

Avaliação Prática de Habilidades Clínicas em Medicina

Dr. Luiz E. A. Troncon
Clínica Médica
CRM-SP 20.735

Organizadores

Iolanda de Fátima Lopes Calvo Tibério

Renata Mahfuz Daud-Gallotti

Luiz Ernesto de Almeida Troncon

Milton de Arruda Martins

 **Atheneu**

EDITORA ATHENEU

São Paulo — Rua Jesuíno Pascoal, 30
Tel.: (11) 2858-8750
Fax: (11) 2858-8766
E-mail: atheneu@atheneu.com.br

Rio de Janeiro — Rua Bambina, 74
Tel.: (21)3094-1295
Fax: (21)3094-1284
E-mail: atheneu@atheneu.com.br

Belo Horizonte — Rua Domingos Vieira, 319 — conj. 1.104

CAPA: produzida pela Equipe Atheneu

PRODUÇÃO EDITORIAL: Equipe Atheneu

PROJETO GRÁFICO/DIAGRAMAÇÃO: Triall Composição Editorial Ltda.

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Avaliação prática de habilidades clínicas em medicina / editor Iolanda de Fátima Calvo Tibério...[et al].
-- São Paulo : Editora Atheneu, 2012.

Outros editores: Renata Mahfuz Daud-Gallotti, Luiz Ernesto de Almeida Troncon, Milton de Arruda Martins

Bibliografia.
ISBN 978-85-388-0321-8

1. Clínica médica 2. Habilidades clínicas 3. Medicina - Estudantes I. Tibério, Iolanda de Fátima Calvo. II. Daud-Gallotti, Renata Mahfuz. III. Troncon, Luiz Ernesto de Almeida. IV. Martins, Milton de Arruda.

12-11399

CDD-610
NLM-WB 100

Índices para catálogo sistemático:

1. Habilidades clínicas em medicina 610

TIBÉRIO, I. F. L. C.; DAUD-GALLOTTI, R. M.; TRONCON, L. E. A.; MARTINS, M. A.
Avaliação Prática de Habilidades Clínicas em Medicina

© EDITORA ATHENEU

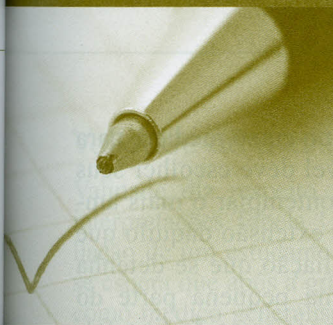
São Paulo, Rio de Janeiro, Belo Horizonte, 2012

Iolanda de Fátima Calvo Tibério
Professora Livre-Docente
Comissão de Internato
Coordenadora do Curso

Renata Mahfuz Daud-Gallotti
Médica Coordenadora
de Medicina da Universidade Nove de Julho

Luiz Ernesto de Almeida Troncon
Professor Titular de Clínica
Preto da Universidade Federal de São Carlos

Milton de Arruda Martins
Professor Titular de Clínica
Universidade de São Paulo
Hospital das Clínicas



capítulo 1

Ruy Guilherme Silveira de Souza

Atributos Fundamentais dos Procedimentos de Avaliação

“...provavelmente existe mais ignorância em aspectos significativos no campo da avaliação, do que em qualquer outra área da educação superior”
(Assessment for Learning in Higher Education. 1995)

“avaliação não é um a questão de medida, mas uma questão de planejamento instrucional”
(Cees P M van der Vleuten & Lambert W T Schuwirth. 2005)

INTRODUÇÃO

Avaliação não é uma atividade a ser realizada no final de um programa ou disciplina. Na verdade, avaliação deve ser parte integrante de todo processo educacional, do início ao fim, pois frequentemente é a parte mais motivadora do processo ensino/aprendizado e, portanto, uma aliada na formação profissional.

Um professor que não tenha clareza do papel da avaliação irá cometer o erro mais comum: deixará essa atividade para o final de seu programa, perdendo um dos principais instrumentos para promover o aprendizado significativo. Desta forma desejamos que o professor passe a pensar a avaliação no dia a dia de sua atividade pedagógica, tornando-a prática e útil. O conceito de utilidade de uma avaliação pode ser representado em uma “fórmula” proposta por Wass¹, onde se utilizam atributos da avaliação para determinar o grau de utilidade de um exame, representada pela equação apresentada na Tabela 1.1.

Esse é o objetivo desse capítulo! Utilizar a discussão de alguns atributos da avaliação como base para o desenvolvimento de um planejamento instrucional que inclua na avaliação uma ferramenta de aprendizagem.

“As únicas coisas que ela quer saber é por que e como e quando e onde ir...”

Belle and Sebastian, If You're Feeling Sinister

AVALIAR POR QUÊ?

Tabela 1.1 Equação da utilidade de uma avaliação.

Equação da utilidade de uma avaliação

Utilidade = confiabilidade x validade x viabilidade x aceitabilidade x impacto educacional

WASS, 2007

Medir e educar são duas tarefas distintas e com atributos próprios.

Frequentemente, quando um professor de Medicina é chamado para realizar uma avaliação, é com o objetivo de garantir que seus estudantes tenham atingido um nível mínimo satisfatório para a progressão dentro de seu curso ou para realizar algum tipo de seleção. Nesses casos, o ofício do professor é somente medir. No entanto, não podemos esquecer que seu principal papel é justamente o de educador e, nesse caso, hierarquizar e selecionar não correspondem exatamente ao que se espera de um processo educativo.

Como educador, o professor médico irá utilizar a avaliação para oferecer ao estudante subsídios para melhorar seu processo de aprendizado, permitindo que o aprendiz monitore seu caminho na aquisição de conhecimentos, habilidades e atitudes, para finalmente atingir o grau de autonomia desejado para o exercício de sua profissão⁶.

Confiabilidade da avaliação

Quando medimos, desejamos que o resultado de nossa medida seja consistente e que permita fazer generalizações, afinal, o estudante graduado na região Sul deveria estar apto para exercer seu ofício, por exemplo, na região Amazônica. Nesse caso, temos que garantir a **confiabilidade** do processo.

Confiabilidade é uma questão primariamente estatística² e, como tal, estará sujeita às mesmas regras do tratamento quantitativo de dados: como ao medirmos o desempenho humano estaremos diante de variáveis com instabilidades inerentes, será frequentemente necessário repetir a medida de diversas formas para se obter algum grau de precisão. Essa precisão estará, portanto, na dependência da amostragem utilizada no teste. As principais amostragens para garantir a estabilidade de um teste dizem respeito à amostragem de conteúdo, de tempo e de examinadores.

A – Amostragem de conteúdo

A avaliação de uma disciplina ou programa que tenha um conteúdo muito extenso, como

normalmente acontece no curso médico, para ser considerada confiável deve escolher seus tópicos de maneira a contemplar o mais amplamente possível toda extensão daquilo que foi ensinado. Uma avaliação que se detenha excessivamente em uma pequena parte do conteúdo poderá ocasionar erros do tipo falso-positivo ou falso-negativo. Um exemplo típico ocorre em testes de desempenho onde o aluno passa um longo tempo com um único paciente, ou um teste cognitivo que aborde somente uma pequena parcela do conteúdo esperado.

B – Amostragem de tempo

Aumentar o tempo de avaliação, independente do tipo de teste, será um dos principais mecanismos para atingir a confiabilidade de um teste. Desta forma, avaliar somente no final do programa, durante poucas horas, não contribui para uma avaliação precisa nem justa. Um estudo analisando a confiabilidade de diferentes tipos de avaliação em função do tempo de teste, demonstrou que mesmo os testes menos estruturados podem se tornar confiáveis se testado por um tempo suficiente³. Aumentaremos a amostragem de tempo de nosso teste se aplicarmos exames várias vezes ao longo do curso.

C – Amostragem de examinadores

Existem algumas situações no ensino médico em que a confiabilidade é extremamente importante. Um exemplo típico são as avaliações que acontecem ao final do sexto ano, quando um fracasso impede o aluno de graduar-se. Tais exames, cujos resultados trarão consequências duradouras e às vezes irreversíveis, são conhecidos como exames de alta aposta. Uma das estratégias utilizadas para aumentar a confiabilidade é justamente aumentando o número de avaliadores.

Podemos concluir que, independentemente do tipo de exame, se testarmos em uma quantidade de tempo suficiente, com uma amostragem de conteúdo representativa e com um número adequado de examinadores, certamente atingiremos uma precisão adequada de nossa medida⁴.

Validade da avaliação

Validade diz respeito às interpretações. Tabela 1.2: Validade de um estudo de caso. Espera-se que após a avaliação de um estudo de caso como médico generalista, o estudante seja capaz de interpretar a validade de um estudo de caso como médico generalista por não saber a última causa de uma pancreatite aguda?

Ao planejar um estudo de caso, o professor deveria saber qual a validade de sua interpretação de acordo com as Diretrizes Nacionais para o curso de Medicina. Formar um médico com níveis de atendimento primário e atendimento de emergência, a competência de um médico generalista, o conhecimento de ressecção cirúrgica, contribuirá para a validade de um estudo de caso?

Enquanto que a validade de um estudo de caso deve-se ter uma amostragem abrangente e representativa do conhecimento de validade o conteúdo de um estudo de caso e adequado às necessidades de um estudo de caso. Conteúdo é, por si só, uma consideração na validade é um conteúdo determinado uso ou avaliação somente com o uso de uma futura prática profissional.

Tabela 1.2 O questionário de validade diária.

Avaliação
<ul style="list-style-type: none"> Seja qual for o tipo de avaliação, a validade é contextualizada e a mais autêntica e autenticidade possível. A busca pela validade de um estudo de caso em ambientes reais e a utilização de vídeo. No Brasil, o estudo de caso é utilizado em avaliações Nacionais e o estudo de caso em escola.

Validade da avaliação

Validade diz respeito à coerência de nossas interpretações Tabela 1.2. O que dizer da avaliação de um estudante de Medicina, de quem se espera que após a graduação possa trabalhar como médico generalista, se ele for reprovado por não saber a última técnica para uma ressecção pancreática?

Ao planejar uma avaliação, o professor deveria saber qual uso ele fará dos resultados de sua interpretação. No caso do Brasil, e de acordo com as Diretrizes Curriculares Nacionais para o curso de Medicina (DCN), deve-se formar um médico para “*atuar nos diferentes níveis de atendimento à saúde, com ênfase nos atendimentos primário e secundário*”, avaliar a competência de um aluno em Medicina, baseado no conhecimento da última técnica de ressecção cirúrgica do pâncreas dificilmente contribuirá para a validade de um teste.

Enquanto que para uma boa confiabilidade deve-se ter uma amostra de conteúdo que seja abrangente e representativa de todo o domínio do conhecimento abordado, para uma boa validade o conteúdo precisa ser significativo e adequado às necessidades futuras do aprendiz. Conteúdo é, portanto, uma das principais considerações na validade de um teste². Como validade é um conceito específico para um determinado uso ou interpretação, deveríamos avaliar somente conteúdos relevantes com a futura prática profissional.

Tabela 1.2 O quesito validade para avaliação diária.

Avaliação no dia a dia: Validade

- Seja qual for o tipo de teste, uma das melhores maneiras de agregar validade ao exame é contextualizando a avaliação e tornando-a a mais autêntica possível. Quanto maior a autenticidade profissional maior a validade.
- A busca pela validade tem aumentado o uso de testes para avaliar competência em ambientes reais, tais como o mini-CEX e a utilização de vídeo.
- No Brasil, o exame válido também é aquele que contempla as Diretrizes Curriculares Nacionais e o projeto político pedagógico da escola.

Outra consideração fundamental na construção da validade de um teste é saber o quanto o resultado de um exame poderá prever o desempenho futuro do aprendiz. Nesse sentido, a melhor estratégia é reproduzir, na avaliação, o máximo de autenticidade, e é por isso que avaliações que utilizem ambientes simulando cenários reais como o OSCE e a utilização de pacientes padronizados tornaram-se tão populares na educação médica.

AVALIAR COMO?

Para responder a essa pergunta deve-se levar em consideração os aspectos listados na Tabela 1.4.

Tabela 1.3 Conceitos fundamentais para avaliação acadêmica do curso de Medicina.

Conceitos Fundamentais

- Definindo objetivos instrucionais
- Viabilidade
- Medidas de aquisição
- Definindo critérios e padrões
- Avaliação critério-referenciada
- Avaliação normo-referenciada

Quais são os meus objetivos educacionais ao planejar minha avaliação?

Sem uma resposta clara para essa pergunta dificilmente podemos desenvolver um programa de sucesso. O que será mais importante para a formação de meu estudante? Conhecimento teórico? Desenvolvimento de habilidades que se revertam em competências? Aquisição de atitude profissional?

Nesse sentido um dos trabalhos de maior impacto na educação médica foi a taxonomia proposta por Miller para avaliação de competências clínicas⁵. Mesmo após duas décadas, a “*Pirâmide de Miller*” (Figura 1.1) continua sendo uma das referências mais citadas para definição de competências a serem exploradas durante a formação médica. A pirâmide tem sido útil para demonstrar ao professor de Medicina que seus objetivos educacionais não podem ficar restritos ao conhecimento teórico, e que o rumo da autenticidade profissional exige que o aluno saia do conhecimento meramente



Figura 1.1 ■ Pirâmide de Miller.

factual (*saber*), mas que também saiba aplicar esse conhecimento (*saber como*) e executá-lo de maneira prática *in vitro*, ou seja, em laboratórios e ambientes simulados (*demonstrar*) e, finalmente, aplicá-lo na vida real (*fazer*).

Familiarizando-se com a pirâmide de Miller, pode-se planejar o nível de competência adequado a determinado saber. Note que os dois primeiros níveis da pirâmide estão relacionados apenas com o domínio cognitivo, enquanto que os dois últimos exigem a aplicação de habilidades psicomotoras. Nos níveis em que somente o domínio cognitivo é exigido, pode-se avaliar utilizando testes tradicionais como teste escrito, de múltipla escolha ou exame oral. Nos dois últimos níveis, quando é necessário que o aprendiz demonstre habilidades e atitudes, a sala de aula tradicional já não oferece mais condições de avaliação. É necessário, então, que a avaliação mude de cenário, acontecendo em ambientes simulados ou em cenário real no ambiente de trabalho, como acontece no internato. A Figura 1.2 mostra as principais modalidades de avaliação no ensino médico em relação à Pirâmide de Miller.

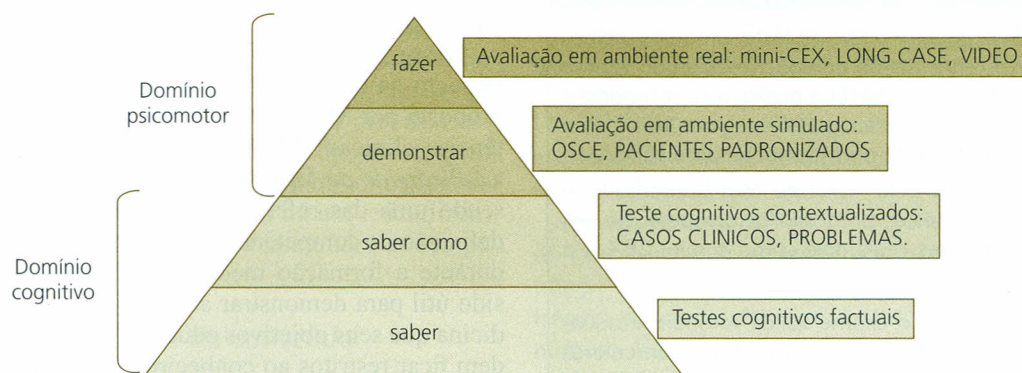


Figura 1.2 ■ Avaliações e pirâmide de Miller.

Viabilidade do teste

Escalar a pirâmide exige planejamento. Para planejar seu programa de avaliação são necessárias algumas considerações práticas. Chega o momento de pensar na **viabilidade** da avaliação. De nada adianta planejar uma avaliação baseada em ambiente simulado se não tivermos recursos para aquisição de equipamentos necessários. Outras considerações de ordem prática envolvem número de pessoal disponível e recursos para desenvolvimento docente. **Uma das missões do planejamento é tornar a avaliação viável de acordo com os recursos disponíveis.** Em um cenário de poucos recursos e muitos estudantes, frequentemente o teste escrito, particularmente o de múltipla escolha, acaba sendo uma opção se considerarmos custo/benefício, no entanto, é possível, através de um bom programa de desenvolvimento docente, aumentar a autenticidade de nossos exames com custos relativamente baixos, como ocorre com o minixame clínico, que aproveita as condições que existem em qualquer internato ou residência médica.

Escolha do teste apropriado

Analise os dois itens abaixo, extraídos do artigo cinco das Diretrizes Curriculares Nacionais (DCN) para os cursos de graduação em Medicina:

Art. 5º A formação do médico tem por objetivo dotar o profissional dos conhecimentos

requeridos para o exercício das competências e habilidades

VI – dominar os conteúdos básicos da natureza biológica subjacentes à prática médica crítica na interpretação e interpretação da natureza da medicina e na sua resolução

XIV – realizar procedimentos cirúrgicos indispensáveis ambulatorial e para o tratamento de urgências e emergências no ciclo biológico.

É óbvio que para o domínio do conhecimento científico cognitivo através de testes escritos é suficiente quando a avaliação cognitiva tem lugar na sala de aula.

O item XIV das diretrizes que o aprendiz seja capaz de executar procedimentos. Para uma avaliação de desempenho, a sala de aula não oferece condições de laboratórios de habilidades reais de trabalho, mais apropriados.

Subindo a pirâmide

Avaliação na sala de aula

No cenário da sala de aula tradicionalmente os testes, por várias razões, a mais importante é o custo/benefício, de melhor custo/benefício, um número relativo de testes em pouco tempo. Já discutimos que testes escritos são adequados e com um custo/benefício bem representativa possível realizar um teste de grau de fidedignidade. O número de testes no tipo de avaliação orais e de questões de múltipla escolha é muito maior do que os problemas de múltipla escolha devido a varia-

requeridos para o exercício das seguintes competências e habilidades específicas:

VI – dominar os conhecimentos científicos básicos da natureza biopsicossocio-ambiental subjacentes à prática médica e ter raciocínio crítico na interpretação dos dados, na identificação da natureza dos problemas da prática médica e na sua resolução;

XIV – realizar procedimentos clínicos e cirúrgicos indispensáveis para o atendimento ambulatorial e para o atendimento inicial das urgências e emergências em todas as fases do ciclo biológico.

É óbvio que para avaliação do domínio do conhecimento científico, uma avaliação cognitiva através de teste escrito pode ser suficiente quando bem construída. A avaliação cognitiva tem como cenário natural a sala de aula.

O item XIV das DCN, no entanto, exige que o aprendiz seja capaz de *realizar procedimentos*. Para uma avaliação adequada, o professor terá que planejar uma maneira de testar a desempenho. A sala de aula passa não mais a oferecer condições adequadas. A utilização de laboratórios de habilidades clínicas ou ambientes reais de trabalho passam a ser cenários mais apropriados.

Subindo a pirâmide!

Avaliação na sala de aula

No cenário da sala de aula os testes escritos são tradicionalmente os mais utilizados por inúmeras razões, a mais importante delas por ser o de melhor custo benefício: é possível testar um número relativamente grande de estudantes em pouco tempo e com recursos escassos. Já discutimos que se testarmos por um tempo adequado e com uma amostragem de conteúdo bem representativa, mesmo em sala de aula é possível realizar uma avaliação com um bom grau de fidedignidade.

O número de estudantes poderá influenciar no tipo de avaliação. A utilização de testes orais e de questões abertas demandará um tempo muito maior do que testes objetivos, além dos problemas de confiabilidade entre avaliadores devido variações nas interpretações. Não

é coincidência, portanto, que testes objetivos do tipo múltipla escolha sejam “*universalmente os mais amplamente utilizados*”⁷.

O teste de múltipla escolha tem sido muito criticado devido sua baixa validade no contexto da educação médica, no entanto, estudos têm demonstrado que é possível colocar autenticidade em todos os níveis da pirâmide³. Para isso, os testes escritos precisam ser construídos utilizando formatos de questões contextualizados, utilizando casos clínicos ou orientados para a resolução de problemas. A contextualização melhora a qualidade do teste escrito⁸.

Da sala de aula para o laboratório

Pode-se afirmar que a maior tendência na avaliação do estudante médico diz respeito aos testes de performance. Tradicionalmente, a avaliação de desempenho do estudante de Medicina se dava no cenário hospitalar, e isso gerava um problema de confiabilidade, uma vez que fica difícil padronizar uma avaliação em um cenário de constante mudança como uma enfermaria por exemplo. Uma solução para esse problema foi proposto por Harden e Gleeson, em 1979, com a criação do Exame Clínico Objetivo Estruturado, o popular *OSCE (Objective Structured Clinical Examination)*⁹, onde estudantes passam por um “rodízio” de estações, cada uma simulando um ambiente aonde uma habilidade específica é testada, com um avaliador diferente em cada estação. Todas as estações são construídas de maneira padronizada, de tal modo que todos os examinados são testados sob as mesmas condições. A popularidade desse exame pode ser demonstrada pela sua utilização mundial: a literatura demonstra que escolas médicas de mais de 50 países já utilizam o *OSCE* como parte de seu programa de avaliação¹⁰.

Uma vez que se tenha decidido utilizar o *OSCE* como modelo de avaliação, uma primeira pergunta a ser respondida diz respeito ao conteúdo que se deve avaliar. Evidentemente que devem ser utilizadas questões que exijam do candidato a demonstração de habilidades e atitudes na resolução de problemas clínicos.

Ao falarmos de avaliação de desempenho, cabe discutir como avaliar o grau de *expertise* ou de maestria em determinada habilidade. Este conceito é derivado de outro conceito fundamental em avaliação que é a avaliação critério – referenciada¹¹⁻¹². Na avaliação critério-referenciada se julga o desempenho de um estudante baseado em um **padrão mínimo** a partir do qual se atinge a meta desejada. Por exemplo, quando se testa a capacidade de um grupo de alunos de realizar uma intubação orotraqueal, o que se deseja é que o candidato realize o procedimento dentro da técnica adequada e com segurança. Alguns poderão fazer mais rápido ou mais habilmente que outros, mas todos que intubarem corretamente passarão. Não existe interesse do professor em hierarquizar quem é melhor através de notas, seu objetivo é garantir que todos saibam realizar o procedimento adequadamente. Em geral, testes critério-referenciados resultam em uma decisão de suficiente ou insuficiente (“*pass/fail*”) ao invés de nota.

As avaliações critério-referenciadas em geral são contrastadas com avaliações normo-referenciadas. Nessas últimas, o desempenho do aprendiz é medido através de notas, para poder comparar os resultados obtidos pelos diversos examinados. De uma maneira geral, a avaliação normo-referenciada só seria necessária para propósitos de seleção, uma tarefa desnecessária no ensino de graduação, já que o único momento em que a comparação é essencial é na seleção dos candidatos para a escola médica ou para um programa de pós-graduação como a residência médica.

Estações de OSCE deveriam obedecer a um modelo de critério-referenciamento, onde se espera que o estudante tenha um desempenho baseado em um padrão pré-estabelecido. Em exames de “alta aposta” podemos garantir a confiabilidade aumentando o número de especialistas envolvidos na definição dos padrões exigidos. Essa definição de padrão tem por objetivo estabelecer o “ponto de corte” a partir do qual o desempenho é adequado. Existem dois tipos principais de critérios:

- **Relativos:** quando se compara o desempenho dos candidatos. Um exemplo típico é uma prova de residência, onde se deverá selecionar, por exemplo, as maiores notas.

- **Absolutos:** este é, na verdade, o verdadeiro teste critério-referenciado, pois avalia-se o desempenho do avaliado baseado nos critérios estabelecidos. Um exemplo típico dessa técnica é o método de “grupos contrastados”, onde o ponto de corte é calculado a partir da análise da interseção da curva de escores do grupo suficiente com a dos escores do grupo insuficiente. Em testes de “alta aposta” têm sido utilizado métodos como o de Angoff e Ebel.

Embora tenham se tornado extremamente populares, as dificuldades logísticas tornam o OSCE uma realidade distante de muitas escolas, afinal, é necessário mobilizar um grande número de examinadores (pelo menos um em cada estação), contar com instalações físicas adequadas que permitam o rodízio de estudantes, além de realizar um preparo minucioso que vai desde o treinamento de pacientes padronizados, até a utilização de manequins. No Brasil, não só a escassez de recursos das escolas é um empecilho para a popularização do teste, mas também toda uma “cultura de avaliação” existente em nossas instituições¹³.

Da sala de aula para a beira do leito

Quanto maior autenticidade maior a validade de um exame. Nesse sentido, nenhum exame será mais válido do que aquele em que se avalia a capacidade de realizar tarefas reais em ambientes de trabalho. A avaliação do aluno na “beira do leito” é tão antiga quanto o ensino da Medicina, mas só nas últimas décadas ganhou atenção adequada para sua sistematização.

Cabe aqui analisarmos um pouco da história da educação médica para percebemos como a busca pela validade e a confiabilidade de um exame promoveram mudanças na maneira como avaliamos.

No início do século XX Flexner já preconizava que a única maneira de avaliar a proficiência do médico era através de *casos concretos*, a partir dos quais o estudante coletasse as informações necessárias e propusesse a solução adequada¹⁴. É fácil perceber, então, que examinar o estudante oferecendo a ele um paciente real, e solicitando que realize toda a cadeia de habilidades que envolve uma consulta (entrevista e coleta de dados,

exame e conduta daquele famoso re- populares casos lo o examinado fica um paciente real. dos casos longos lado, deixam a de tões de confiabili completa com um e, portanto o caso gem pobre com número de exam do conteúdo é at questão crucial nho¹⁵. Um estuda penho em um co desempenho exce

Uma solução desenvolvimento exame clínico, se porém mais breve encontros, maior e em cenários d teúdos diferente: “novo” exame f confiabilidade ac ginou o “mini e ca”, conhecido abordado em out diferencial do n tempo em que o “paciente real” de um *feedback* formativo à aval

ONDE AVALI

A avaliação pode lados e/ou no loc

A ten
curri

estim
avaliação de p

Ensin

na verdade, o verdadeiro preferenciado, pois a avaliação do avaliado baseado em selecionados. Um exemplo é o método de “grupos” e o ponto de corte é calculado a partir da análise da interseção de cada grupo suficiente com o grupo insuficiente. Em alguns casos, “grupos” têm sido utilizados por Angoff e Ebel.

ornado extremamente as dificuldades logísticas tornam o teste bastante de muitas escolhas. Mobilizar um grande número de alunos (pelo menos um em cada instalação física) em um rodízio de estudo para um preparo minucioso de pacientes e a utilização de manequins. A falta de recursos das instituições para a popularização de toda uma “cultura de avaliação” em nossas instituições¹³.

A beira do leito

de maior a validade do teste, nenhum exame que seja realizado em que se avalie as tarefas reais em ambientes reais. A avaliação do aluno na prática quanto ao ensino das últimas décadas ganhou uma sistematização.

um pouco da história para percebermos como a confiabilidade de um teste é avaliada na maneira como

XX Flexner já preferia a maneira de avaliar a prática através de *casos* reais o estudante com as competências necessárias e propunha¹⁴. É fácil perceber, portanto, que o estudante oferecendo solicitando que realidades que envolve a coleta de dados,

exame e conduta) tem suas raízes desde aquele famoso relatório, e originou os então populares casos longos (“*long cases*”) onde o examinado fica cerca de 50 minutos com um paciente real. De fato, a autenticidade dos casos longos é inquestionável; por outro lado, deixam a desejar com relação às questões de confiabilidade. Realizar uma tarefa completa com um paciente demanda tempo, e, portanto o caso longo teria uma amostragem pobre com relação ao conteúdo e ao número de examinadores. A especificidade do conteúdo é atualmente considerada uma questão crucial na avaliação de desempenho¹⁵. Um estudante que tenha baixo desempenho em um conteúdo específico pode ter desempenho excelente em outros.

Uma solução para esse problema foi o desenvolvimento de um novo modelo de exame clínico, semelhante aos casos longos, porém mais breve, com um maior número de encontros, maior número de examinadores e em cenários diferentes (e, portanto, conteúdos diferentes). Fica evidente que esse “novo” exame foi uma tentativa de agregar confiabilidade ao “antigo” caso longo, e originou o “mini exercício de avaliação clínica”, conhecido como “mini-CEX”¹⁶, que é abordado em outro capítulo deste livro. Um diferencial do mini-CEX é que ao mesmo tempo em que oferece uma avaliação com “paciente real” possibilita a oportunidade de um *feedback* imediato conferindo caráter formativo à avaliação¹⁷.

ONDE AVALIAR?

A avaliação pode ser feita em ambientes simulados e/ou no local de trabalho.

A tendência atual em desenvolvimento curricular em direção a um currículo baseado em competências tem estimulado uma atenção crescente na avaliação de performance no local de trabalho

Assessment in Medical Education

Ensinar não é transmitir conhecimento

Paulo Freire

Nossa jornada em direção ao ápice da Pirâmide de Miller termina no local de trabalho

A avaliação no local de trabalho parece ser a “nova” tendência no ensino médico, e com razão!

Cabe aqui um momento de acordo semântico: frequentemente temos observado críticas acaloradas à primeira menção da palavra “competência”. Tais críticos, com razão, apontam para o uso reducionista e com foco no mercado, quando a palavra é frequentemente usada. Fica aqui combinado que utilizamos o termo para designar o conjunto de conhecimentos, habilidades e atitudes que caracterizam a prática efetiva do médico.

Embora seja de nossa opinião que a avaliação no local de trabalho deve acontecer durante todo o ensino de graduação, tradicionalmente ela se torna um fato durante o internato e a residência médica.

Em nenhum outro local teremos maior autenticidade, o que contribui de maneira inigualável para sua validade. E mais: a riqueza do ambiente pode suprir o exame da amostragem necessária para a confiabilidade do teste.

Norcini e colaboradores¹⁸ propõem uma avaliação da *performance* no local de trabalho baseado em três processos:

- 1. avaliação do resultado do cuidado**, onde o principal critério é o resultado do tratamento. Este modelo reflete muito a prática médica no país de origem dos autores, e pode entrar em choque particularmente com a filosofia da prática médica em nosso país, que defende a Medicina como uma profissão de meios e não de resultados;
- 2. avaliação do processo do cuidado**, onde se exige do aprendiz que os meios necessários para um cuidado satisfatório tenham sido realizados corretamente;
- 3. avaliação do volume da prática** onde são registrados o número de procedimentos e atendimentos realizados.

De uma maneira geral podemos dizer que a avaliação no local de trabalho acontece de duas maneiras: 1 – registro e 2 – observação pelo preceptor (Tabela 1.3).

Tabela 1.3 Aquisição de Autonomia para Punção Lombar durante a Residência Médica.

Fases	Níveis de autonomia
Fase 1	Descrever adequadamente a técnica de punção lombar e suas possíveis complicações
Fase 2	Observar uma punção lombar realizada por especialista
Fase 3	Realizar pelo menos três procedimentos sob observação direta, com parecer favorável de especialista
Fase 4	Autonomia

REGISTRO

Diários (logbooks)

O registro em diário, comumente identificado pelo seu equivalente em língua inglesa como “*logbook*”, permite que o aprendiz registre todos os encontros e procedimentos significativos que teve em sua experiência, e é particularmente útil na residência médica, onde o nosso objetivo educacional é promover um grau crescente de autonomia (*expertise*) do aprendiz. Desta forma o professor deve determinar um conjunto de habilidades dispostas em ordem crescente de autonomia (cognição ⇒ observação ⇒ realizar sob supervisão ⇒ realizar de maneira autônoma), e determinar a quantidade de encontros ou experiências em cada fase. Na Tabela 1.3 apresentamos um modelo de aquisição de autonomia.

Nesse método, o tempo de formação fica na dependência da aquisição de autonomia nas habilidades necessárias para o desempenho profissional esperado. Na verdade, esse tópico desperta uma das mais importantes perguntas que se deve fazer no planejamento educacional de um currículo médico: *Qual o nível de autonomia/expertise que desejo de meu graduando ou residente ao final da formação?*

Acreditamos que esse deve ser o próximo “fórum” de discussão entre os diversos interessados na formação de médicos.

Portfólios

Em nossa instituição temos utilizado portfólios, principalmente no programa de integração com a comunidade. Portfólios, quando bem aplicados, permitem que se registre o “caminho cognitivo” do estudante em áreas de difícil avaliação, tais como profissionalismo e atitude.

Nesse registro que o aprendiz faz dos fatos significativos de seu aprendizado, é possível avaliar as dificuldades enfrentadas, as iniciativas tomadas para resolução de problemas, as atitudes diante de situações críticas, novos aprendizados e suas evidências, e até mesmo o grau de satisfação do paciente.

Portfólios trazem dois problemas inerentes: o primeiro de confiabilidade, já que registra a subjetividade do aprendiz que coloca em seu registro aquilo que deseja e acha significativo, além de apresentar problemas na confiabilidade entre avaliadores; o segundo, de viabilidade, uma vez que exige um tempo significativo para a correção.

OBSERVAÇÃO PELO PRECEPTOR

Diversos tipos de observações podem ser feitas, incluindo avaliação por vídeo ou observação direta de habilidades. De todos os modelos, o mini-CEX ganhou crescente popularidade justamente porque oferece uma forma tradicional de avaliação sob um novo formato estruturado, permitindo maior confiabilidade entre examinadores.

O cenário da educação médica mudou. O aluno já não depende mais de professores para a transmissão de fatos, e muitos alunos vão argumentar, com razão, que para assistir projeções de *slides* (muitas vezes sem a interação necessária entre estudante e professor) é mais produtivo ficar diante da tela do computador. No entanto, o papel do professor é insuperável na transmissão de experiências e para o ensino de habilidades e atitudes que somente o ambiente de trabalho pode oferecer.

AVALIAR QUANDO?

Ensinar exige querer bem aos educandos.

Paulo Freire

Existe um momento em que ele passa se por de educador. Nesse momento, ele precisa preparar conteúdos sobre a ter foco no aprendizado frequentemente não acontece.

É momento de levar o pacto educacional de especialistas poderão dizer que já é capaz de promover isso, o educador que não está especializará desprezando as ferramentas disponíveis.

Baseado em tudo que poderíamos dizer que uma provávelmente válida) já ter impacto educacional basta. Para ter impacto e ação, além de válida, precisa sobre a prática e, a partir disso criar consciência sobre o aprendizado com o objetivo *formance* futura. Aí então que promove progressivamente e, portanto, de grande

Cabe aqui abordarmos o mental em educação. O conteúdo que justamente permeia seu aprendizado e com habilidades, suas fraquezas e para suprir suas necessidades. Estamos falando de metas.

O termo *metacognição* toriedade a partir do trabalho de Flavell, que tentava explicar “*cognição e saber sobre o cognitivo*”²⁰. Esse conceito, que mente novo para aquele com Piaget, diz respeito ao diz de planejar intelectualdo, ou seja, ele tem consciência

Tabela 1.4 Avaliação for

Quando?
Atributos fundamentais
Estratégias

Existe um momento na vida do professor em que ele passa se preocupar com seu papel de educador. Nesse momento, ele para de despejar conteúdos sobre seus estudantes e passa a ter foco no aprendizado. Ensinar e aprender frequentemente não andam juntos!

É momento de levar em consideração o impacto educacional de suas atividades. Alguns céticos poderão dizer que a avaliação por si só já é capaz de promover aprendizado¹⁹, no entanto, o educador que não utilize todo seu potencial estará desprezando uma das mais poderosas ferramentas disponíveis na sua tarefa de educar.

Baseado em tudo que já foi discutido, poderíamos dizer que uma avaliação autêntica (e provavelmente válida) já seria capaz de produzir impacto educacional, no entanto, isso só não basta. Para ter impacto educacional uma avaliação, além de válida, precisa produzir reflexão sobre a prática e, a partir daí, permitir ao aprendiz criar consciência sobre as necessidades de aprendizado com o objetivo de melhorar a *performance* futura. Aí então temos uma avaliação que promove progressivamente melhora da prática e, portanto, de grande impacto educacional.

Cabe aqui abordarmos outro conceito fundamental em educação. O conceito de uma habilidade que justamente permite ao aluno monitorar seu aprendizado e com isso identificar suas fortalezas, suas fraquezas e saber traçar um roteiro para suprir suas necessidades de aprendizagem. Estamos falando de metacognição.

O termo *metacognição* ganhou alguma notoriedade a partir do trabalho pioneiro de John Flavell, que tentava explicar o termo como “*cognição e saber sobre um fenômeno cognitivo*”²⁰. Esse conceito, que pode não ser exatamente novo para aqueles mais familiarizados com Piaget, diz respeito à capacidade do aprendiz de planejar intelectualmente seu aprendizado, ou seja, ele tem consciência dos resultados

de suas práticas, percebe o que ainda falta para torná-las efetivas e é capaz de traçar uma estratégia para suprir essas necessidades. Este conceito pós-moderno à concepção socrática da inteligência, diz que não basta saber suas deficiências, mas também saber traçar uma estratégia intelectual para atingir a tarefa cognitiva.

Só existe uma maneira de garantir essa metacognição: garantir ao aluno um *feedback* efetivo sobre sua prática e permitir que ele revise o cenário do erro para melhorar a prática. Estamos então falando de **avaliação formativa**.

Avaliação formativa é o mais importante conceito em avaliação (pelo menos para o educador), sem o qual nenhuma das estratégias anteriormente abordadas terá impacto educacional efetivo. Muito se tem falado a respeito do assunto e, conseqüentemente, muitas definições confusas têm tornado complexo um assunto simples. Preferimos, por isso, a definição simples de Philippe Perrenoud: “*Proponho considerar como formativa toda prática de avaliação contínua que pretenda contribuir para melhorar as aprendizagens em curso*”. Desta forma, gostaríamos de reforçar a ideia de que, independente do tipo de avaliação, todo professor que avalia ao longo do curso, informando o estudante sobre os erros e acertos de sua *performance* e permitindo que o aluno revise esse cenário para melhorar a prática, poderá ser considerado formativo.

Avaliação formativa, em geral, se contrapõe à outra forma de avaliação conhecida como avaliação somativa que tipicamente ocorre ao final de um programa e é utilizada para saber se um candidato atingiu os objetivos esperados. É usada primariamente para fins de certificação e para notas finais de cursos². É fácil perceber que a avaliação somativa não tem a mesma “perspectiva” educacional. A Tabela 1.4 mostra algumas diferenças entre os dois tipos de avaliação.

Tabela 1.4 Avaliação formativa x avaliação somativa.

	Formativa	Somativa
Quando?	• Ao longo do curso	• Ao final do curso
Atributos fundamentais	• Validade • Impacto educacional	• Confiabilidade
Estratégias	• Garantir <i>feedback</i> • Permitir revisar o cenário do erro	• Garantir amostragem

utilizado portfólios, a de integração com quando bem aplicados registre o “caminho áreas de difícil avalismo e atitude.

rendiz faz dos fatos dizado, é possível rentadas, as iniciação de problemas, ões críticas, novos cias, e até mesmo o te.

problemas inerentes: e, já que registra a e coloca em seu re- a significante, além confiabilidade entre iabilidade, uma vez ivo para a correção.

RECEPTOR

ões podem ser fei- vídeo ou observa- todos os modelos, ente popularidade ma forma tradicio- ovo formato estru- onfiabilidade entre

médica mudou. O e professores para muitos alunos vão para assistir pro- es sem a interação professor) é mais a do computador. ssor é insuperável as e para o ensino e somente o am- cer.

m aos educandos.

Paulo Freire

Podemos então concluir que o papel da avaliação somativa é medir, enquanto que a avaliação formativa está comprometida com o aprendizado significativo (Tabela 1.5).

Cagne, ao aplicar os conceitos da ciência cognitiva à instrução, estabelece que o papel do professor seja:

1. chamar atenção sobre o assunto;
2. informar os resultados ao aprendiz e
3. estimular o retorno aos objetivos ainda não alcançados²¹. Chegamos, então, à estratégia fundamental da avaliação formativa, sem a qual não teremos sucesso: o *feedback*.

O *feedback* sozinho pode ser o mais poderoso facilitador de aprendizado na educação médica²² (Tabela 1.6). Não se trata simplesmente de informar resultados, isso frequentemente produz um comportamento defensivo do aprendiz sem promover reflexão. Na definição de Archer, o *feedback* efetivo é aquele no qual a *informação sobre a performance prévia é utilizada para promover desenvolvimento positivo e desejado*²³. Deve ser encarado com um diálogo entre o aprendiz e um professor treinado. Mais uma vez, portanto, reforçamos a necessidade de um programa de desenvolvimento docente que contemple esse treinamento.

CONCLUSÃO

Ensinar exige rigorosidade metódica, pesquisa e criticidade.

Paulo Freire

Finalmente cabe dividir com o leitor uma conclusão. Só existe um papel para o professor: o de educador. Nesse sentido nossa abordagem tem

Tabela 1.5 Avaliação formativa x Avaliação somativa.

Avaliação no dia a dia: Você é "formativo" ou "somativo"?
Independente do método de avaliação, se você avalia seus alunos desde o início do programa, discute a avaliação apontando pontos positivos e informando o que ainda precisa ser atingido, e finalmente se permite ao aluno visitar o cenário de suas necessidades, então você é "formativo"!

sido sempre o de colocar a avaliação a serviço do aprendizado, e não o contrário. Acreditamos que o primeiro passo em direção a essa visão acontece quando o professor começa a refletir sobre a validade de seu exame, e então faz a pergunta reveladora: *estou medindo aquilo que realmente importa para a futura prática de meu estudante?*

A revelação acontece quando então o professor mergulha na pesquisa educacional. A abundância de estudos no campo da educação médica aponta o caminho da utilização de novas metodologias, particularmente aquelas que agregam autenticidade profissional aos exames. Convidamos os professores a considerarem a educação médica com o *status* da ciência que ela realmente é, e partirem para o campo da experimentação científica e do desenvolvimento docente. Utilizem os momentos de exames de "baixa aposta" para introduzirem novas técnicas, analisem os dados levando em consideração a validade, a confiabilidade e o impacto educacional. Após consolidada a experiência, aprimorem a técnica nos exames de alta aposta, construindo a confiabilidade necessária. No caminho da experiência educacional na avaliação, o professor pode contar com um parceiro de todas as horas para ajudá-lo: o estudante de Medicina.

Aproveitando uma advertência de Machado de Assis, que por sua vez cita Shakespeare:

*Our doubts are traitors
And make us lose the good
we often might win,
y fearing to attempt.*

Tabela 1.6 Treinamento para feedback efetivo.

Avaliação no dia a dia: Treinamento para feedback efetivo
<ul style="list-style-type: none"> Existem diversas técnicas para <i>feedback</i> efetivo que podem ser utilizadas em um programa de desenvolvimento docente. Um modelo altamente efetivo e prático é o "preceptor minuto". O material pode ser obtido on line: One Minute Preceptor handout http://www.im.org/Resources/FacultyDev/GIMFDP/Documents/Strategies%20Tampa%20Sarkin.htm (Acesso em 06 Jan 2010)

Não há p
Este é u
em um país
médica com
como em in
Resta ag

REFERÊN

1. Wass V
cípios o
Editor.
and Tra
Radclif
2. Lunn R
sessmer
sey. Edi
3. Van der
professi
program
309-317
4. Wass V,
Medical
5. Miller C
compet
65(supp
6. Coles C
J Contin
7. Jackson
ment it
Guide.
cliffe P
8. Jozefow
Galbrai
quality
nations.
9. Harden
clinical
structur
Medica
10. Hart I.
Dent J,
guide fo
Londres
11. Swartz J
of vide
rating c

ar a avaliação a serviço do
ontrário. Acreditamos que
lreção a essa visão acon-
or começa a refletir sobre
e, e então faz a pergunta
ndo aquilo que realmente
rática de meu estudante?

ce quando então o pro-
esquisa educacional. A
s no campo da educa-
aminho da utilização de
particularmente aquelas
idade profissional aos
os professores a consi-
médica com o *status* da
ente é, e partirem para
entação científica e do
ente. Utilizem os mo-
“baixa aposta” para in-
cas, analisem os dados
ão a validade, a confia-
ucacional. Após conso-
primorem a técnica nos
construindo a confiabi-
aminho da experiência
ção, o professor pode
de todas as horas para
e Medicina.

divertência de Machado
z cita Shakespeare:

*Our doubts are traitors
make us lose the good
we often might win,
y fearing to attempt.*

para feedback efetivo.

dia a dia:
feedback efetivo

as para feedback efeti-
adas em um programa
cente. Um modelo al-
co é o “preceptor mi-

ptido on line: One Mi-
ut <http://www.im.org/GIMFDP/Documents/20Sarkin.htm> (Acesso

Não há porque não tentar!

Este é um convite à pesquisa educacional em um país com um acúmulo em educação médica como nenhum outro, tanto em volume como em inovações.

Resta agora registrarmos.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Wass V, Bowden R, Jackson N. The principles of assessment design. In: Jackson N. Editor. *Assessment in Medical Education and Training: A practical guide*. Oxford, Radcliff Publishing, 2007.
2. Lunn R, Grolund N. *Measurement and Assessment in Teaching*. Oitava ed. New Jersey. Editora Merrill/Prentice Hall, 1995.
3. Van der Vleuten C, Schuwirth L. Assessing professional competence: from methods to programmes. *Medical Education*. 2005; 39: 309-317.
4. Wass V, van der Vleuten C. The Long Case. *Medical Education*. 2004; 38: 1176-1180.
5. Miller G. The assessment of clinical skills/competence/performance. *Acad Med*. 1990; 65(suppl):S63-7.
6. Coles C. Developing professional judgement. *J Contin Educ Health Prof*. 2002;22 (1):3-10.
7. Jackson N, Jamieson A, Khan A. *Assessment in Medical Education: A Practical Guide*. Primeira edição, 2007. Oxon. Radcliffe Publishing Ltda.
8. Jozefowicz RF, Koeppen BM, Case SM, Galbraith R, Swanson DB, Glew RH. The quality of in-house medical school examinations. *Acad Med*. 2002;77 (2):156-61.
9. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical Education*. 1979; 13: 41-54.
10. Hart I. Objective clinical examinations. In: Dent J, Harden RM. Editores. *A practical guide for medical teachers*, segunda edição. Londres. Elsevier/Churchill Livingstone.
11. Swartz MH, Colliver JA *et al*. Global ratings of videotaped performance versus global rating of actions recorded on checklists: a criterion for performance assessment with standardized patients. *Acad Med*. 1999; 74:1028-32.
12. Friedman BDM. Principles of Assessment. In: Dent J, Harden RM. Editores. *A practical guide for medical teachers*, Segunda edição. Londres. Elsevier/Churchill Livingstone.
13. Troncon LE. Avaliação de habilidades clínicas: limitações à introdução de um exame clínico objetivo estruturado (“OSCE”) em uma escola médica brasileira tradicional. *Sao Paulo Med J*. 2004; vol.122 nº.1.
14. Flexner A. *Medical Education in the United States and Canada*. Bethesda, Maryland: Science and Health Publications, 1910.
15. Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ*. 1996;1:41-67.
16. Norcini JJ, Blank LL, Arnold GK, Kimball HR. The Mini-CEX (Clinical Evaluation Exercise): A preliminary investigation. *Ann Intern Med*. 1995; 123: 795-99.
17. Norcini JJ. The Mini Clinical Evaluation Exercise (mini-CEX). *The Clinical Teacher* 2005 | Volume 2 | No 1| (Disponível em www.theclinicalteacher.com).
18. Norcini JJ. Work based assessment. *British Medical Journal*. 2003; 326: 753-755.
19. Larsen DP, Butler AC, Rodiger HL III. Test-enhanced learning in medical education. *Med Educ*. 2008; 42:959-966.
20. Hacker D, Dunlosky J, Graesser A. *Metacognition in Educational Theory and Practice*. New Jersey: Editora Lawrence Erlbaum Associates, inc. 1998.
21. Cagne M. The conditions of Learning. In: Dent J. & Haren, R. Editores. *A Practical Guide for Medical Teachers*. Editora Elsevier, 2005; p. 191.
22. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians’ clinical performance. *Med Teach*. 2006;28 (2):117-28.
23. Archer J. State of the science in health professional education: effective feedback. *Med Educ*. 2010;44: 101-108.