

CAPÍTULO 13

—

# Avaliação com o uso de Checklists e Escalas de Avaliação Global

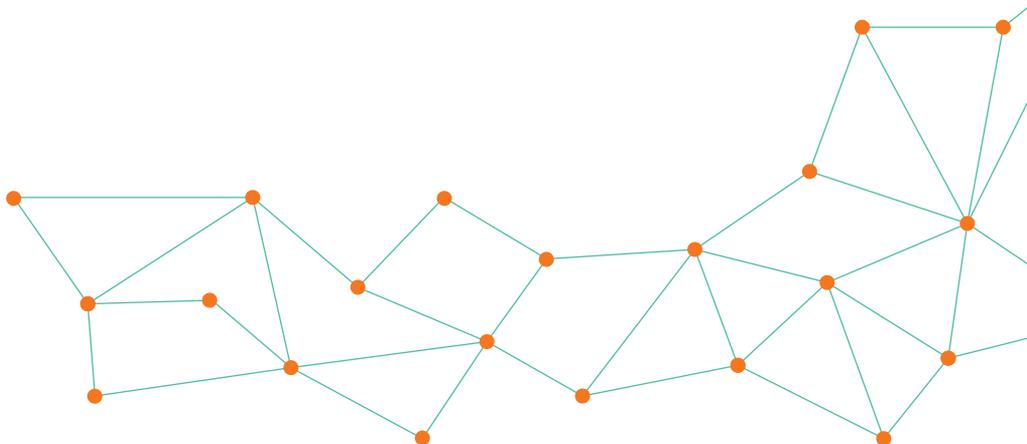


**Renato Soleiman Franco**

Psiquiatra  
Professor Adjunto – Curso  
de Medicina da PUCPR  
Coordenador do Programa de Residência  
Médica em Psiquiatria da Prefeitura  
Municipal de Curitiba, SMS/FEAS.

**Camila Ament Giuliani  
dos Santos Franco**

Médica de Família e Comunidade  
Professora Adjunta - Curso  
de Medicina - PUCPR  
Professora no Curso de Medicina  
nas Faculdades Pequeno  
Príncipe, Curitiba, PR.  
Preceptora responsável pela  
disciplina de Comunicação Clínica na  
Residência Médica em Medicina de  
Família e Comunidade - Residência  
Integrada, Curitiba, PR.



## 1. INTRODUÇÃO

Os checklists e as escalas globais são utilizados para avaliação em diversos contextos na simulação e podem estar presentes tanto em avaliações formativas quanto somativas. Ambos têm como objetivo oferecer um parâmetro objetivo do que é esperado que seja realizado. É preciso que sejam elaborados com bastante cuidado e atenção, uma vez que, muitas vezes, são a base para discutir os elementos de aprendizagem tanto com professores (ou tutores) quanto com os trainees. De forma geral, checklists consistem em itens detalhados de desempenho, como, por exemplo, “lavou as mãos”, “introduz-se para o paciente” ou “mede a pressão arterial com um manguito de tamanho apropriado.” Esses itens normalmente são qualificados em escalas binárias, como “concluído” ou “não concluído”, “realizado

ou não realizado”, escalas com gradação, escalas tipo Likert, entre outras (SWANSON; VAN DER VLEUTEN, 2013).

Assim, os checklists avaliam o desempenho de cada elemento observável de determinada competência. Por outro lado, a avaliação global é uma avaliação geral ou subdividida em subdomínios de competências (ILGEN et al., 2015). Por exemplo: o exame físico e as habilidades de comunicação podem ser dois subdomínios diferentes. Ao utilizar o checklist, o avaliador irá definir cada comportamento observável no exame físico e nas habilidades de comunicação. Já na avaliação global, haverá uma impressão geral sobre o desempenho quanto às habilidades de comunicação e quanto ao exame físico.

## 2. PREPARAÇÃO DE CHECKLISTS E ESCALAS GLOBAIS

Cada item, nas escalas globais, compreende uma série de comportamentos, habilidades e atitudes. Dessa forma, as ferramentas que utilizam Escalas de Avaliação Global (*do inglês – Global Rate Scale – GRS*) apresentam número bem menor de itens – em torno de 6, diferente dos checklists, em torno de 19 (ILGEN et al., 2015). As GRSs podem ser dicotômicas, descritivas (o avaliador descreve a sua avaliação) ou escalonadas (de baixa até alta performance; novato/intermediário/expert; afirmativas, com – concordo totalmente até discordo totalmente – ou até mesmo entre 0 e 100) (MA et al., 2014).

Os checklists e as GRSs podem ser utilizados para as mais variadas habilidades/competências, como, por exemplo, habilidades de comunicação, exames clínicos ou procedimentos cirúrgicos (VAN HOVE et al., 2010; CÖMERT et al., 2016). A escolha dos itens em um checklist deve estar de acordo com o objetivo de aprendizagem e o cenário de ensino (o quê, para quem, para o quê, por quem e como) (Tabela 1 e 2):

## 3. GRADAÇÃO DO CHECKLIST

A partir da definição do objetivo de aprendizagem e do cenário de ensino, é possível definir com maior clareza como será o checklist. Há uma maleabilidade grande quanto ao checklist. Por exemplo, em uma revisão sistemática sobre cateterismo venoso central, os autores identificaram o uso de checklists

que variam entre 1 e 63 itens (MA et al., 2014). Nessa mesma revisão, os autores observaram que os checklists eram em sua maioria binários, ou seja, do tipo suficiente/insuficiente, realizou (sim/não), entre outros. Mas também observaram escalas do tipo Likert, por exemplo, com gradação entre 1 e 4.

No caso das escalas do tipo Likert, é possível transcrever os objetivos de aprendizagem como uma afirmativa e graduar o nível de concordância do avaliador. Por exemplo: “O estudante realizou de forma adequada a palpação do pulso radial” – frente a isso, o avaliador pode graduar: concordo totalmente (1), concordo parcialmente (2), nem concordo ou discordo (3), discordo parcialmente (4) e discordo totalmente (5). Alguns estudos incluem uma categoria de gradação, como o “não realizado”, para elementos não observados pelo examinador. Apesar de haver inúmeras possibilidades, cada uma delas terá pontos positivos e negativos.

Entre os tipos de gradação de checklist, o modelo binário demonstra maior confiabilidade entre examinadores e é mais fácil de ser reproduzido. No entanto, traz uma diferenciação menor entre os participantes e capta com menor precisão diferentes níveis de habilidades e competências. Muitas vezes, os avaliadores gostariam de classificar diferentes níveis de expertise dentro da realização da competência ou habilidade proposta e não conseguem alcançar tal classificação com modelos binários. Não há um modelo certo e outro errado, mas deve-se ter noção de que, ao escolher um modelo de gradação no checklist, um grande passo está sendo tomado. Por exemplo, em uma avaliação que tem como objetivo uma maior reprodutibilidade, esca-

**Tabela 1** - Objetivos de aprendizagem para elaboração do checklist e da GRS

O que será ensinado?	Objetivo da atividade – Quais competências a serem desenvolvidas?
Para quem será ensinado?	Sujeitos – Qual o nível de treinamento/especialização pretendido?
Para o quê (motivo)?	Motivo do uso do Checklist – Qual o propósito? O checklist será um apoio para avaliação formativa, somativa, com ou sem feedback?
Por quem será realizado o checklist?	O checklist pode ser completado por professores, preceptores, pacientes, pacientes simulados, monitores, pares ou pelo próprio aprendiz (autoaplicáveis). É possível que haja mais de um avaliador da mesma categoria ou de categorias diferentes.
Como será ensinado?	Tempo – Qual a duração da atividade?
	Local – Qual contexto (simulação, OSCE, role play, entre outros) e recursos disponíveis?

Dos autores, Renato Soleiman Franco e Camila Ament Giuliani dos Santos Franco.

**Tabela 2** - Roteiro preparatório para elaboração de Checklist e GRS

Preparação para realização do Checklist	
1) Objetivo da atividade	
Objetivo geral – Nessa atividade, o aluno deverá ser capaz de:	
Domínios envolvidos:	Habilidades/Competências específicas
( ) Habilidades de comunicação	
( ) Exame físico	
( ) Raciocínio Clínico	
( ) Procedimentos	
( ) Outro - Qual?	
2) Sujeitos	
Número de sujeitos envolvidos:	Trabalho em equipe: ( ) sim ( ) não
Categorias profissionais:	
Grau de especialização:	( ) estudante/graduação ( ) residentes ( ) especialistas
( ) Outro grau - Qual?	
3) Motivo(s) do uso do Checklist:	
( ) Avaliação formativa	( ) Avaliação somativa
( ) Feedback	( ) Processo de seleção
( ) Suporte ao ensino	( ) Outro motivo - Qual?
4) Por quem será realizado o checklist?	
( ) pares	( ) pacientes simulados
( ) professores	( ) preceptores
( ) outros - quem?	
5) Ambiente	
Qual contexto: ( ) simulação, ( ) OSCE, ( ) role play ( ) outros - Quais?	
Recursos disponíveis/utilizados - Quais?	
Tempo:	

Referência - Camila Ament Giuliani dos Santos Franco e Renato Soleiman Franco - baseado no processo de validação dos OSCEs do artigo: Franco et al. (2015).

las binárias podem ser melhores; já para utilizar a escala para auxiliar no feedback, modelos com mais itens de gradação podem ser mais apropriados. No entanto, isso não é uma regra universal e o professor/facilitador deve ponderar riscos/benefícios e sempre reavaliar se sua escolha tem alcançado seus objetivos.

#### 4. AVALIAÇÃO DE HABILIDADES/COMPETÊNCIAS

Utilizamos o acrônimo SMART (simples, mensurável, atingível, realista e tempo adequado) para definir os itens a serem avaliados no checklist e nas GRSs (CHATTERJEE; CORRAL, 2017). Os itens podem ser: idênticos/semelhantes a checklists/GRS já

validados, adaptados a partir de modelos teóricos ou teórico-práticos consagrados ou definidos a partir da necessidade da atividade que será realizada.

Os itens devem estar dispostos em uma sequência clara que facilite a avaliação, não devem ser muito longos e precisam ser claros. Um exemplo de item longo seria: realizou a palpação abdominal com as manobras corretas para percussão e palpação hepáticas com ênfase no volume do fígado. Ao ter que marcar sim ou não, o avaliador ficará muito confuso. Caso forem realizadas algumas manobras e outras não, ou então, se não forem realizadas as manobras corretas, mas for avaliado o tamanho de uma forma não convencional, também haverá dificuldade para o avaliador. A presença de muitas habilidades ou procedimentos em um mesmo item deve ser cuidadosamente avaliada pelos professores/facilitadores.

No entanto, se a presença de todos esses elementos for essencial e for necessária uma associação entre eles, justifica-se a presença em conjunto. Contudo, devem-se considerar os riscos e benefícios desse modelo de item mais longo e complexo. Nesse caso, se o item fosse: Realizou a hepatimetria utilizando a manobra X, haveria maior clareza, objetividade e menor dúvida na qualificação. Um exercício interessante é buscar em cada área de interesse escalas e checklists já utilizados, mas é igualmente prudente adaptá-los aos cenários de simulação e objetivos das atividades a serem realizadas.

Muitas vezes, por dificuldades quanto à disponibilidade de simuladores, professores, salas ou até mesmo porque é importante que competências diversas sejam envolvidas, os checklists podem se tornar extensos e envolver diversos domínios de competências ou diferentes habilidades. Isso é possível de ser realizado, mas deve-se estar atento para o tempo disponibilizado para os cenários, assim como com a expertise do examinador. Quanto maior o número de itens no checklist, maiores a complexidade e a exigência para o examinador.

É importante lembrar que estações que avaliam habilidades muito específicas exigem menos itens. No exemplo acima, o objetivo poderia estar relacionado com “demonstra ser capaz de realizar a hepatimetria”. Agora, se o cenário é um ambiente de trauma, que envolve desde manejo do ambiente, habilidades de comunicação, habilidades ligadas a manobras de reanimação cardiopulmonar e raciocínio clínico, certamente o checklist será mais extenso e, conseqüentemente, o tempo para realizar a estação também). É importante destacar que

o checklist deve contemplar todos os domínios de competência a serem avaliados.

Além de utilizar modelos validados, é possível construir checklists e escalas globais. Ao elaborar um modelo para a avaliação, é possível enfatizar determinada competência ou habilidade a ser desenvolvida, como, por exemplo, quanto às competências de comunicação (Figura 1).

## 5. IMPORTÂNCIA DE VALIDAR CHECKLIST E GRS

A primeira etapa na construção do checklist/GRS é discutir o seu objetivo de aprendizagem e cenário de ensino (o quê, para quem, para quê, por quem e como); subsequente a isso, é importante realizar um estudo aprofundado do tema a ser avaliado pelo checklist/GRS e decidir se:

- A) aquela habilidade e/ou competência é passível de ser mensurada a partir do checklist/GRS;
- B) existem checklists/GRS que sejam referências já validadas – em caso positivo, para qual contexto, ou se o checklist/GRS será elaborado de acordo com as necessidades dos objetivos do cenário – baseados ou não em modelos prévios, e
- C) realizar um piloto com o uso do checklist/GRS.

A validação do conteúdo e o consenso entre os avaliadores são elementos essenciais para construção do checklist e da GRS. Existem várias formas de realizar essa validação, mas o objetivo é ter um checklist e uma GRS de boa qualidade, e que sejam confiáveis. Se possível, incluir nessa atividade especialistas na área – para validar o conteúdo – e não especialistas. **Os não especialistas** valorizam a forma e **apresentam uma impressão global**, assim como, quando o cenário for desenvolvido para residentes ou estudantes, auxiliam muito **na categorização dos graus de dificuldade**. Geralmente, as pessoas envolvidas na validação do checklist/GRS recebem antes a descrição do cenário, os objetivos e os checklists/GRS, e podem assim contribuir com apontamentos e sugestões para uma reunião de consenso. A reunião de consenso e validação pode ser realizada a partir da gravação de um vídeo e da solicitação de preenchimento pelos avaliadores, ou mesmo em ambiente simulado em tempo real. Na reunião de consenso, em geral, são apresentados o objetivo de aprendizagem e o cenário. A partir disso, os avaliadores discutem (Tabela 3):

- 1) Existem algumas habilidades/competências que poderiam ser incluídas e que não estejam representadas nesses itens?

### Checklist Estação 1 - Suzana

Nº identificação Aluno: \_\_\_\_\_ Avaliador (professor): \_\_\_\_\_

Competência Apresentada	Não	Sim
Cumprimenta o paciente e pergunta o seu nome		
Identifica-se (nome e função)		
Demonstra respeito e empatia		
Inicia a entrevista com pergunta aberta		
Identifica e confirma o motivo da consulta (pergunta o motivo da consulta e se tem mais alguma outra preocupação)		
Estimula que o paciente conte a sua história (por exemplo: pergunta aberta, silêncio, linguagem não verbal, facilitadores...)		
Explora as preocupações do paciente sobre o problema [Considera além das queixas e sintomas ( <i>disease</i> ), levando em conta as preocupações ( <i>illness</i> )]		
Encoraja o paciente a expressar seus sentimentos/emoções		
Demonstra linguagem não verbal apropriada (tom de voz, postura, olhar...)		

De uma forma geral, como avalia o desempenho do estudante:

0 - 10 - 20	21 - 30 - 40	41 - 50 - 60	61 - 70 - 80	81 - 90 - 100
Muito ruim	Ruim	Regular	Bom	Muito bom

No geral, esse estudante comunicou-se adequadamente com o paciente simulado:

( ) Sim ( ) Não

Observações:

**Figura 1** - Modelo utilizado pelos autores para avaliação de comunicação clínica

Modelo de checklist e GRS elaborado por Camila Ament Giuliani dos Santos, baseado no Guia Calgary-Cambridge de Comunicação.

**Tabela 3** - Roteiro para validação do conteúdo do Checklist ou GRS

Nome - Avaliador(a):
Checklist (Exemplo): Gradação de cada item: sim/não - Nessa estação, o estudante:
Cumprimentou a paciente
Disse seu nome e função
Iniciou a entrevista com perguntas abertas
Investigou se havia ideação suicida
Investigou se havia planejamento suicida
Perguntou sobre suporte familiar
Fez perguntas que possibilitassem avaliar a afetividade da paciente
Fez perguntas que possibilitassem avaliar a sensopercepção
Pediu que a paciente chamasse por ajuda
Solicitou que a paciente esperasse, acompanhada de um membro da equipe, até que seu/sua responsável chegasse

Referência - Camila Ament Giuliani dos Santos Franco e Renato Soleiman Franco - baseado no processo de validação dos OSCEs do artigo: Franco et al. (2015).

Tabela 3 - Continuação...

Deixou-se disponível e certificou que a paciente não tinha mais nada a dizer
Avalie de forma Global: Investigação do Risco de Suicídio foi adequada?
Avalie de forma Global: A conduta frente ao Risco de Suicídio foi adequada?
Avalie de forma Global: Houve uma condução adequada do caso, de acordo com as diretrizes nacionais para acolhimento dos (as) pacientes com risco de suicídio?
Favor realizar uma análise do checklist abaixo e responder:
1) Existem algumas habilidades/competências que poderiam ser incluídas nesse checklist e não estão representadas?
2) Quais itens geraram dúvida no preenchimento?
3) Há itens extensos que precisariam ser desmembrados?
4) Foi possível avaliar o objetivo do cenário com esse checklist?
5) O nível exigido é o adequado para os sujeitos em treinamento?
6) Como foi o tempo para o preenchimento?
7) Sugestões

Referência - Camila Ament Giuliani dos Santos Franco e Renato Soleiman Franco - baseado no processo de validação dos OSCEs do artigo: Franco et al. (2015).

- 2) Quais itens geraram dúvida no preenchimento?
- 3) Há itens extensos que precisariam ser desmembrados?
- 4) Foi possível avaliar o objetivo do cenário com esse checklist?
- 5) O nível exigido é o adequado para os sujeitos em treinamento?
- 6) Como foi o tempo para o preenchimento?
- 7) Após isso, faz-se uma checagem quanto à concordância (sugestão - não realizar a checagem da concordância antes de avaliar o construto de uma forma global, como apresentado nos itens acima).
- 8) Após a correção, pode-se seguir uma nova rodada de correções até que o checklist/GRS seja considerado adequado.

Em avaliações que exigem alta fidelidade, como exames finais ou concursos, pode-se realizar um teste-piloto. No teste-piloto, além da reunião de consenso e dos passos realizados anteriormente, programa-se com um grupo de cerca de 5 a 10 sujeitos (ou maior, conforme necessidade) a realização das estações/cenários com o preenchimento dos checklists/GRS. Como, geralmente, o número de participantes no piloto é pequeno, uma análise qualitativa dos preenchimentos é mais apropriada.

## 6. CONSIDERAÇÕES FINAIS

Apesar de o Checklist e a GRS serem modelos de escalas diferentes, a correlação entre eles é boa. A confiabilidade entre os avaliadores para ambos os tipos de escala era similarmente elevada. As GRSs apresentam maior confiabilidade entre os itens e entre as estações. Os valores de confiabilidade para ambos os modelos são menores para habilidades menos objetivas, como as de comunicação. Geralmente, os checklists são validados com base em opiniões de especialistas, já as GRS têm maior fundamento em modelos teóricos/referenciais (ILGEN et al., 2015). Um exemplo é a Avaliação Objetiva Estruturada de Habilidades Técnicas (OSATS), que tem sido modificada para diferentes propósitos, como descrito em um estudo Brasileiro por Barreto et al. (2019).

Há autores que consideram os checklists melhores para discriminar eventuais dificuldades do estudante em determinados comportamentos/habilidades, além de possibilitarem um feedback mais objetivo (ILGEN et al., 2015). Por outro lado, no preenchimento da GRS, nota-se uma correlação maior com o objetivo da estação. O foco não é na observação de cada passo ou cada habilidade, mas sim em alcançar o resultado de uma forma global. Não ra-

ramente, os estudantes conseguem fazer isso por vias não previamente determinadas e que, portanto, não estariam no checklist (WALZAK et al., 2015). Por outro lado, tem sido mais fácil o balizamento dos checklists do que das avaliações globais, lembrando que avaliadores menos experientes preferem os checklists (SEO; THOMAS, USPAL, 2019).

Fica claro que GRS e checklists têm seus pontos positivos e negativos. Na prática, sugerimos que

sejam inseridos, após os itens dos checklists, itens de avaliação global. O essencial é alinhar o modelo de instrumento desenvolvido com os objetivos de aprendizagem e com as razões pelas quais decidiu utilizar o instrumento (por exemplo, para feedback, para avaliação de um caso longo, para um OSCE, entre outros). A partir disso, podem-se fazer avaliações e análises psicométricas em testes-piloto e após seu uso.

## REFERÊNCIAS

- BARRETO, R. M. S. et al. Validação de um Programa de Treinamento Simulado de Habilidades Laparoscópicas por Residentes de Cirurgia. **Revista Brasileira de Educação Médica**, Brasília, v. 43, n. 2, p. 106-113, 2019. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-55022019000200106&lng=pt&tlng=pt](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-55022019000200106&lng=pt&tlng=pt)>. Acesso em: 7 dez. 2018.
- CHATTERJEE, D.; CORRAL, J. How to write well-defined learning objectives. *The Journal of Education in Perioperative Medicine*: JEPM, Richmond, v. 19, n. 4, p. E610, 2017. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/29766034>>. Acesso em: 7 dez. 2018.
- CÖMERT, M. et al. Assessing communication skills of medical students in objective structured clinical examinations (OSCE) - A Systematic Review of Rating Scales. **PLoS One**, San Francisco, v. 11, n. 3, p. e0152717, 2016. Disponível em: <<http://dx.plos.org/10.1371/journal.pone.0152717>>. Acesso em: 7 dez. 2018.
- FRANCO, C. A. G. S. et al. OSCE para competências de comunicação clínica e profissionalismo: relato de experiência e meta-avaliação. **Revista Brasileira de Educação Médica**, Brasília, v. 39, n. 3, p. 433-441, 2015.
- ILGEN, J. S. et al. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. **Medical Education**, Oxford, v. 49, n. 2, p. 161-173, 2015.
- MA, I. W. et al. Measuring competence in central venous catheterization: a systematic-review. **SpringerPlus**, USA, v. 3, n. 1, p. 33, 2014. Disponível em: <<https://springerplus.springeropen.com/articles/10.1186/2193-1801-3-33>>. Acesso em: 7 dez. 2018.
- SEO, S.; THOMAS, A.; USPAL, N. A global rating scale and checklist instrument for pediatric laceration repair. **MedEdPORTAL – The Journal of Teaching and Learning Resources**, Washington, v. 15, p. 10806, 2019. Disponível em: <<https://www.mededportal.org/publication/10806>>. Acesso em: 7 dez. 2018.
- SWANSON, D. B.; VAN DER VLEUTEN, C. P. M. Assessment of clinical skills with standardized patients: state of the art revisited. **Teaching and Learning in Medicine**, Hillsdale, v. 25, p. S17-S25, 2013. Supplement 1.
- VAN HOVE, P. D. et al. Objective assessment of technical surgical skills. **British Journal of Surgery**, Bristol, v. 97, n. 7, p. 972-987, 2010.
- WALZAK, A. et al. Diagnosing technical competence in six bedside procedures. **Academic Medicine**, Philadelphia, v. 90, n. 8, p. 1100-1108, 2015. Disponível em: <<http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landinpage&an=00001888-201508000-00027>>. Acesso em: 7 dez. 2018.