial antes de sua ratificação. Com referência à testagem e à avaliação, **protocolo** normalmente significa o formulário ou a folha ou o folheto no qual as respostas de um testando são inseridas. O termo também pode ser usado para descrever um conjunto de procedimentos relacionados a teste ou avaliação, como na frase “O examinador seguiu precisamente o protocolo completo para a entrevista de estresse”.



Pelo menos desde o início do século XIX, unidades militares ao redor do mundo se basearam em testes psicológicos e outros testes para seleção de pessoal, validação de programas e razões relacionadas (Hartmann et al., 2003). Em algumas culturas em que o serviço militar é altamente valorizado, estudantes fazem cursos preparatórios na esperança de serem aceitos nas unidades militares de elite. Esse é o caso em Israel, onde o treinamento rigoroso, como o retratado aqui, prepara estudantes do ensino médio para testes físicos e outros relacionados nos quais apenas um em 60 recrutas militares será aprovado.

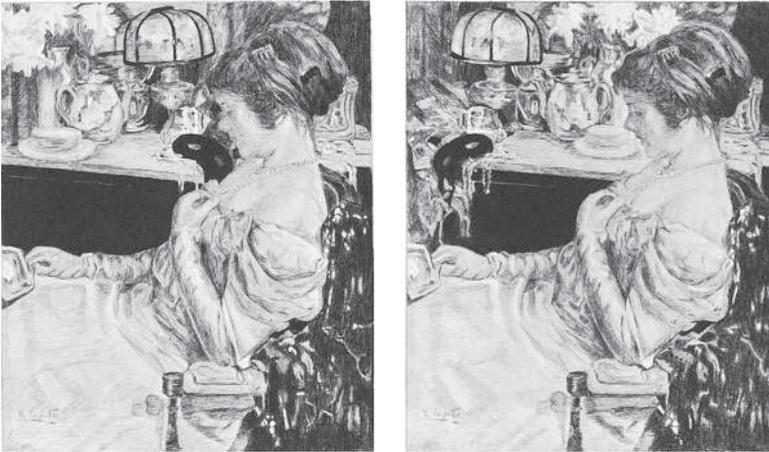


A evidência sugere que algumas pessoas com transtornos alimentares possam na verdade ter um transtorno da autopercepção; ou seja, elas se veem como mais pesadas do que realmente são (Thompson e Smolak, 2001). J. Kevin Thompson e colaboradores criaram o aparelho de feixe de luz ajustável para medir distorções da imagem corporal. Os avaliados ajustam quatro feixes de luz para indicar o que acreditam ser a largura de suas bochechas, seus quadris, sua cintura e suas coxas. Uma medida da precisão dessas estimativas é então obtida.



Herman Witkin e colaboradores (Witkin e Goodenough, 1977) estudaram variáveis relacionadas à personalidade de algumas formas inovadoras. Por exemplo, identificaram pessoas dependentes do campo (ou contexto) e independentes do campo por meio deste dispositivo "sala inclinada-cadeira inclinada". Perguntas eram feitas aos avaliados visando avaliar sua dependência ou independência de sugestões visuais.

Figura 1.6 O extenso mundo da mensuração.



Figuras como estes exemplos do Teste de Julgamento Artístico de Meier podiam ser usadas para avaliar a percepção estética das pessoas. Qual destas duas interpretações você considera esteticamente mais agradável? A diferença entre as duas figuras envolve o posicionamento dos objetos sobre a estante.



O prejuízo de certas funções sensoriais pode indicar déficit neurológico. Para fins de diagnóstico, bem como para medir o progresso do tratamento, a bola de treinamento do desenvolvimento neurológico pode ser útil na avaliação do senso de equilíbrio da pessoa.



O Stresseraser é um dispositivo de *biofeedback* portátil e autoadministrado que visa facilitar a mudança no relaxamento corporal. O funcionamento do nervo vago é monitorado pela pulsação do dedo indicador e apresentado ao usuário por meio de imagens em uma tela. Os usuários podem então alterar a respiração e o foco mental para afetar o funcionamento simpático-parassimpático a fim de obter benefícios terapêuticos. A unidade tem a vantagem da portabilidade; ela pode ser usada para facilitar o relaxamento em uma variedade de situações.



Figura 1.7 Condições de testagem abaixo do ideal.

Em 1917, novos recrutas do exército sentavam-se no chão enquanto era administrado o primeiro teste de inteligência em grupo – condições de testagem não ideais pelos padrões atuais.

Os aplicadores de testes têm a responsabilidade de assegurar que a sala na qual o teste será conduzido seja adequada e propícia à testagem. À medida do possível, distrações como ruído excessivo, calor, frio, interrupções, luz solar ofuscante, aglomeração, ventilação inadequada, e assim por diante, devem ser evitadas. Naturalmente, criar um ambiente de testagem ideal nem sempre é possível (ver Fig. 1.7).

Durante a administração do teste, e sobretudo na testagem individual ou de grupo pequeno, o *rapport* entre o examinador e o examinando pode ser de fundamental importância. Nesse contexto, *rapport* pode ser definido como um relacionamento de trabalho entre o examinador e o examinando. Tal relacionamento às vezes pode ser alcançado com algumas palavras trocadas quando ambos são apresentados. Se apropriado, algumas palavras sobre a natureza do teste e por que é importante que os examinandos façam o melhor possível também pode ser útil. Em outros casos – por exemplo, com uma criança assustada – a obtenção do *rapport* poderia abranger técnicas mais elaboradas, como envolver a criança no brinquedo ou em alguma outra atividade até que ela se adapte ao examinador e ao ambiente. É importante que as tentativas de estabelecer o *rapport* com o testando não comprometam qualquer regra das instruções de administração do teste.

Após um teste ter sido aplicado, os aplicadores também têm muitas obrigações. Estas variam de salvaguardar os protocolos do teste a transmitir seus resultados de forma claramente compreensível. Se outras pessoas estiveram presentes durante a testagem ou se aconteceu qualquer outra coisa que pudesse ser considerada fora do comum nessa ocasião, é responsabilidade do aplicador do teste anotar esses eventos no relatório da testagem. Os responsáveis pela contagem dos pontos do teste também têm obrigações. Por exemplo, se um teste deve ser pontuado por pessoas, a pontuação precisa estar de acordo com critérios preestabelecidos. Os aplicadores que têm a responsabilidade de interpretar

as pontuações ou outros resultados do teste devem fazê-lo de acordo com procedimentos e diretrizes éticos estabelecidos.

Avaliação de pessoas com deficiências Pessoas com deficiências são avaliadas exatamente pelas mesmas razões que o são aquelas sem deficiências: para obter emprego, ganhar uma credencial profissional, ser avaliado para psicopatologia, e assim por diante. Inúmeras leis foram promulgadas que afetam as condições sob as quais os testes são administrados a pessoas com condições incapacitantes. Por exemplo, uma lei obriga o desenvolvimento e a implementação de programas de “avaliação alternativa” para crianças que, como resultado de uma deficiência, não poderiam de outro modo participar das avaliações estaduais e distritais gerais. A definição do exato significado de “avaliação alternativa” ficou a cargo de cada estado ou de seus distritos escolares locais. Essas autoridades definem quem requer avaliação alternativa, como essas avaliações devem ser conduzidas e como inferências significativas devem ser feitas a partir dos dados da avaliação.

Em geral, a avaliação alternativa é normalmente realizada por meio de alguma *acomodação* feita para o avaliando. O verbo *acomodar* pode ser definido como “adaptar, ajustar ou tornar adequado”. No contexto da testagem e da avaliação psicológica, **acomodação** pode ser definida como a adaptação de um teste, um procedimento ou uma situação, ou a substituição de um teste por outro, para tornar a avaliação mais apropriada a um avaliando com necessidades excepcionais.

À primeira vista, o processo de acomodar estudantes, empregados ou outros testandos com necessidades especiais poderia parecer claro e direto. Por exemplo, o indivíduo que tem dificuldade para ler as letras pequenas de um determinado teste pode ser acomodado com uma versão de letras grandes do mesmo teste ou com um ambiente de teste especialmente iluminado. Um estudante com um prejuízo auditivo pode responder ao teste na linguagem de sinais. Indivíduos com TDAH poderiam ter um tempo de avaliação estendido, com intervalos frequentes durante os períodos de avaliação. Embora tudo possa parecer simples a princípio, na verdade essa situação pode se tornar bastante complicada. Considere, por exemplo, o caso de um estudante com um prejuízo de visão que esteja programado para receber um teste escrito de múltipla escolha. Há diversos procedimentos alternativos possíveis para administrá-lo. Por exemplo, o teste poderia ser traduzido para Braille e aplicado nesse formato, ou poderia ser administrado por meio de fitas de áudio. Entretanto, alguns estudantes podem se sair melhor com uma administração em Braille e outros com fitas de áudio. Estudantes com habilidades de atenção e memória de curto prazo superiores para estímulos auditivos parecem ter uma vantagem com a administração em fitas de áudio. Estudantes com habilidades hápticas (sentido do tato) e perceptomotoras superiores poderiam ter uma vantagem com a administração em Braille. E portanto, mesmo nesse exemplo bastante simples, pode ser facilmente percebido que o desempenho (e a pontuação) de um testando pode ser afetado pela maneira da administração alternativa do teste. Essa realidade da avaliação alternativa levanta questões importantes sobre o quanto esses métodos são realmente equivalentes. De fato, visto que os procedimentos alternativos foram adaptados ao indivíduo, existem raras pesquisas convincentes apoiando a equivalência. As diretrizes governamentais para a avaliação alternativa evoluirão para incluir as formas de traduzir os procedimentos de mensuração de um formato para outro. Outras diretrizes podem sugerir a substituição de um instrumento de avaliação por outro. Atualmente, há muitas formas de acomodar pessoas com deficiências em uma situação de avaliação (ver a seção *A psicometria no cotidiano* neste capítulo) e muitas definições diferentes de *avaliação alternativa*. Só para registro, oferecemos nossa própria definição geral desse termo evasivo. **Avaliação alternativa** é um procedimento, ou processo

REFLITA...

Que incidentes inesperados poderiam ocorrer durante uma sessão de teste? Tais incidentes devem ser anotados no relatório da sessão?

NO BRASIL

Aqui é abordada a legislação dos Estados Unidos. No Brasil, as diretrizes são fixadas pelo Conselho Federal de Psicologia (CPF, www.cfp.org.br) ou por legislação federal.

A PSICOMETRIA NO COTIDIANO

Acomodações diárias

Foi estimado que um sétimo dos norte-americanos tem uma deficiência que interfere nas atividades da vida diária. Nos últimos anos, a sociedade tem reconhecido mais do que nunca as necessidades especiais dos cidadãos desafiados por deficiências físicas e/ou mentais. Os efeitos desse crescente reconhecimento são visíveis: rampas de acesso especiais ao lado de lances de escadas, programação de televisão legendada para deficientes auditivos e jornais, livros e revistas com letras grandes para os deficientes visuais. Em geral, tem havido uma tendência no sentido de alterar os ambientes para que os indivíduos com condições de desvantagem se sintam menos desafiados.

Dependendo da natureza da deficiência de um testando e de outros fatores, modificações – referidas como acomodações – podem ser necessárias em um teste psicológico (ou procedimento de mensuração) a fim de que a avaliação seja possível. A acomodação pode assumir diferentes formas. Um tipo de acomodação geral envolve a forma como o teste é apresentado ao testando, tal como ocorre quando um teste escrito é definido em um tipo maior para apresentação a um testando com deficiência visual. Outro tipo diz respeito à forma como as respostas aos testes são obtidas. Por exemplo, poderia ser permitido que um indivíduo com deficiência de fala escrevesse as respostas em um exame que, de outro modo, seria administrado oralmente. Estudantes com deficiências de aprendizagem podem ser acomodados sendo permitido que leiam as perguntas do teste em voz alta (Fuchs et al., 2000).

A modificação do ambiente físico no qual um teste é conduzido é outro tipo de acomodação geral. Por exemplo, um teste que costuma ser administrado a grupos em um local central pode ocasionalmente ser aplicado de forma individual a uma pessoa deficiente em sua casa. As modificações do ambiente interpessoal no qual um teste é conduzido é outra possibilidade (ver Fig. 1).

Quais dos muitos tipos diferentes de acomodação devem ser empregados? Uma resposta a essa pergunta é em geral abordada pela consideração de pelo menos quatro variáveis:

1. as capacidades do avaliando;
2. o propósito da avaliação;
3. o significado atribuído às pontuações do teste; e
4. as capacidades do avaliador.

As capacidades do avaliando

Qual dos vários meios alternativos de avaliação é mais apropriado às necessidades e capacidades do avaliando? Dados da história de caso, registros de avaliações anteriores e entrevistas com amigos, família, professores e outros que conhecem o avaliando podem, todos, fornecer uma riqueza de informações úteis relativas a qual dos diversos meios de avaliação alternativos é o mais adequado.

O propósito da avaliação

A acomodação é adequada sob algumas circunstâncias e inadequada sob outras. Em geral se examina o propósito da avaliação



Figura 1 Modificação do ambiente interpessoal.

Um testando individual que necessita da ajuda de um auxiliar ou cão guia pode requerer a presença de uma terceira pessoa (ou animal) se um teste em particular for administrado. Em alguns casos, devido à natureza da deficiência do testando e das demandas de um determinado teste, um teste mais adequado poderia substituir o habitual se uma avaliação significativa tiver que ser conduzida.

e as consequências da acomodação a fim de julgar a adequação de modificar um teste para acomodar uma pessoa com uma deficiência. Por exemplo, modificar um teste de direção escrito – ou um exame de direção prático – para que uma pessoa cega pudesse ser testada para tirar a carteira de motorista é evidentemente inadequado. Para sua própria segurança, bem como a do público, os cegos são proibidos de dirigir automóveis. Entretanto, mudar a forma da maioria dos outros testes escritos a fim de que uma pessoa cega pudesse fazê-los é uma coisa bem diferente. Em geral, a acomodação é apenas uma forma de ser fiel a uma política social

que promove e garante oportunidades e tratamento iguais para todos os cidadãos.

O significado atribuído às pontuações do teste

O que acontece ao significado de uma pontuação em um teste quando ele não foi administrado da maneira que deveria ter sido? Com muita frequência, quando as instruções de administração do teste são modificadas (alguns diriam “comprometidas”), o significado de suas pontuações se torna no mínimo questionável. Os aplicadores de testes são deixados à própria sorte para interpretar esses dados. O julgamento profissional, a perícia e, francamente, as suposições podem, todos, entrar no processo de fazer inferências a partir das pontuações nos testes modificados. De fato, um registro preciso de como exatamente um teste foi modificado para fins de acomodação deve ser feito no relatório do teste.

As capacidades do avaliador

Embora a maioria das pessoas encarregadas da responsabilidade de avaliar gostem de pensar que podem administrar uma avaliação profissional a qualquer pessoa, esse não é de fato o caso. É importante reconhecer que alguns avaliadores podem experimentar um nível de desconforto na presença de pessoas com determi-

nadas deficiências, e esse desconforto pode afetar sua avaliação. Também é importante reconhecer que outros podem necessitar de treinamento adicional antes de conduzirem certas avaliações, incluindo experiência supervisionada com membros de determinadas populações. Como alternativa, o avaliador pode encaminhar essas avaliações a outro avaliador que teve treinamento e experiência com membros dessas populações.

Uma literatura acadêmica florescente tem se focalizado em vários aspectos da acomodação, incluindo questões relacionadas a políticas gerais (Burns, 1998; Nehring, 2007; Shriner, 2000; Simpson et al., 1999), método de administração do teste (Calhoon et al., 2000; Danford e Steinfeld, 1999), compatibilidade da pontuação (Elliott et al., 2001; Johnson, 2000; Pomplun e Omar, 2000, 2001), documentação (Schulte et al., 2000) e a motivação dos testandos para requerer acomodação (Baldrige e Veiga, 2006). Antes que uma decisão sobre acomodação seja tomada para qualquer testando individual, deve ser dada a devida consideração a questões relativas ao significado das pontuações derivadas de instrumentos modificados e à validade das inferências que podem ser feitas a partir dos dados derivados.

avaliativo ou diagnóstico, que varia da forma habitual, costumeira ou padronizada da qual uma medida é derivada, ou em virtude de alguma acomodação especial feita para o avaliando ou por meio de métodos alternativos destinados a medir a(s) mesma(s) variável(s). Tendo considerado alguns dos *quem, o que, como e por que* da avaliação, vamos agora considerar as fontes para mais informações com relação a todos os aspectos do empreendimento da avaliação.

Onde buscar informações oficiais: fontes de referência

Existem muitas fontes de referência para conhecer mais sobre os testes publicados e as questões relacionadas à avaliação. Essas fontes variam com respeito ao detalhe. Algumas meramente fornecem descrições de testes, outras, informações detalhadas sobre aspectos técnicos, e outras, ainda, revisões críticas completas com discussão dos prós e contras do uso.

Catálogos de testes Talvez uma das fontes de informação mais facilmente acessível seja um catálogo distribuído pelo editor do teste.

Visto que a maioria dos editores de testes disponibilizam catálogos de suas ofertas, essa fonte de informação pode ser obtida por um simples telefonema, *e-mail* ou bilhete. Como se poderia esperar, entretanto, esses catálogos em geral contêm apenas uma breve descrição do teste e raras vezes contêm o tipo de informação técnica detalhada que um futuro aplicador poderia necessitar. Além disso, o objetivo do catálogo é vender o teste. Por essa razão, revisões com altas críticas de um teste raramente, ou nunca, são encontradas no catálogo de um editor.

Manuais do teste Informações detalhadas relativas ao desenvolvimento de um determinado teste e informações técnicas sobre ele devem ser encontradas no manual, que geralmente pode ser comprado do editor do teste. Porém, para fins de segurança, o editor costuma solicitar a documentação de treinamento profissional antes de preencher um pedido

REFLITA...

Existe algum tipo de avaliação para a qual nenhum procedimento alternativo deva ser desenvolvido?

NO BRASIL

Consultar o SATEPSI: <http://www.satepsi.cfp.org.br>.

para um manual de teste. São boas as chances de que sua universidade mantenha uma coleção de manuais de testes populares, talvez na biblioteca ou no centro de orientação. Se o manual do teste que você procura não está disponível lá, pergunte a seu professor a melhor maneira de obter uma cópia de referência. Ao examinar os vários manuais, é provável que você veja que variam não apenas nos detalhes de como os testes foram desenvolvidos e julgados psicometricamente sólidos, mas também na franqueza com que eles descrevem as limitações de seus próprios testes.

Livros de referência O Buros Institute of Mental Measurements fornece um “balcão único” (*one-stop shopping*) para uma grande quantidade de informação relacionada aos testes. A versão inicial do que viria a ser o *Mental Measurements Yearbook* foi compilada por Oscar Buros (Fig. 1.8) em 1933. Até o momento, a edição mais atual dessa compilação oficial de revisões de testes é o *18º Annual Mental Measurements Yearbook*, publicado em 2010 (embora o décimo nono possa não estar muito longe). O Buros Institute também distribui *Tests in Print*, que lista todos os testes de língua inglesa disponíveis comercialmente que estão sendo editados. Esse livro, que também recebe atualizações periódicas, fornece informações detalhadas de cada teste listado, incluindo o editor, o autor, o propósito, a população visada e o tempo de administração.

Artigos de jornais Os artigos nos jornais atuais podem conter revisões do teste, estudos atualizados ou independentes de sua solidez psicométrica ou exemplos de como o instrumento foi usado na pesquisa ou em um contexto aplicado. Esses artigos podem aparecer em uma ampla série de revistas de ciência do comportamento, como *Psychological Bulletin*, *Psychological Review*, *Professional Psychology: Research and Practice*, *Journal of Personality and Social Psychology*, *Psychology & Marketing*, *Psychology in the Schools*, *School Psychology Quarterly* e *School Psychology Review*. Há publicações que se concentram especificamente em questões relativas à testagem e à avaliação. Por exemplo, dê uma olhada em publicações como *Journal of Psychoeducational Assessment*, *Psychological Assessment*, *Educational and Psychological Measurement*, *Applied Measurement in Education* e *Journal of*



Figura 1.8 Oscar Krisen Buros (1906-1978).

Buros é mais lembrado como criador do *Anuário de medidas mentais* (MMY), um tipo de *Relatórios do consumidor* para testes e uma fonte muito necessária de “policimento psicométrico” (Peterson, 1997, p. 718). Sua obra sobrevive no Buros Institute of Mental Measurements (Instituto Buros de Medidas Mentais), localizado na University of Nebraska, em Lincoln. Além do MMY, que é atualizado periodicamente, o instituto publica uma variedade de outros materiais de teste.



Personality Assessment. Publicações como *Psychology, Public Policy, and Law* e *Law and Human Behavior* com frequência contêm artigos bastante informativos sobre questões legais e éticas e sobre controvérsias relacionadas à testagem e à avaliação psicológica. Publicações como *Computers & Education, Computers in Human Behavior* e *Cyberpsychology, Behavior, and Social Networking* frequentemente trazem artigos perspicazes sobre medidas relacionadas a computadores e a internet.

NO BRASIL

Para revisões de testes brasileiros, a principal revista é a *Avaliação Psicológica*, disponível *on-line* (www.pepsic.bvsalud.org).

Banco de dados da internet Um dos bancos de dados bibliográficos mais amplamente usado para publicações relacionadas a testes é mantido pelo Educational Resources Information Center (ERIC). Patrocinado pelo U.S. Department of Education e operado da University of Maryland, a página do ERIC em www.eric.ed.gov contém fartos recursos e novidades sobre testes, testagem e avaliação. Há trechos de artigos, artigos originais e *links* para outras páginas úteis. O ERIC tenta fornecer informação equilibrada relativa à avaliação educacional e a recursos que encorajem o uso responsável dos testes.

A American Psychological Association (APA) mantém uma série de bancos de dados úteis para localizar informações relacionadas à psicologia em artigos de revistas, capítulos de livros e dissertações de doutorado. O PsycINFO é um banco de dados de artigos completos a partir de 1887. O ClinPSYC é um banco de dados derivado do PsycINFO que focaliza artigos de natureza clínica. PsycSCAN: Psychopharmacology contém artigos a respeito de psicofarmacologia. O PsycARTICLES é um banco de dados de artigos completos a partir de 1894. O Health and Psychosocial Instruments (HAPI) contém uma listagem de medidas criadas ou modificadas por estudos de pesquisa específicos, mas não disponíveis comercialmente; ele está disponível em muitas bibliotecas de universidades por meio da BRS Information Technologies e também em CD-ROM (atualizado duas vezes por ano). O PsycLAW é um banco de dados livre, acessível a qualquer pessoa, que contém discussões de temas selecionados envolvendo psicologia e lei. Ele pode ser acessado em www.apa.org/psyclaw. Para mais informação sobre qualquer um desses bancos de dados, visite a página da APA na internet em www.apa.org.

A maior instituição privada de mensuração do mundo é o Educational Testing Service (ETS). Essa companhia, com sede em Princeton, Nova Jersey, mantém uma equipe de cerca de 2.500 pessoas, incluindo aproximadamente mil profissionais da mensuração e especialistas em educação. Essas são as pessoas que oferecem o Teste de Aptidão Escolar (SAT) e o Graduate Record Exam (GRE) entre muitos outros testes. As descrições desses e de muitos outros testes desenvolvidos por essa companhia podem ser encontradas em sua página da internet, www.ets.org.

Outras fontes A biblioteca de sua escola contém inúmeras outras fontes que podem ser usadas para obter informações sobre testes e assuntos relacionados. Por exemplo, duas fontes para explorar o mundo dos testes e medidas não publicados são o *Directory of Unpublished Experimental Mental Measures* (Goldman e Mitchell, 2007) e o *Tests in Microfiche*, administrado pelo ETS. Alguns dos prós e contras das várias fontes de informação que listamos estão resumidos na Tabela 1.4.

Bibliotecas de muitas universidades também proporcionam o acesso a bancos de dados da internet, como o PsycINFO e revistas eletrônicas. A maioria dos ensaios científicos pode ser baixada direto para o computador de qualquer pessoa usando um serviço de internet. Esse é um recurso extremamente valioso para os estudantes, uma vez que os não assinantes desses bancos de dados podem ter de pagar taxas pesadas para esse acesso.

Munidos de fartas informações práticas sobre testes e outros instrumentos de avaliação, vamos explorar os aspectos históricos, culturais e legais/éticos do empreendimento da avaliação no próximo capítulo.

Tabela 1.4 Fontes de informação sobre testes: alguns prós e contras

| Fonte de informação | Prós | Contras |
|--|---|--|
| Catálogo de testes disponíveis do editor do teste bem como de distribuidores associados | Contém a descrição geral do teste, incluindo para o que se destina e com quem deve ser usado. Facilmente disponível para a maioria das pessoas que solicitam um catálogo. | Concebidos a princípio para vender o teste a seus aplicadores e raramente contêm alguma revisão crítica. A informação não é detalhada o suficiente para embasar uma decisão de usar o teste. |
| Manual do teste | Em geral a fonte mais detalhada disponível para informação relativa à amostra de padronização e instruções de administração do teste. Também pode conter informação útil sobre no que se baseia a teoria do teste, se for o caso. Normalmente contém pelo menos alguma informação relativa à solidez psicométrica do teste. | Detalhes relativos à solidez psicométrica do teste geralmente são oportunistas e escritos com base em estudos conduzidos por seu autor e/ou por seu editor. Pode ser difícil para os estudantes obterem o próprio manual do teste, já que sua distribuição pode ser restrita a profissionais qualificados. |
| Livros de referência, como o <i>Annual Mental Measurements Yearbook</i> , disponíveis na forma de livro ou na internet | Semelhante a um <i>Relatório do consumidor</i> para testes, contém descrições e revisões críticas de um teste escrito por terceiros que presumivelmente não têm nada a ganhar ou a perder elogiando ou criticando o instrumento, sua amostra de padronização e sua solidez psicométrica. | Poucas desvantagens se o revisor estiver genuinamente tentando ser objetivo e for competente, mas, como qualquer revisão, pode fornecer um quadro enganador se este não for o caso. Além disso, para relatos muito detalhados da amostra de padronização e questões relacionadas, é melhor consultar o próprio manual do teste. |
| Artigos de revistas | Fonte atualizada de revisões e estudos de solidez psicométrica. Pode fornecer exemplos práticos de como um instrumento é usado na pesquisa ou em contextos aplicados. | Da mesma forma que os livros de referência, as revisões são valiosas na medida em que são informativas e, tanto quanto possível, imparciais. O leitor deve pesquisar o máximo de artigos possível ao tentar aprender como o instrumento é realmente usado; qualquer artigo isolado pode fornecer um quadro atípico. |
| Bancos de dados da internet | Os bancos de dados da internet amplamente conhecidos e respeitados, como o ERIC, são “minas de ouro” virtuais de informações úteis, contendo quantidades variáveis de detalhes. Embora alguns testes psicológicos legítimos possam estar disponíveis para autoadministração e pontuação pela internet, a grande maioria não está. | Consumidor, cuidado! Alguns sites disfarçados de bancos de dados para testes psicológicos visam mais entreter ou vender alguma coisa do que informar. Esses sites com frequência oferecem testes que você pode fazer pela internet. À medida que aprender mais sobre testes, provavelmente você se tornará mais crítico quanto valor desses “testes psicológicos” autoadministrados e autopontuados. |

Autoavaliação

Teste seu entendimento dos elementos deste capítulo vendo se você pode explicar cada um dos seguintes termos, expressões e abreviações:

acomodação
aplicador do teste
autópsia psicológica
avaliação alternativa
avaliação dinâmica
avaliação informal
avaliação psicológica
avaliação psicológica colaborativa

avaliação psicológica terapêutica
CAPA
catálogo de testes
dados da história de caso
demência
desenvolvedor de testes
diagnóstico
dramatização

entrevista
entrevista em painel
estudo de caso
facilitação social
formato
história de caso
manual do teste
observação comportamental
observação naturalista

| | | |
|---|-------------------------------------|-----------------------|
| outras pessoas envolvidas na avaliação psicológica | psicologia da saúde | solidez psicométrica |
| pensamento grupal | psicometria | teleprocessamento |
| pontos de corte | psicometrista | testagem psicológica |
| pontuação | qualidade de vida | testando |
| pontuar | <i>rapport</i> | teste |
| portfólio | relatório consultivo | teste de dramatização |
| processamento central | relatório de pontuação estendido | teste de realização |
| processamento local | relatório de pontuação simples | teste diagnóstico |
| protocolo | relatório integrativo | teste psicológico |
| pseudodemência | relatório interpretativo | utilidade |