

Avaliação do raciocínio clínico de médicos e estudantes de Medicina: uma revisão integrativa

Evaluation of clinical reasoning in physicians and Medical students: an integrative review

Maurício Prätzel Ellwanger¹ 

ellwangemp@gmail.com

Fernando Tureck¹ 

fernandotureck@gmail.com

RESUMO

Introdução: O desenvolvimento do raciocínio clínico é crucial para estudantes e médicos, pois permite uma prática médica de qualidade e investigação racional de casos. Esse processo combina raciocínio intuitivo e analítico, adaptando-se à complexidade dos casos. Durante a formação médica, os alunos passam por fases de elaboração de hipóteses diagnósticas até o desenvolvimento de *scripts* das doenças. Com o avanço tecnológico na medicina, compreender como o raciocínio clínico se adapta é essencial para garantir cuidados de saúde de qualidade.

Objetivo: Este estudo visa analisar práticas de avaliação do raciocínio clínico, identificando lacunas e áreas de melhoria nos métodos avaliativos para promover currículos mais eficazes.

Método: Este estudo revisa as principais formas de avaliação do raciocínio clínico em médicos e estudantes de Medicina. Utilizando uma metodologia de revisão integrativa, foram selecionados artigos e ensaios publicados nos últimos 20 anos, em qualquer idioma, que abordavam comparações ou discussões sobre instrumentos de avaliação.

Resultado: Do total de 6.150 trabalhos inicialmente identificados, após a eliminação de duplicatas e a avaliação dos títulos e resumos, selecionaram-se 244 artigos que foram submetidos à leitura completa. Na revisão, mantiveram-se 122 artigos. Os métodos mais utilizados são questões de múltipla escolha (QME), teste de concordância de *scripts* (TCS) e exame clínico objetivo estruturado (OSCE), e mais de dez formatos de avaliação foram encontrados e discutidos.

Conclusão: O instrumento de avaliação, além de poder providenciar notas objetivas para os alunos, molda o comportamento e, portanto, é fundamental para a formação cada vez melhor de médicos. Inúmeros métodos têm sido pesquisados no momento, e outros, como o TCS, estão sendo implementados em larga escala. Este estudo contribui para o esforço de compilar todos os dados existentes da literatura e descobrir quais métodos, no agregado, são considerados mais confiáveis e válidos.

Palavras-chave: Raciocínio Clínico; Educação Médica; Avaliação; Tomada de Decisão Clínica; Revisão.

ABSTRACT

Introduction: The development of clinical reasoning is crucial for medical students and practitioners, as it enables quality medical practice and rational case investigation. This process combines intuitive and analytical reasoning, adapting to the complexity of cases. During medical training, students progress from formulating diagnostic hypotheses to developing "scripts" of diseases. With the technological advancements in medicine, understanding how clinical reasoning adapts is essential to ensuring high-quality healthcare.

Objective: This study aims to analyze clinical reasoning assessment practices, identifying gaps and areas for improvement in evaluation methods to promote more effective curricula.

Method: This study reviews the primary methods of assessing clinical reasoning in doctors and medical students. Using an integrative review methodology, articles and essays published in the last twenty years, in any language, that discussed or compared assessment tools were selected.

Result: Of a total of 6,150 initially identified studies, after removing duplicates and evaluating titles and abstracts, 244 were selected. These articles underwent full reading, and a total of 122 articles were included in the review. The most commonly used methods are multiple-choice questions (MCQs), script concordance testing (SCT), and objective structured clinical examinations (OSCE), with over ten different assessment formats found and discussed.

Conclusion: Assessment instruments, besides providing objective grades for students, shape behavior and are thus fundamental for the increasingly better training of doctors. Numerous methods are currently being researched, while others, such as the script concordance test, are being widely implemented. This study contributes to the effort of compiling all existing literature data and discovering which methods, in aggregate, are considered the most reliable and valid.

Keywords: Clinical Reasoning; Education, Medical; Anthropometry; Clinical Decision-Making; Review.

¹ Universidade do Contestado, Marombá, Santa Catarina, Brasil.

Editora-chefe: Rosiane Viana Zuza Diniz.

Editor associado: Mauricio Peixoto.

Recebido em 16/04/24; Aceito em 04/04/25.

Avaliado pelo processo de double blind review.

INTRODUÇÃO

O raciocínio clínico é um conceito complexo, definido como o processo pelo qual médicos integram teoria e prática para formular hipóteses diagnósticas e planejar intervenções. Esse processo envolve coleta de dados, revisão de hipóteses e adaptação contínua até atingir confiança na decisão. O raciocínio clínico é influenciado pelo contexto, pela experiência e por fatores institucionais, tornando a tomada de decisão complexa e incerta^{1,2}.

Avaliar o raciocínio clínico na educação médica é desafiador. Métodos estruturados são essenciais para formar médicos aptos a tomar decisões fundamentadas. Daniel et al.¹³⁸, em sua revisão de escopo, exploraram uma ampla variedade de estratégias de avaliação aplicadas a médicos e estudantes de Medicina, destacando desde simulações até avaliações práticas de desempenho. Este trabalho visa complementar essa abordagem, com foco na seleção rigorosa de estudos com alto controle metodológico, a fim de aprofundar a análise sobre como essas ferramentas impactam a formação clínica.

Os erros no raciocínio clínico podem levar a diagnósticos errados, tratamentos inadequados e custos elevados para o sistema de saúde³. As estatísticas mostram que a taxa de erro diagnóstico em emergências médicas, por exemplo, pode chegar a 10%-15% e que muitos dos erros são de origem cognitiva, envolvendo vieses como o fechamento prematuro (parar de buscar por novas hipóteses uma vez fechado o diagnóstico) e a busca por confirmação⁴. O raciocínio clínico pode ser entendido a partir de modelos de duplo processamento: o sistema 1 (raciocínio intuitivo e não analítico) é rápido e se baseia em padrões, e o sistema 2 (raciocínio analítico) é mais lento, deliberado e lógico³. O raciocínio clínico intuitivo pode ser vantajoso em situações de alta familiaridade com o caso clínico, mas também está suscetível a vieses cognitivos, como a disponibilidade e a representatividade⁴. Já o raciocínio analítico, embora mais exaustivo, pode ser limitado pela capacidade de memória de trabalho, gerando erros devido ao excesso de informações³. Ambos os tipos de raciocínio estão presentes na prática clínica, e estratégias educacionais que promovem a combinação dos dois processos têm mostrado benefícios moderados na redução de erros³.

Dentro desse contexto, o raciocínio clínico emerge como fundamental para lidar eficazmente com os erros diagnósticos. Por conta do papel crucial do raciocínio clínico, as instituições de pesquisa e ciência reconhecidas internacionalmente têm defendido melhorias tanto no ensino quanto na avaliação desse tipo de raciocínio ao corroborarem a importância dele na educação médica. No entanto, apesar desse reconhecimento, deficiências na compreensão dos conceitos de raciocínio clínico persistem entre muitos estudantes de Medicina e residentes,

destacando possíveis lacunas tanto na graduação quanto na pós-graduação em Medicina⁵.

A pirâmide de Miller descreve quatro níveis de competência clínica: “sabe”, “sabe como”, “mostra como” e “faz”, destacando a progressão do conhecimento teórico para a prática. As avaliações devem acompanhar essa evolução, testando desde o conhecimento factual até a aplicação em cenários reais. Além disso, revisões recentes, como a de Cruess et al.⁶, propuseram a inclusão de um quinto nível, “é”, que reflete a identidade profissional do médico, destacando que ser médico envolve não apenas “fazer”, mas também incorporar as atitudes, os valores e os comportamentos que definem a prática médica. Essa adição à pirâmide ressalta a importância da formação da identidade profissional, um processo que vai além do simples desempenho técnico, abrangendo a internalização dos valores éticos e morais da profissão. A evolução das habilidades de raciocínio clínico na educação médica destaca a importância do equilíbrio entre o treinamento formal e a experiência prática, reconhecendo o valor do conhecimento experiencial ao lado dos currículos tradicionais⁷⁻⁹.

É importante ressaltar que um dos objetivos principais de uma avaliação nesse contexto de ensino-aprendizagem é promover estímulos para desenvolver certos comportamentos nos alunos, de forma que com esses hábitos formados eles sejam bons médicos, e esse processo começa tanto na sala de aula no começo de uma graduação quanto no processo de educação continuada que ocorre após a formatura até a aposentadoria. De acordo com Bacchi et al.¹⁰, o comportamento dos estudantes é alterado com base no formato de avaliação que realizarão. Segundo Bacchi et al.¹⁰, os discentes relataram que preferem estudar por meio de livros e sites para as provas de questões de múltipla escolha, mas valorizam o ensino presencial de consultores locais para que possam preparar para o Teste de Concordância de Script (TCS), ou seja, alterações do formato de avaliação podem mudar hábitos e, por consequência, resultar em uma modificação significativa nas formas de aprendizado e ensino de medicina.

No atual contexto, em que a medicina está passando por transformações significativas devido ao avanço tecnológico, em especial com o crescimento do uso de ferramentas de inteligência artificial na área da saúde¹¹, avaliar como o raciocínio clínico está evoluindo e se adaptando a essas mudanças é de extrema importância para garantir a qualidade da formação médica e dos cuidados de saúde.

Os objetivos deste trabalho é revisar a literatura quanto às práticas adotadas nas formas de avaliação do desenvolvimento do raciocínio clínico de médicos e estudantes, bem como identificar lacunas nos formatos avaliativos e possíveis áreas de melhoria, contribuindo para o aprimoramento dos currículos e

da metodologia de ensino e avaliação, a fim de promover um desenvolvimento mais efetivo do raciocínio clínico dos futuros médicos, fomentar práticas clínicas mais eficientes e assegurar uma assistência de saúde de qualidade.

MÉTODO

Trata-se de uma revisão integrativa da literatura realizada com o objetivo de responder à seguinte pergunta orientadora:

- Quais são as principais formas de avaliação do desenvolvimento do raciocínio clínico de médicos e estudantes de Medicina?

Nesta revisão integrativa, utilizou-se, devidamente adaptada, a metodologia proposta por Mendes et al.¹² e adotada por diversos outros autores. Essa metodologia consiste em seis etapas: 1. elaboração da pergunta da pesquisa; 2. definição dos critérios de inclusão e exclusão; 3. elenco de informações a serem extraídas; 4. avaliação dos estudos incluídos; 5. interpretação dos resultados; e 6. apresentação da revisão.

Realizou-se a busca de publicações científicas nas bases de dados PubMed, Scopus, Embase, Web of Science, Lilacs, SciELO Citation Index, conforme mostra o Quadro 1.

Na sequência, eliminaram-se os estudos duplicados. Depois, avaliaram-se os estudos por meio de seus títulos e resumos utilizando a plataforma eletrônica Rayyan¹³ para verificar se atendiam aos critérios de inclusão e exclusão.

Foram selecionados artigos dos últimos 20 anos, sem filtros de idioma, que comparavam ou discutiam a avaliação do raciocínio clínico. As bases de dados consultadas, apresentadas no Quadro 1, incluem predominantemente artigos com títulos e resumos em inglês e português. Além disso, as palavras-chave utilizadas foram formuladas em inglês e português, o que pode ter limitado a recuperação de artigos publicados em outros idiomas.

Excluíram-se artigos não relacionados com a pergunta orientadora, editoriais, comentários, outras profissões da saúde e estudos sobre ensino sem métodos avaliativos.

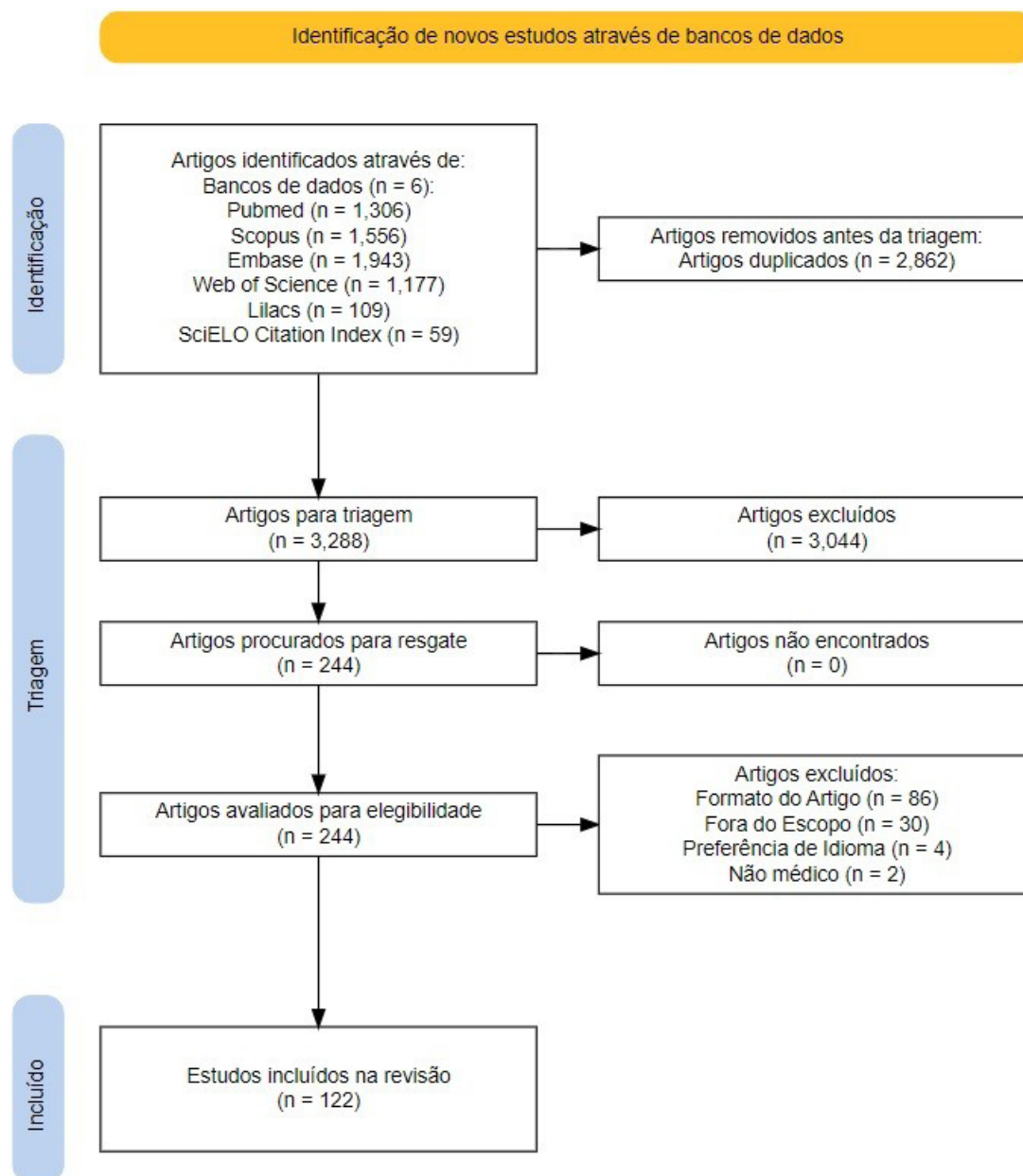
RESULTADOS E DISCUSSÃO

A busca identificou 6.150 trabalhos. Após a eliminação de duplicatas (2.862 artigos) e a avaliação de títulos e resumos, selecionaram-se 244 para leitura completa. A Figura 1 apresenta o processo utilizado¹⁴.

Quadro 1. Chaves de busca.

Base de dados	Chaves de busca	Campo
PubMed	(((“clinical reasoning”) and (“measurement” or “evaluation” or “assessment” or “examination” or “assessing” or “exam” or “test” or “tests” or “testing” or “judgment” or “appraisal” or “analysis” or “performance”) and (“medical education” or “resident” or “residents” or “medical student” or “medical students” or “physician” or “physicians” or “medical doctor” or “medical doctors” or “medical school” or “medical schools” or “internship” or “intern” or “interns” or “residency” or “clerkship” or “attending” or “attendings”)))	All fields
Scopus		Article title, abstract, keywords
Embase		All fields
Web of Science		All fields
Lilacs	(((“clinical reasoning”) and (“measurement” or “evaluation” or “assessment” or “examination” or “assessing” or “exam” or “test” or “tests” or “testing” or “judgment” or “appraisal” or “analysis” or “performance”) and (“medical education” or “resident” or “residents” or “medical student” or “medical students” or “physician” or “physicians” or “medical doctor” or “medical doctors” or “medical school” or “medical schools” or “internship” or “intern” or “interns” or “residency” or “clerkship” or “attending” or “attendings”)) OR (((“raciocínio clínico”) and (“avaliação” or “teste” or “performance” or “desempenho” or “julgamento” or “análise” or “acurácia” or “prova” or “exame” or “verificação”) and (“estudante de medicina” or “estudantes de medicina” or “internato” or “interno” or “internos” or “residência” or “residente” or “residentes” or “médico” or “médicos” or “educação médica” or “ensino médico” or “escola médica” or “escolas médicas”)))	Título, resumo, assunto
SciELO Citation Index		All fields

Fonte: Elaborado pelos autores.

Figura 1. Resultados.

Fonte: Elaborada pelos autores com base na ferramenta de Haddaway et al.¹³⁷.

Dentre os artigos selecionados para a revisão, percebeu-se que o ano de 2022 obteve o maior número de artigos selecionados ($n = 14$) e que os artigos que compreenderam o período de 2014-2023 representaram 68,85% do total de artigos, comparado com 31,15% dos anos 2003-2013. Essa crescente tendência de publicações ao longo dos anos sugere um interesse crescente e uma maior reflexão sobre os métodos de avaliação do raciocínio clínico no cenário científico contemporâneo. O crescimento dos estudos desde 2014 reflete a maior valorização da avaliação do raciocínio clínico.

A validade e a confiabilidade são essenciais na avaliação de instrumentos. A confiabilidade mede a consistência, e a validade depende dela. O alfa de Cronbach é amplamente usado para essa medição¹³⁶.

A avaliação de instrumentos educacionais exige confiabilidade e validação para assegurar resultados consistentes e precisos. No contexto deste estudo, utilizou-se o coeficiente alfa de Cronbach, amplamente empregado para medir a consistência interna de instrumentos na área da educação médica, garantindo que o instrumento aplicado reflita de forma confiável as habilidades de raciocínio clínico dos estudantes e profissionais de medicina¹³².

Os métodos mais comuns foram TCS, OSCE e QME. Também se identificam avaliações digitais, mas a maioria dos métodos foi citada poucas vezes. Mais adiante, descreveremos os diferentes métodos, e, para um melhor entendimento do leitor, segue a lista de abreviaturas:

- ART-R: Ferramenta reconstruída de avaliação do

- raciocínio.
 - ASCLIRE: Avaliação de raciocínio clínico (teste eletrônico).
 - CBCRT: Sistema de treinamento de raciocínio clínico baseado em computadores.
 - CBA: Avaliação baseada em casos multimídia.
 - CCS: Simulações de casos baseados em computadores.
 - CDI: Interpretação de dados clínicos.
 - CIP: *Puzzles* integrativos.
 - CIVA: Avaliação clínica com imagem e vídeo.
 - CRANAPL: Raciocínio clínico nas notas de admissão – avaliação e planejamento.
 - CRI-HT-S: Indicadores de raciocínio clínico – escala de anamnese.
 - CRC-PD: Casos de raciocínio clínico (plataforma digital).
 - CPX: Exame de *performance* clínica.
 - crSBA: Perguntas ricas em contexto de única melhor resposta.
 - CRT: Tarefa de raciocínio clínico.
 - CSE: Exame de habilidades clínicas.
 - DTI: Inventário de raciocínio diagnóstico.
 - DPR: Reconhecimento de padrões diagnósticos.
 - EMQ: Questões estendidas de correspondência.
 - GCE: Exame de competência de graduação.
 - HFS: Simulação de alta fidelidade.
 - IDEA: Resumo interpretativo, diagnóstico diferencial, justificativa de raciocínio e alternativas.
 - IGT: Testes de obtenção de informação.
 - KF: Características-chave.
 - LC: Avaliação de caso longo.
 - MEQ: Questão de ensaio modificada.
- NAR: Notas de admissão de residentes.
 - NPE: Notas pós-encontro do paciente padronizado.
 - ORT: Formulário de avaliação de observação.
 - OSCE: Exame clínico objetivo estruturado.
 - PD-R: Prova descritiva com gabarito por meio de rubrica.
 - PCE: Encontros com pacientes reais.
 - PERT: Ferramenta de avaliação pós-encontro.
 - PP: Pacientes padronizados.
 - PV: Pacientes virtuais.
 - QME: Questão de múltipla escolha.
 - QRC: Questões de resposta curta.
 - REACT: Ferramenta de avaliação rápida do raciocínio clínico.
 - SAVE: Exame de vinheta de respostas curtas.
 - SCBD: Discussão baseada em casos padronizados.
 - SCOE: Exame oral clínico estruturado.
 - SF: Formação de cenários.
 - SG: Jogo de simulação.
 - SRL: Aprendizado autorregulado.
 - SSAR: Rubrica de avaliação de resumo.
 - TCS: Teste de concordância de *script*.
 - TCS_e: Teste de concordância de *script* em evolução.

Teste de concordância de *script*

Os TCS são compostos por breves cenários clínicos que envolvem incerteza, replicando o processo de raciocínio clínico. Os participantes são apresentados a uma série de perguntas que exploram como novas informações influenciam suas decisões. As respostas dos participantes são então comparadas com as de um painel de especialistas, permitindo avaliar a variabilidade nas respostas dos profissionais em diferentes contextos clínicos^{32,33,53-55,61,66,68,73,82,87,89,105,112,132}.

Quadro 2. Principais resultados.

Autores	País	Ferramenta de avaliação	Pontos principais
Cohen Aubart et al. ²⁰	França	TCS	Alfa de Cronbach entre 0,08 e 0,59.
Bhardwaj et al. ²⁷	Estados Unidos	TCS	Fraca correlação entre o TCS e várias provas de competência médica.
Blunk et al. ²⁹	Estados Unidos	TCS	O TCS para psiquiatria melhora o raciocínio clínico dos estudantes.
Boulouffe et al. ³⁰	Bélgica	TCS	Pontuações do TCS dependem do painel usado para referência.
Carrière et al. ³²	Canadá	TCS	Os residentes aprovaram a representação de casos no TCS.
Charlin et al. ³³	Canadá	TCS	O TCS demonstrou eficácia na avaliação do raciocínio clínico.
Compère et al. ³⁷	França	TCS	Resultados dos residentes foram mais baixos do que dos especialistas.
Cooke et al. ³⁸	Estados Unidos	TCS _e	O TCS _e é considerado envolvente e próximo de casos reais.

Continua...

Quadro 2. Continuação.

Autores	País	Ferramenta de avaliação	Pontos principais
Couraud et al. ³⁹	França	TCS	Os TCS são uma avaliação adequada para a formação contínua.
Ducos et al. ⁴³	França	TCS	O TCS é confiável para avaliar raciocínio clínico em residentes.
Duggan et al. ⁴⁵	Austrália	TCS	As razões para as perguntas do TCS serem inadequadas não são óbvias.
Enser et al. ⁴⁷	França	TCS	O TCS mediu que ruído afeta o raciocínio clínico de residentes.
Gagnon et al. ⁵³	França	TCS	Painel de 20 membros para exames importantes; mais de dez para exames aceitáveis.
Gagnon et al. ⁵⁴	Canadá	TCS	Testes considerados confiáveis geralmente incluem poucas perguntas, mas abrangem um número maior de casos clínicos (20-30), enquanto testes menos confiáveis tendem a conter menos casos (de quatro a cinco perguntas) para cada conjunto de 25-30 casos clínicos.
Gagnon et al. ⁵⁵	Canadá	TCS	Os resultados sugerem robustez da metodologia, requerendo confirmação em outros contextos.
Iglesias Gómez et al. ⁵⁸	Espanha	TCS	O TCS é eficaz na medição do raciocínio clínico.
Iglesias Gómez et al. ⁵⁹	Espanha	TCS	O TCS supervisionado pode melhorar cuidados pediátricos em residência.
Goos et al. ⁶⁰	Alemanha	TCS	O TCS evidencia discrepância entre especialistas e estudantes em cirurgia visceral.
Goulet et al. ⁶¹	Estados Unidos	TCS	O coeficiente alfa de Cronbach para o TCS foi de 0,90, indicando um teste de alta confiabilidade.
Hamui et al. ⁶⁴	Argentina	TCS	Primeira experiência de TCS viável em escala nacional (Argentina) em pediatria.
Humbert et al. ⁶⁶	Estados Unidos	TCS	O TCS diferencia estudantes de diferentes anos e especialistas.
Humbert et al. ⁶⁷	Estados Unidos	TCS	Os alunos passaram a lidar melhor com as incertezas em cenários clínicos, medidos pelo TCS.
Humbert et al. ⁶⁸	Estados Unidos	TCS	O TCS demonstra ser promissor na avaliação do raciocínio clínico.
Iravani et al. ⁷¹	Irã	TCS	O TCS é uma ferramenta confiável para avaliar o raciocínio clínico.
Kania et al. ⁷³	França	TCS	Os participantes consideraram o teste realista, interessante, relevante e intuitivo.
Kaur et al. ⁷⁴	Índia	TCS	Os estudantes de Medicina demonstraram alta satisfação com esse método.
Kazour et al. ⁷⁵	Líbano	TCS	Os alunos melhoraram 11% nas pontuações do TCS após o treinamento.
Kün-Darbois et al. ⁷⁹	França	TCS	A opinião global de estudantes e professores de Medicina sobre o TCS foi globalmente negativa.
Lambert et al. ⁸²	Canadá	TCS	O TCS é útil para discriminar participantes de acordo com seu nível de experiência.
Lineberry et al. ⁸³	Argentina	TCS	Explicita duas situações hipotéticas com potencial para interpretação errônea das respostas do TCS.
Talvard et al. ⁸⁶	França	TCS	Embora o TCS seja um método útil, sua elaboração pode ser difícil.
Marie et al. ⁸⁷	França	TCS	O TCS permite diferenciar grupos de candidatos com base em seu nível de competência.
Mathieu et al. ⁸⁸	França	TCS	Limitações: Custo e tempo de desenvolvimento e recrutamento.
Meterissian et al. ⁸⁹	Canadá	TCS	Teste útil para formação e avaliação em cirurgia geral.

Continua...

Quadro 2. Continuação.

Autores	País	Ferramenta de avaliação	Pontos principais
Mzoughi et al. ⁹¹	Tunísia	TCS	Média de notas no TCS superou prova de referência. Correlação ausente.
Nomura et al. ⁹³	Japão	TCS	Foi criado e validado um TCS em japonês para avaliar a educação clínica no Japão.
Nouh et al. ⁹⁴	Canadá	TCS	O teste foi capaz de diferenciar residentes juniores (R1-R2) de residentes seniores (R3-R5).
Omega et al. ⁹⁵	Indonésia	TCS	O TCS foi capaz de discriminar grupos de diferentes experiências clínicas.
Petrucchi et al. ⁹⁸	Canadá	TCS	O artigo apresenta resultados contraditórios.
Peyrony et al. ⁹⁹	França	TCS	As pontuações do TCS variaram estatisticamente, porém as diferenças foram fracas.
Piovezan et al. ¹⁰⁰	Brasil	TCS	O TCS é viável e fácil de desenvolver, administrar e corrigir.
Power et al. ¹⁰²	Canadá	TCS	O estudo mostrou que o TCS pode ser melhorado com um protocolo escrito de <i>think-aloud</i> .
Ruiz et al. ¹⁰⁵	Estados Unidos	TCS	O TCS pode ser útil como <i>feedback</i> formativo para estagiários e médicos.
Sibert et al. ¹¹²	França	TCS	A validade de construto do TCS <i>online</i> não é direta.
Steinberg et al. ¹¹⁶	Estados Unidos	TCS	O TCS tem utilidade limitada na avaliação do raciocínio clínico entre os residentes.
Subra et al. ¹¹⁷	França	TCS	A pontuação dos especialistas foi mais alta do que a dos alunos em todas as avaliações ($p < 0,001$).
Van den Broek et al. ¹²²	Holanda	TCS	Incluir diagnóstico diferencial resultou em menor confiabilidade, contrariando as expectativas.
Wan et al. ¹²⁵	Austrália	TCS	A pontuação dos candidatos foi maior com o aumento do nível de experiência clínica.
Wan et al. ¹²⁶	Austrália	TCS	As respostas do teste refletem pensamentos, validando o processo do TCS.
Wilson et al. ¹²⁸	Estados Unidos	TCS	Ambiguidade na dimensionalidade do TCS prejudica a interpretação das pontuações.
Wilson et al. ¹²⁹	Estados Unidos	TCS	Pontuação em escala de 5 pontos é mais confiável que em escala de 3 pontos.
Yassin et al. ¹³⁰	Iraque	TCS	A pontuação dos residentes juniores no TCS não diferiu muito da dos residentes seniores.
Zavaleta-Hernández et al. ¹³²	México	TCS	Os residentes apresentaram melhora significativa no desempenho com experiência clínica.

Fonte: Elaborado pelos autores.

Dentre as ferramentas de avaliação do raciocínio clínico, a mais utilizada nos estudos que fazem parte da revisão foi o TCS. Testou-se o TCS em diversas áreas da medicina, como psiquiatria²⁹, otorrinolaringologia^{71,73}, anestesiologia^{36,43,47,95}, cirurgia^{60,89}, pediatria^{32,59}, entre outras. Diversos estudos^{37,60,67,75,92,94,110,117,125} relatam que o TCS foi capaz de distinguir com sucesso os alunos de acordo com seus diferentes estágios de formação, como entre um aluno do quinto ano e um iniciante, ou entre um residente e um aluno, e reconhecer as habilidades de um especialista em relação a um residente, por exemplo. Isso é uma característica importante, visto que reflete a expectativa de que estudantes e médicos com maior grau de experiência devem apresentar melhor desempenho do que aqueles que

tiveram um menor tempo de treinamento. Alguns autores relataram também em seus resultados que o TCS pode ser uma avaliação com grau de confiabilidade alto medido pelo alfa de Cronbach^{17,29,30,33,36,37,47,53,59-61,71,73,75,82,84,87-89,91,93,94,98,100,105,106,112,129}.

Nomura et al.⁹³ concluíram que o TCS funciona em línguas diferentes da inglesa, como o japonês⁹³. Além desses achados, inúmeros autores sugerem que o TCS é um instrumento válido para avaliar o raciocínio clínico^{33,36,39,53-55,58,74,82,84,87,100,105}.

Apesar de uma gama abrangente de estudos indicar sucesso ao utilizar o TCS, outros estudos apontam resultados divergentes. Alguns estudos mostram que o TCS não conseguiu distinguir os diferentes graus de instrução dos médicos¹³⁰. Muitos relataram que é muito oneroso e/

ou difícil construir as perguntas para o TCS^{45,86,88,98,98,122} e que a confiabilidade do teste medido pelo alfa de Cronbach é baixa^{20,27,39,43,63,64,67,92,95,99,116,122,130}.

Alguns *feedbacks* de utilização tanto por parte dos alunos quanto por alguns professores não foram positivos^{64,76,79}. Esse fato pode ser explicado pela novidade do ambiente do TCS e pela natureza incomum do raciocínio médico exigido⁷⁹. Kelly et al.⁷⁶ mostram em seu trabalho que os alunos preferiram as QME ao TCS e citaram a familiaridade e a percepção de objetividade com o formato QME.

De acordo com outros *feedbacks*, o TCS pode ser mais útil para o ensino do que para a avaliação⁶³, já que a validade do construto é limitada ou ambígua, o que compromete o significado e a interpretação das pontuações do TCS^{112,116,129}.

Outro ponto a ser mencionado é que houve baixa correlação entre o TCS e outros formatos de avaliação, e isso significa que uma boa nota no TCS não implicará necessariamente boa nota caso aplique outro estilo de avaliação. Amini et al.¹⁷ encontraram uma fraca correlação entre as notas do TCS e de QME. Humbert et al.⁶⁸ constataram fraca correlação entre as notas do TCS e do Step 2 CK. Bhardwaj et al.²⁷ observaram fraca correlação entre as notas do TCS e as provas do USMLE Step 1 e Step 2, e outras.

Outros achados incluem os de Mzoughi et al.⁹¹ que não encontraram correlação entre as notas do TCS e da prova de cardiologia de referência. Ducos et al.⁴³ explicam em seu estudo que o TCS mede diferentes domínios de outros testes. Assim também Groves et al.⁶³ mostraram que o TCS e o CRP também medem diferentes aspectos do raciocínio clínico. Isso é extremamente relevante, pois, em teoria, todos os instrumentos de avaliação medem o raciocínio clínico, mas a literatura revela que há divergências no que poderia ser considerado raciocínio clínico, e, embora raciocínio clínico uma expressão especialmente abrangente, falta ainda uma definição mais precisa. Durante nossa revisão de 122 artigos, observamos que alguns autores tentaram definir e caracterizar o construto em suas partes desde a anamnese, o exame físico, o diagnóstico etc.^{10,21,23,44,48,52,75,85,104,113,114,123}. Outros assumem que é somente o diagnóstico, e para a maioria está vagamente classificado dentro do grande guarda-chuva que essa expressão pode englobar.

Inúmeros outros estudos acerca do TCS foram realizados ao longo dos anos, como testes na escala de Likert com 3 pontos a 5 pontos, sendo preferível a escala de 5 pontos¹²⁹. Foi investigado se os alunos selecionavam as respostas no TCS que refletiam seu raciocínio clínico real, e, em geral, descobriu-se que sim¹²¹. Com base nisso, sugeriu-se incorporar uma técnica de pensamento em voz alta⁹⁷ para aprimorar a avaliação do TCS. Dificuldades adicionais foram analisadas,

como divergências na interpretação da pergunta e resposta⁸³, a vulnerabilidade na utilização de valores extremos¹¹⁰ na escala de Likert, a possibilidade de responder às questões do TCS sem ler a vinheta-problema⁹⁰, a importância de divulgar a nota do painel e o desvio padrão para ter referência de notas³⁰, e até mesmo uma evolução do TCS³⁸.

Os resultados desta revisão indicam que métodos como o TCS e o OSCE são amplamente empregados na avaliação do raciocínio clínico em médicos e estudantes de Medicina. Daniel et al.¹³⁸ também identificaram e classificaram esses métodos em sua revisão de escopo, destacando a importância de diferentes tipos de avaliação, como avaliações práticas e simulações. A principal diferença entre os estudos está na abordagem metodológica: enquanto Daniel et al.¹³⁸ mapearam uma ampla gama de métodos a partir de múltiplas fontes, incluindo literatura cinzenta, nosso trabalho utiliza uma revisão integrativa com critérios de seleção mais restritivos, priorizando estudos de maior controle metodológico.

Exame clínico objetivo estruturado

Os OSCE consistem em várias estações em que os examinandos realizam diferentes tarefas clínicas, usando pacientes padronizados, avaliações de observadores, anotações escritas e outros métodos para uma avaliação abrangente. Esse foi o segundo tipo de prova mais encontrado dentro da nossa revisão da literatura. De acordo com autores, o OSCE não é uma prova realizada para medir exclusivamente o raciocínio clínico, envolve habilidades de anamnese, exame físico, comunicação, raciocínio clínico, procedimentos e profissionalismo¹¹⁴. Volkan et al.¹²³, em seus estudos, conseguiram distinguir dois grandes fatores que se ajustaram bem aos dados. O primeiro fator estava relacionado ao exame físico e à anamnese, rotulado como coleta de informações, enquanto o segundo fator estava relacionado ao diagnóstico diferencial/raciocínio clínico e à interação com o paciente, rotulado como raciocínio e disseminação de informações. Sim et al.¹¹⁴ destacaram que os alunos analisados tiveram notas mais elevadas durante a etapa procedimental do OSCE e notas mais baixas na temática de raciocínio clínico, indicando que esse formato de ensino, além de abrangente, não é tão direto.

Inúmeros pesquisadores consideram o uso desse instrumento como positivo^{23,24,46,57,85,123}, e muitos obtiveram bons resultados^{48,50} de confiabilidade medidos pelo alfa de Cronbach. Volkan et al.¹²³ realizaram o exame com nove estações separadas (categorias de obstetrícia/ginecologia, pediatria, exame de mama, patologia, psiquiatria, asma, radiologia, cardiologia e neurologia) e obtiveram boa confiabilidade, exceto nas estações de obstetrícia/ginecologia e patologia.

Quadro 3. Principais resultados.

Autores	País	Ferramenta de avaliação	Pontos principais
Behrens et al. ²³	Chile	OSCE	Exame colaborativo viável para estudantes de Medicina chilenos.
Berger et al. ²⁴	Estados Unidos	OSCE	Leigos podem avaliar raciocínio clínico com treinamento adequado.
Durning et al. ⁴⁶	Estados Unidos	OSCE	O OSCE é um formato válido para avaliar alunos do segundo ano de Medicina.
Fleiszer et al. ⁵⁰	Canadá	OSCE	O raciocínio clínico melhora com a prática clínica; a avaliação é viável.
Gallagher et al. ⁵⁶	Estados Unidos	OSCE	Notas escritas no OSCE foram associadas ao desempenho em rotação clínica.
Gilkes et al. ⁵⁷	Austrália	OSCE	Feedback padronizado é útil para ensinar e avaliar estudantes.
Lukas et al. ⁸⁵	Estados Unidos	OSCE	Os participantes do OSCE obtiveram posteriormente pontuações mais altas.
Park et al. ⁹⁶	Coreia do Sul	OSCE	A pontuação do OSCE pode não refletir a habilidade de raciocínio clínico de um estudante de Medicina.
Sim et al. ¹¹⁴	Malásia	OSCE	O OSCE mede habilidades de anamnese, exame físico, comunicação, raciocínio clínico, procedimentais e profissionalismo.
Volkan et al. ¹²³	Estados Unidos	OSCE	Um modelo com dois fatores correlacionados se ajustou bem aos dados.

Fonte: Elaborado pelos autores.

Quadro 4. Principais resultados.

Autores	País	Ferramenta de avaliação	Pontos principais
Silberman et al. ¹¹³	Estados Unidos	QME	Ferramenta pode ser uma avaliação formativa útil do raciocínio psiquiátrico.
Surry et al. ¹¹⁹	Estados Unidos	QME	O artigo avalia a disposição cognitiva dos estudantes ao realizarem uma QME.

Fonte: Elaborado pelos autores.

De acordo com o estudo realizado por Berger et al.²⁴, os avaliadores leigos, em posse de uma rubrica, podem fornecer resultados similares a avaliadores médicos. Esse achado é muito interessante, pois o valor financeiro despendido em um médico é muito alto, o que poderia reduzir custos na hora de dar as notas. Gilkes et al.⁵⁷ também utilizaram uma rubrica para avaliar e apoiam a utilidade de uma ferramenta de *feedback* simples e padronizada para ensino e avaliação de estudantes de Medicina.

Nem todos os estudos foram positivos com relação ao OSCE. Na percepção de Park et al.⁹⁶, as notas do OSCE podem não refletir a habilidade de raciocínio clínico de um estudante de Medicina. Os autores adicionam ainda que esforços devem ser feitos para melhorar a avaliação das habilidades de raciocínio clínico dos estudantes de Medicina usando OSCE. Fleiszer et al.⁵⁰ relatam que ainda há limitações na utilização de rubricas para avaliação, pois perceberam que o raciocínio clínico, conforme medido por algumas de suas características “procedimentais”, melhora ao longo do ano do estágio clínico.

Rubricas podem ser criadas para avaliar objetivamente questões “procedimentais”, mas não questões “semânticas”, ou seja, aspectos como a eficácia do raciocínio clínico, a tomada de decisões e a execução de procedimentos médicos seriam mais fáceis de medir utilizando uma rubrica do que a compreensão, interpretação e precisão na comunicação.

Questões de múltipla escolha

Uma QME na medicina pode apresentar um cenário clínico. Nesse caso, há uma pergunta com várias opções de resposta, usualmente cinco, em que somente uma é correta.

As QME são atualmente o método mais comum⁷⁶ para avaliação nas faculdades de Medicina. Percebe-se que os estudos sobre questões de múltipla escolha têm um caráter mais aprofundado, visto que ele já é tão difundido, ou então um caráter comparativo, pois tenta-se comparar diferentes métodos com o padrão ouro para encontrar alguma correlação.

Silberman et al.¹¹³ prepararam uma ferramenta que pode ser uma avaliação formativa do raciocínio psiquiátrico, como

um complemento às diversas metodologias já disponíveis nessa área difícil da educação. Surry et al.¹¹⁹ avaliaram as disposições cognitivas dos estudantes ao realizarem uma prova de QME. Coderre et al.³⁵ obtiveram resultados que apoiam a ideia de que QME bem elaboradas podem, de fato, testar o raciocínio clínico de ordem superior. Além disso, ressaltam que, ao testar o raciocínio clínico, o enunciado da pergunta ou o conteúdo permanecem mais importantes do que o número de alternativas. Um dos estudos¹³⁴ obteve um grau questionável de confiabilidade medido pelo alfa de Cronbach, indicando que, mesmo uma prova padrão ouro, dependendo do contexto em que foi desenvolvida ou aplicada, pode não ser um instrumento de medida tão confiável.

Avaliações digitais, online e simulações

Nos recentes anos, inúmeras formas de avaliação foram introduzidas com base em multimídia e vídeo, computadores ou *online*. No entanto, percebeu-se que os estudos estão em menor quantidade e podem ser material de estudos nos próximos anos. Isso foi útil, em casos como a pandemia global de Covid-19 na qual autores como Duffy et al.⁴⁴ adaptaram a versão do *long case* para uma versão 100% digital eliminando o exame físico. No entanto, concluíram que, em última análise, o uso de pacientes simulados isoladamente para desenvolver habilidades clínicas não substitui encontros com pacientes reais. Resultados similares foram vistos no trabalho de Fink et al.⁴⁹ em que a autenticidade percebida e a precisão diagnóstica foram superiores para os pacientes padronizados (PP) quando comparados aos pacientes virtuais (VP), apesar de a carga cognitiva na realização da atividade ser equivalente.

A utilização de uma plataforma digital possui inúmeras vantagens, como a automação de vários serviços. Waechter et al.¹²⁴ estudaram uma plataforma digital com casos de raciocínio clínico, em que, com apenas dois instrutores, conseguiram fornecer *feedback* detalhado em 376 casos, demonstrando carga de trabalho viável e sustentável. Schaye

et al.¹⁰⁹ desenvolveram e validaram um modelo de *Machine Learning* de alto desempenho que classifica a qualidade do raciocínio clínico na documentação nas notas de admissão de residentes no ambiente clínico, ou seja, incluíram algoritmos sofisticados para avaliação que reduzem, e muito, a mão de obra do processo avaliativo. Kotwal et al.⁷⁸ realizaram estudo similar, em que analisaram a ferramenta de raciocínio clínico nas notas de admissão – avaliação e planejamento (CRANAPL) – e constataram que, em razão da facilidade de uso e da versatilidade da ferramenta, as divisões de hospitalistas em ambientes acadêmicos e não acadêmicos podem usar a ferramenta CRANAPL para avaliar e fornecer *feedback* sobre a documentação do raciocínio clínico dos hospitalistas. Todos esses estudos, respondem a maneiras sobre como educadores e médicos conseguem fornecer *feedback* mais detalhado e consistente de forma viável e sustentável.

Zuo et al.¹³⁵ utilizaram um sistema de treinamento de raciocínio clínico baseado em computadores (CBCRT) e constataram que as pontuações do teste aumentaram com os anos de treinamento, ou seja, há indícios de que o *software* esteja avaliando de forma coerente, conseguindo diferenciar os alunos em seus diferentes estágios de formação.

Dong et al.⁴² utilizaram simulações de casos baseado em computadores (CCS) com o *software* DxR Clinician e sugeriram ações de treinamento, pois as notas dos residentes foram inferiores às notas internacionais. Essa implicação é extremamente relevante, visto que, com sistemas de *softwares* padronizados, é possível analisar e avaliar diferentes instituições no mundo e criar sistemas para garantir um ensino de medicina com qualidade universal. Fida et al.⁴⁸ também utilizaram o mesmo *software* e relataram que ele apresenta boa confiabilidade de consistência interna e que mede a competência clínica de um aluno em um construto relacionado aos medidos pelos OSCE. No entanto, parece que as pontuações do CCS medem um construto diferente daquele medido por meio de exames de

Quadro 5. Principais resultados.

Autores	País	Ferramenta de avaliação	Pontos principais
Dong et al. ⁴²	China	CCS	Pontuações de CCS de residentes diminuíram com anos de treinamento.
Jerant et al. ⁷²	Estados Unidos	CCS	Pontuações de <i>software</i> de CCS (D&X Clinician) parecem não ter validade de critério.
Kotwal et al. ⁷⁸	Estados Unidos	CRANAPL	Hospitalistas podem usar a ferramenta CRANAPL para avaliar documentação clínica.
Lai et al. ⁸¹	Taiwan	PP	Obter o “diagnóstico mais provável” correlaciona-se com diferentes habilidades.
Pottier et al. ¹⁰¹	França	PP	Prefere-se a avaliação global por conta da rapidez e eficácia.
Zuo et al. ¹³⁵	China	CBCRT	As pontuações aumentaram com treinamento.

Fonte: Elaborado pelos autores.

encontro com pacientes reais, ou seja, o CCS é confiável e mede competência clínica, mas pode não capturar completamente as habilidades avaliadas em encontros com pacientes reais. Outro estudo, mais antigo, de Jerant et al.⁷² relata que as pontuações automatizadas de raciocínio diagnóstico geradas por um *software* de CCS amplamente disponível (D&X Clinician) parecem não ter validade de critério. Isso implica que, apesar do grande potencial da ferramenta, mais estudos são necessários para garantir que a utilização desses novos métodos de estudo sejam confiáveis para a utilização em escala global.

Uma nova vertente foi encontrada no trabalho de Fonteneau et al.⁵¹ que utilizaram conceitos de gamificação para construir um jogo de simulação (SG) e o compararam a uma simulação de alta fidelidade (HFS) e QME. Os autores concluíram que o SG usado refletiu melhor a competência clínica dos alunos em HFS do que uma QME no mesmo caso clínico de exacerbação de asma pediátrica. Além disso, acrescentam que os alunos apreciaram o jogo como método de avaliação.

Sulaiman et al.¹¹⁸ realizaram uma avaliação clínica com imagem e vídeo (CIVA) e constataram forte correlação entre o novo método (CIVA) e o OSCE ($r = 0,83$, $p < 0,001$). Kim et al.⁷⁷ fizeram uma avaliação baseada em casos multimídia (CBA) e concluíram que ela estimula mais fortemente a análise e síntese de dados do que outros instrumentos avaliativos, como questões de ensaio modificadas (MEQ) e exame de desempenho clínico (CPX).

Kunina-Habenicht et al.⁸⁰ realizaram um estudo utilizando um formulário eletrônico chamado avaliação de raciocínio clínico (ASCLIRE) e testaram a validade do construto. Calcularam-se correlações com variáveis externas ao teste em estudo, como QME e resultados do OSCE, e não foram encontradas correlações fortes. De fato, a ASCLIRE se correlacionou aproximadamente na mesma intensidade com QME e os resultados do OSCE, assim como esses dois se correlacionam entre si. Dado que a maioria concorda que QME e OSCE medem diferentes aspectos do raciocínio clínico (conhecimento declarativo e procedimental, respectivamente), argumenta-se que a ASCLIRE mensura um terceiro construto. Schaubert et al.¹⁰⁷ também utilizaram o mesmo formulário eletrônico e concluíram que não há um formato de resposta geralmente “mais válido” do que outro para avaliar o raciocínio

clínico. O uso exclusivo de um formato pode prejudicar a validade do teste. A diversidade de formatos de resposta é essencial na concepção de avaliações em educação médica, pois não é o formato sozinho, mas sim a tarefa como um todo, que influencia os processos de raciocínio.

Essa afirmação de Schaubert et al.¹⁰⁷ é extremamente relevante, pois a avaliação é um método para condicionar o comportamento. O ato de combinar diferentes formatos foi testado inúmeras vezes por diferentes autores em olimpíadas de medicina no Irã^{16,90,106}. Amini et al.¹⁶ construíram um instrumento de avaliação que combinava os seguintes tipos de questão: características-chave (KF), TCS, problemas de raciocínio clínico (CRP) e *puzzles* integrativos abrangentes (CIP). Destaca-se que o novo teste que engloba os diferentes formatos de perguntas obteve valores excelentes de confiabilidade pelo método do alfa de Cronbach. Sadeghi et al.¹⁰⁶ reproduziram algo similar ao instrumento elaborado por Amini et al.¹⁶ e obtiveram resultados semelhantes. Monajemi et al.⁹⁰ construíram um instrumento de avaliação que combinava os seguintes tipos de questão: KF, TCS, CRP, CIP, testes de obtenção de informação (IGT) e formação de cenários (SF), alguns inovadores. Porém, constataram problemas como a infamiliaridade com os testes e muito tempo para construir perguntas do tipo IGT. Além disso, construir perguntas do TCS não é um processo trivial, visto que as questões têm que ser muito bem formuladas, pois muitas vezes não é necessário ler o cenário para responder a elas.

Métodos conjuntos de avaliação

A combinação de diferentes métodos de avaliação tem sido explorada como uma estratégia para obter uma visão mais abrangente e confiável do raciocínio clínico de estudantes e médicos. Essa abordagem integrativa permite aproveitar as vantagens e compensar as limitações de cada método, fornecendo um quadro mais completo das habilidades clínicas. O Quadro 7 apresenta uma síntese dos principais estudos que aplicaram métodos conjuntos, destacando a confiabilidade e os pontos-chave de cada combinação.

Como mostrado no Quadro 7, a combinação de métodos melhora a confiabilidade da avaliação, identifica habilidades e lacunas educacionais, e reforça a importância da diversidade metodológica.

Quadro 6. Principais resultados.

Autores	País	Ferramenta de avaliação	Pontos principais
Kunina-Habenicht et al. ⁸⁰	Alemanha	ASCLIRE	Argumenta-se que a ASCLIRE mensura um terceiro construto.
Schauber et al. ¹⁰⁷	Alemanha	ASCLIRE	A diversidade de formatos de resposta é essencial na avaliação.

Fonte: Elaborado pelos autores.

Quadro 7. Principais resultados.

Autores	País	Ferramenta de avaliação	Pontos principais
Amini et al. ¹⁶	Irã	KF, TCS, CRP, CIP e Novo	A combinação de testes mostra alta confiabilidade.
Amini et al. ¹⁷	Irã	TCS e QME	O estudo mostrou fraca correlação entre TCS e QME.
Bacchi et al. ¹⁰	Austrália	TCS e QME	Preferência por livros e <i>sites</i> para QME; ensino presencial para TCS.
Coderre et al. ³⁵	Canadá	QME e EMQ	As QME testam raciocínio clínico de ordem superior.
Compère et al. ³⁶	França	TCS, QME e QRC	O TCS indica que a intervenção educacional melhora o raciocínio clínico em residentes.
Covin et al. ⁴⁰	Estados Unidos	CRT, PNS, SSAR e CDI	Diferentes instrumentos medem componentes distintos do raciocínio clínico.
Derakhshandeh et al. ⁴¹	Irã	CRP e QME	O CRP pode ser uma alternativa para medir habilidades de raciocínio clínico.
Fida et al. ⁴⁸	Barein	CCS, QME, QRC, OSCE e PCE	Pontuações do CCS têm boa confiabilidade, mas diferem de OSCE.
Fink et al. ⁴⁹	Alemanha	PP e PV	PP é superior a PV em autenticidade e precisão diagnóstica; carga cognitiva equivalente.
Fonteneau et al. ⁵¹	França	HFS, SG e QME	SG reflete competência clínica melhor que QME.
Fürstenberg et al. ⁵²	Alemanha	CRI-HT-S	CRI-HT-S é consistente para avaliar raciocínio clínico.
Groves et al. ⁶²	Estados Unidos	DTI e CRP	DTI pode limitar a diversidade diagnóstica.
Groves et al. ⁶³	Austrália	TCS e CRP	CRP e TCS complementam-se, mais úteis para ensino do que para avaliação.
Haring et al. ⁶⁵	Holanda	ORT e PERT	ORT e PERT: confiáveis, válidos, correlação fraca entre eles.
Huwendiek et al. ⁶⁹	Alemanha	crSBA, KF e OSCE	KF é mais realista, confiável e eficiente que crSBA.
Kelly et al. ⁷⁶	Estados Unidos	TCS e QME	Apenas uma minoria de estudantes preferiu o TCS à QME.
Kim et al. ⁷⁷	Coreia do Sul	CPX, CBA e MEQ	Avaliações distintas incentivam diferentes aspectos do raciocínio clínico dos participantes.
Loh et al. ⁸⁴	Singapura	TCS e EMQ	Experiência e formação predizem habilidade de raciocínio pós-graduação.
Monajemi et al. ⁹⁰	Irã	KF, TCS, CRP, CIP, IGT e SF	Problemas incluem familiaridade, formato extenso e formulação do teste.
Nazim et al. ⁹²	Paquistão	TCS e EMQ	TCS e EMQ discriminam aprendizes pela experiência clínica em urologia.
Sadeghi et al. ¹⁰⁶	Irã	KF, TCS, CRP, CIP e Novo	Combinação de testes é recomendada para exames de alto risco.
See et al. ¹¹⁰	Singapura	TCS e QME	O TCS distingue competências, mas é vulnerável a valores extremos.
Williams et al. ¹²⁷	Estados Unidos	DPR e CDI	O desempenho dos alunos aumenta com anos de treinamento médico.
Zia et al. ¹³⁴	Paquistão	QME e KF	Tempo limitado afeta a confiabilidade do exame.

Fonte: Elaborado pelos autores.

Outros instrumentos de avaliação

Outros métodos foram encontrados em poucas referências, sugerindo que podem ser métodos não tão comuns nos cenários de ensino médico ou que representam áreas promissoras para futuros estudos. Por exemplo, KF^{69,133} são eficazes para avaliar decisões em situações clínicas específicas, mas sua aplicabilidade pode ser limitada em contextos mais amplos, nos quais se busca avaliar o raciocínio clínico de forma

geral. Provas descritivas¹¹⁵ e orais^{103,111} oferecem oportunidades de avaliação direta do pensamento crítico, embora enfrentem limitações práticas, como subjetividade e falta de padronização.

O Exame de Vinheta de Respostas Curtas (SAVE)²⁸ e o método “preceptor de um minuto”¹¹⁵ proporcionam *feedback* imediato, mas carecem de robustez em termos de mensuração objetiva. Métodos como resumo interpretativo, diagnóstico diferencial, justificativa de raciocínio e alternativas (IDEA)^{21,108}

e o Inventário de Raciocínio Diagnóstico (DTI)^{25,62} são úteis para captar nuances do raciocínio diagnóstico, embora suas implementações exijam recursos especializados para análise detalhada dos resultados.

Instrumentos como a Ferramenta de Avaliação Rápida do Raciocínio Clínico (REACT)⁹⁷ e o Medindo o Pensamento Analítico na Assistência à Saúde Clínica (MATCH)²² trazem inovações ao integrarem raciocínio e análise prática, mas

sua aplicação ainda depende de validação adicional em diferentes contextos culturais e educacionais. Instrumentos como o crSBAs (Perguntas ricas em contexto de única melhor resposta)⁶⁹ e as Questões Estendidas de Correspondência (EMQ)²⁶ facilitam a avaliação padronizada de conhecimentos, enquanto CIP³¹ e PP^{81,101} são ferramentas que integram habilidades práticas e teóricas, embora demandem tempo e recursos para sua implementação.

Quadro 8. Principais resultados.

Autores	País	Ferramenta de avaliação	Pontos principais
Aagaard et al. ¹⁵	Estados Unidos	OMP	Preceptores: OMP melhora diagnóstico, confiança, eficácia e eficiência.
Anderson et al. ¹⁸	Austrália	CRP	É necessária a validação adicional para aplicação.
Artino et al. ¹⁹	Estados Unidos	SRL	Os alunos são cientes, mas poucos aplicam estratégias de raciocínio diagnóstico.
Baker et al. ²¹	Estados Unidos	IDEA	Ferramenta válida e confiável para avaliação formativa de baixo risco.
Baño et al. ²²	Chile	MATCH	A pontuação aumentou com a experiência.
Beullens et al. ²⁵	Bélgica	DTI	As notas, avaliadas com DTI, aumentaram após os seminários clínicos.
Beullens et al. ²⁶	Bélgica	EMQ	EMQ distingue especialistas de menos especializados.
Blazek et al. ²⁸	Estados Unidos	SAVE	SAVE é útil na avaliação de estudantes de Medicina em psiquiatria.
Capaldi et al. ³¹	Estados Unidos	CIP	CIP individual não se correlacionou com exames de múltipla escolha.
Cheung et al. ³⁴	Estados Unidos	NPE	A criação de regras e diretrizes de pontuação é intensiva em recursos.
Duffy et al. ⁴⁴	Irlanda	LC sem exame físico	Pacientes simulados não substituem encontros com pacientes reais.
Im et al. ⁷⁰	Coreia do Sul	CPX	CPX: alta confiabilidade, problemas com validade.
Peterson et al. ⁹⁷	Estados Unidos	REACT	Objetivos educacionais competem com situações clínicas reais.
Rajeswaran et al. ¹⁰³	Canadá	SCOE	Desempenho positivo dos clínicos no SCOE, sem correlação com outras avaliações.
Reinert et al. ¹⁰⁴	Estados Unidos	CSE	CSE é válido para avaliar o desempenho dos alunos no estágio de cirurgia.
Schaye et al. ¹⁰⁸	Estados Unidos	IDEA	Ferramenta confiável e fácil para <i>feedback</i> em raciocínio clínico.
Schaye et al. ¹⁰⁹	Estados Unidos	NAR	Modelo de <i>Machine Learning</i> classifica qualidade de raciocínio clínico.
Shenoy et al. ¹¹¹	Estados Unidos	SOE	SOE para <i>feedback</i> formativo tem vantagens educacionais, notaram <i>fellows</i> e professores.
Simpkins et al. ¹¹⁵	Estados Unidos	PD-R	Rubrica auxiliou alunos e professores em <i>feedback</i> considerado significativo.
Sulaiman et al. ¹¹⁸	Emirados Árabes Unidos	CIVA	A correlação mais forte foi encontrada entre CIVA e OSCE ($r = 0,83$, $p < 0,001$).
Sutherland et al. ¹²⁰	Austrália	SCBD	SCBD tem consistência interna aceitável, mede construto relacionado.
Thammasitboon et al. ¹²¹	Estados Unidos	ART-R	ART-R demonstrou validade como ferramenta de avaliação com cinco domínios.
Waechter et al. ¹²⁴	Canadá	CRC-PD	Dois instrutores, com uma carga de trabalho viável, forneceram <i>feedback</i> detalhado em 376 casos, contribuindo para o aprimoramento do desempenho dos estudantes no raciocínio clínico.
Yudkowsky et al. ¹³¹	Estados Unidos	GCE	Avaliadores não clínicos pontuam notas com confiabilidade comparável a clínicos.
Zegota et al. ¹³³	Alemanha	KF	Validação do exame de raciocínio clínico utilizando teoria de resposta.

Fonte: Elaborado pelos autores.

Esses métodos, apesar de variados, ainda enfrentam desafios, como a complexidade de implementação, a necessidade de recursos e limitações na avaliação de habilidades mais complexas, evidenciando a importância de pesquisas futuras que possam validar e adaptar esses instrumentos para diferentes contextos na educação médica.

Este estudo apresenta limitações. O raciocínio clínico é amplo e complexo, tornando difícil abordá-lo completamente em uma única revisão. Além disso, por se tratar de uma revisão integrativa, não realizamos uma análise metodológica detalhada de cada artigo, o que pode impactar a validade das evidências. Apesar disso, o estudo contribui ao organizar e sintetizar os modelos de avaliação do raciocínio clínico.

CONSIDERAÇÕES FINAIS

A revisão destacou que métodos de avaliação como o TCS, o OSCE e QME desempenham papéis fundamentais na formação médica, cada um contribuindo de maneira única para o desenvolvimento e a avaliação do raciocínio clínico. O TCS se mostrou eficaz para medir a capacidade de raciocínio clínico em cenários de incerteza, enquanto o OSCE avalia a aplicação de conhecimento teórico em contextos simulados de prática clínica. A inteligência artificial traz avanços na avaliação médica, mas ainda carece de padronização e adaptação. Recomendamos que futuras pesquisas investiguem a integração dessas tecnologias para aprimorar a avaliação do raciocínio clínico, promovendo práticas mais eficazes e alinhadas com as demandas atuais da educação médica.

CONTRIBUIÇÃO DOS AUTORES

Fernando Tureck participou da concepção e supervisão da pesquisa, da coleta, análise e interpretação dos dados, da revisão crítica do manuscrito e da aprovação final da versão a ser publicada. Maurício Prätzel Ellwanger participou da coleta, análise e interpretação dos dados, da redação do artigo e da aprovação final da versão a ser publicada.

CONFLITO DE INTERESSES

Declaramos não haver conflito de interesses.

FINANCIAMENTO

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – Brasil.

DISPONIBILIDADE DE DADOS

Os dados de pesquisa estão disponíveis no corpo do documento.

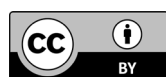
REFERÊNCIAS

1. Norman GR, Van Der Vleuten CPM, Newble DI, Dolmans DHJM, Mann KV, Rothman A, et al., organizers. International handbook of research in medical education. Dordrecht: Springer Netherlands; 2002. v. 7 [acesso em 11 set 2024]. Disponível em: <http://link.springer.com/10.1007/978-94-010-0462-6>.
2. Gruppen L. Clinical reasoning: defining it, teaching it, assessing it, studying it. *West J Emerg Med*. 2017;18(1):4-7.
3. Norman GR, Monteiro SD, Sherbino J, Ilgen JS, Schmidt HG, Mamede S. The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking. *Acad Med*. 2017;92(1):23-30.
4. Norman GR, Eva KW. Diagnostic error and clinical reasoning: diagnostic error and reasoning. *Med Educ*. 2010;44(1):94-100.
5. Connor DM, Durning SJ, Rencic JJ. Clinical reasoning as a core competency. *Acad Med*. 2020;95(8):1166-71.
6. Cruess RL, Cruess SR, Steinert Y. Amending Miller's pyramid to include professional identity formation. *Acad Med*. 2016;91(2):180-5.
7. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med J Assoc Am Med Coll*. 1990;65(9 Suppl):S63-67.
8. Witheridge A, Ferns G, Scott-Smith W. Revisiting Miller's pyramid in medical education: the gap between traditional assessment and diagnostic reasoning. *Int J Med Educ*. 2019;10:191-2.
9. Lee Y-S. OSCE for the medical licensing examination in Korea. 2008;24(12).
10. Bacchi S, Tan Y, Chim I, Dabarno M, Lubarsky S, Duggan P. Script concordance test examinations: student perception and study approaches. 2019;20(2).
11. Haug CJ, Drazen JM. Artificial intelligence and Machine Learning in clinical medicine. *N Engl J Med*. 2023;388(13):1201-8.
12. Mendes KDS, Silveira RCD, Galvão CM. Revisão integrativa: método de pesquisa para a incorporação de evidências na saúde e na enfermagem. *Texto Contexto Enferm*. 2008;17(4):758-64.
13. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan – a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210.
14. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*. 2021;(160).
15. Aagaard E, Teherani A, Irby DM. Effectiveness of the one-minute preceptor model for diagnosing the patient and the learner: proof of concept. *Acad Med*. 2004;79(1):42-9.
16. Amini M, Moghadami M, Kojuri J, Abbasi H, Abadi AAD, Molaee NA, et al. An innovative method to assess clinical reasoning skills: clinical reasoning tests in the second national medical science Olympiad in Iran. *BMC Res Notes*. 2011;4(1):418.
17. Amini M, Shahabi A, Moghadami M, Shams M, Anooshirvani A, Rostamipour H, et al. Psychometric characteristics of script concordance test (SCT) and its correlation with routine multiple choice question (MCQ) in internal medicine department. *Biomed Res*. 2017;28(19).
18. Anderson K, Peterson R, Tonkin A, Cleary E. The assessment of student reasoning in the context of a clinically oriented PBL program. *Med Teach*. 2008;30(8):787-94.
19. Artino AR, Cleary TJ, Dong T, Hemmer PA, Durning SJ. Exploring clinical reasoning in novices: a self-regulated learning microanalytic assessment approach. *Med Educ*. 2014;48(3):280-91.
20. Cohen Aubart F, Papo T, Hertig A, Renaud MC, Steichen O, Amoura Z, et al. Are script concordance tests suitable for the assessment of undergraduate students? A multicenter comparative study. *Rev Médecine Interne*. 2021;42(4):243-50.
21. Baker EA, Ledford CH, Fogg L, Way DP, Park YS. The IDEA assessment tool: assessing the reporting, diagnostic reasoning, and decision-making skills demonstrated in medical students' hospital admission notes. *Teach Learn Med*. 2015;27(2):163-73.
22. Baño G, Di Lalla S, Domínguez P, Noel Seoane M, Wainsztein R, Ossorio MF, et al. Evaluación de una prueba para valorar el razonamiento clínico en diferentes niveles de capacitación médica. *Rev Médica Chile*. 2011;139(4):455-61.

23. Behrens C, Morales V, Parra P, Hurtado A, Fernández R, Giaconi E, et al. Diseño e implementación de OSCE para evaluar competencias de egreso en estudiantes de medicina en un consorcio de universidades chilenas. *Rev Médica Chile*. 2018;146(10):1197-204.
24. Berger AJ, Gillespie CC, Tewksbury LR, Overstreet IM, Tsai MC, Kalet AL, et al. Assessment of medical student clinical reasoning by "lay" vs physician raters: inter-rater reliability using a scoring guide in a multidisciplinary objective structured clinical examination. *Am J Surg*. 2012;203(1):81-6.
25. Beullens J, Struyf E, Van Damme B. Diagnostic ability in relation to clinical seminars and extended-matching questions examinations. *Med Educ*. 2006;40(12):1173-9.
26. Beullens J, Struyf E, Van Damme B. Do extended matching multiple-choice questions measure clinical reasoning? *Med Educ*. 2005;39(4):410-7.
27. Bhardwaj P, Black EW, Fantone JC, Lopez M, Kelly M. Script concordance tests for formative clinical reasoning and problem-solving assessment in general pediatrics. *MedEdPORTAL*. 2022;11274.
28. Blazek M, Bess J, Hirshbein L, Chiang C, Sastry D, Ravindranath D. The Short-Answer Vignette Examination (SAVE): an assessment tool for the core psychiatry clerkship. *Acad Psychiatry*. 2014;38(5):615-8.
29. Blunk DI, Tonarelli S, Gardner C, Quest D, Pettitt D, Leiner M. Evaluating medical students' clinical reasoning in psychiatry using clinical and basic science concepts presented in session-level integration sessions. *Med Sci Educ*. 2019;29(3):819-24.
30. Boulouffe C, Doucet B, Muschart X, Charlin B, Vanpee D. Assessing clinical reasoning using a script concordance test with electrocardiogram in an emergency medicine clerkship rotation. *Emerg Med J*. 2014;31(4):313-6.
31. Capaldi VF, Durning SJ, Pangaro LN, Ber R. The clinical integrative puzzle for teaching and assessing clinical reasoning: preliminary feasibility, reliability, and validity evidence. *Mil Med*. 2015;180(Suppl 4):54-60.
32. Carrière B, Gagnon R, Charlin B, Downing S, Bordage G. Assessing clinical reasoning in pediatric emergency medicine: validity evidence for a script concordance test. *Ann Emerg Med*. 2009;53(5):647-52.
33. Charlin B, Gagnon R, Pelletier J, Coletti M, Abi-Rizk G, Nasr C, et al. Assessment of clinical reasoning in the context of uncertainty: the effect of variability within the reference panel. *Med Educ*. 2006;40(9):848-54.
34. Cheung JHH, Park YS, Aufderheide K, Holden J, Yudkowsky R. Optimizing clinical reasoning assessments with analytic and holistic ratings: examining the validity, reliability, and cost of a simplified patient note scoring procedure. *Acad Med*. 2022;97(11S):S15-21.
35. Coderre SP, Harasym P, Mandin H, Fick G. The impact of two multiple-choice question formats on the problem-solving strategies used by novices and experts. *BMC Med Educ*. 2004;4(1):23.
36. Compère V, Moriceau J, Gouin A, Guitard PG, Damm C, Provost D, et al. Residents in tutored practice exchange groups have better medical reasoning as measured by the script concordance test: a pilot study. *Anaesth Crit Care Pain Med*. 2015;34(1):17-21.
37. Compère V, Abily J, Moriceau J, Gouin A, Veber B, Dupont H, et al. Residents in tutored practice exchange groups have better medical reasoning as measured by script concordance test: a controlled, nonrandomized study. *J Clin Anesth*. 2016;32:236-41.
38. Cooke S, Lemay JF, Beran T. Evolutions in clinical reasoning assessment: the evolving script concordance test. *Med Teach*. 2017;39(8):828-35.
39. Couraud S, Girard P, Girard N, Souquet PJ, Coiffard B, Charlin B, et al. Évaluation des connaissances sur le dépistage du cancer par test de concordance de script. *Rev Mal Respir*. 2016;33(5):333-42.
40. Covin Y, Longo P, Wick N, Gavinski K, Wagner J. Empirical comparison of three assessment instruments of clinical reasoning capability in 230 medical students. *BMC Med Educ*. 2020;20(1):264.
41. Derakhshandeh Z, Amini M, Kojuri J, Dehbozorgian M. Psychometric characteristics of Clinical Reasoning Problems (CRPs) and its correlation with routine multiple choice question (MCQ) in Cardiology department. *J Adv Med Educ Prof*. 2018;6(1):37-42.
42. Dong L, Li W, Ma D, Lv C, Chen C. The evaluation and comparison on different types of resident doctors in training through DxR Clinician system. *J Phys Conf Ser*. 2020;1549(4):042076.
43. Ducos G, Lejus C, Sztark F, Nathan N, Fourcade O, Tack I, et al. The Script Concordance Test in anesthesiology: validation of a new tool for assessing clinical reasoning. *Anaesth Crit Care Pain Med*. 2015;34(1):11-5.
44. Duffy B, Tully R, Stanton AV. An online case-based teaching and assessment program on clinical history-taking skills and reasoning using simulated patients in response to the Covid-19 pandemic. *BMC Med Educ*. 2023;23(1):4.
45. Duggan P, Charlin B. Summative assessment of 5th year medical students' clinical reasoning by script concordance test: requirements and challenges. *BMC Med Educ*. 2012;12(1):29.
46. Durning SJ, Artino A, Boulet J, La Rochelle J, Van Der Vleuten C, Arze B, et al. The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. *Med Teach*. 2012;34(1):30-7.
47. Enser M, Moriceau J, Abily J, Damm C, Occhiali E, Besnier E, et al. Background noise lowers the performance of anaesthesiology residents' clinical reasoning when measured by script concordance: a randomised crossover volunteer study. *Eur J Anaesthesiol*. 2017;34(7):464-70.
48. Fida M, Kassab SE. Do medical student's scores using different assessment instruments predict their scores in clinical reasoning using a computer-based simulation? *Adv Med Educ Pract*. 2015;135.
49. Fink MC, Reitmeier V, Stadler M, Siebeck M, Fischer F, Fischer MR. Assessment of diagnostic competences with standardized patients versus virtual patients: experimental study in the context of history taking. *J Med Internet Res*. 2021;23(3):e21196.
50. Fleiszer D, Hoover ML, Posel N, Razek T, Bergman S. Development and validation of a tool to evaluate the evolution of clinical reasoning in trauma using virtual patients. *J Surg Educ*. 2018;75(3):779-86.
51. Fonteneau T, Billion E, Abdoul C, Le S, Hadchouel A, Drummond D. Simulation game versus multiple choice questionnaire to assess the clinical competence of medical students: prospective sequential trial. *J Med Internet Res*. 2020;22(12):e23254.
52. Fürstenberg S, Helm T, Prediger S, Kadmon M, Berberat PO, Harendza S. Assessing clinical reasoning in undergraduate medical students during history taking with an empirically derived scale for clinical reasoning indicators. *BMC Med Educ*. 2020;20(1):368.
53. Gagnon R, Charlin B, Coletti M, Sauve E, Van Der Vleuten C. Assessment in the context of uncertainty: how many members are needed on the panel of reference of a script concordance test? *Med Educ*. 2005;39(3):284-91.
54. Gagnon R, Charlin B, Lambert C, Carrière B, Van Der Vleuten C. Script concordance testing: more cases or more questions? *Adv Health Sci Educ*. 2008;14(3):367-75.
55. Gagnon R, Lubarsky S, Lambert C, Charlin B. Optimization of answer keys for script concordance testing: should we exclude deviant panelists, deviant responses, or neither? *Adv Health Sci Educ*. 2011;16(5):601-8.
56. Gallagher BD, Nematollahi S, Park H, Kurra S. Comparing students' clinical grades to scores on a standardized patient note-writing task. *J Gen Intern Med*. 2020;35(11):3243-7.
57. Gilkes L, Kealley N, Frayne J. Teaching and assessment of clinical diagnostic reasoning in medical students. *Med Teach*. 2022;44(6):650-6.
58. Iglesias Gómez C, González Sequeros O, Sarquella Brugada G, Padilla Del Rey ML, Salmerón Martínez D. Usefulness of SCT in detecting clinical reasoning deficits among pediatric professionals. *Prog Pediatr Cardiol*. 2021;61:101340.
59. Iglesias Gómez C, González Sequeros O, Salmerón Martínez D. Clinical reasoning evaluation using script concordance test in primary care residents. *An Pediatr (Engl Ed)*. 2022;97(2):87-94.
60. Goos M, Schubach F, Seifert G, Boeker M. Validation of undergraduate medical student script concordance test (SCT) scores on the clinical assessment of the acute abdomen. *BMC Surg*. 2016;16(1):57.
61. Goulet F, Jacques A, Gagnon R, Charlin B, Shabah A. Poorly performing physicians: does the script concordance test detect bad clinical reasoning? *J Contin Educ Health Prof*. 2010;30(3):161-6.
62. Groves M, O'Rourke P, Alexander H. The clinical reasoning characteristics of diagnostic experts. *Med Teach*. 2003;25(3):308-13.

63. Groves M, Dick ML, McColl G, Bilszta J. Analysing clinical reasoning characteristics using a combined methods approach. *BMC Med Educ.* 2013;13(1):144.
64. Hamui M, Ferreira JP, Torrents M, Torres F, Ibarra M, Ossorio MF, et al. Script Concordance Test: first nationwide experience in pediatrics. *Arch Argent Pediatr.* 2018;116(1):e151-5.
65. Haring CM, Klaarwater CCR, Bouwmans GA, Cools BM, Van Gurp PJM, Van Der Meer JWM, et al. Validity, reliability and feasibility of a new observation rating tool and a post encounter rating tool for the assessment of clinical reasoning skills of medical students during their internal medicine clerkship: a pilot study. *BMC Med Educ.* 2020;20(1):198.
66. Humbert AJ, Johnson MT, Miech E, Friedberg F, Grackin JA, Seidman PA. Assessment of clinical reasoning: a script concordance test designed for pre-clinical medical students. *Med Teach.* 2011;33(6):472-7.
67. Humbert AJ, Miech EJ. Measuring gains in the clinical reasoning of medical students: longitudinal results from a school-wide Script Concordance Test. *Acad Med.* 2014;89(7):1046-50.
68. Humbert AJ, Besinger B, Miech EJ. Assessing Clinical reasoning skills in scenarios of uncertainty: convergent validity for a Script Concordance Test in an emergency medicine clerkship and residency. *Acad Emerg Med.* 2011;18(6):627-34.
69. Huwendiek S, Reichert F, Duncker C, De Leng BA, Van Der Vleuten CPM, Muijtjens AMM, et al. Electronic assessment of clinical reasoning in clerkships: a mixed-methods comparison of long-menu key-feature problems with context-rich single best answer questions. *Med Teach.* 2017;39(5):476-85.
70. Im S, Kim DK, Kong HH, Roh HR, Oh YR, Seo JH. Assessing clinical reasoning abilities of medical students using clinical performance examination. *Korean J Med Educ.* 2016;28(1):35-47.
71. Iravani K, Amini M, Doostkam A, Dehbozorgian M. The validity and reliability of script concordance test in otolaryngology residency training. 4(2).
72. Jerant AF, Azari R. Validity of scores generated by a web-based multimedia simulated patient case software: a pilot study. *Acad Med.* 2004;79(8):805-11.
73. Kania RE, Verillaud B, Tran H, Gagnon R, Kazitani D, Huy PTB, et al. Online Script Concordance Test for clinical reasoning assessment in otorhinolaryngology: the association between performance and clinical experience. *Arch Otolaryngol Neck Surg.* 2011;137(8):751-5.
74. Kaur M, Singla S, Mahajan R. Script Concordance Test in pharmacology: maiden experience from a medical school in India. *J Adv Med Educ Prof.* 2020;8(3):115-20.
75. Kazour F, Richa S, Zoghbi M, El-Hage W, Haddad FG. Using the Script Concordance Test to evaluate clinical reasoning skills in psychiatry. *Acad Psychiatry.* 2017;41(1):86-90.
76. Kelly W, Durning S, Denton G. Comparing a Script Concordance Examination to a multiple-choice examination on a core internal medicine clerkship. *Teach Learn Med.* 2012;24(3):187-93.
77. Kim S, Choi I, Yoon BY, Kwon MJ, Choi S-jin, Kim SH, et al. Medical students' thought process while solving problems in 3 different types of clinical assessments in Korea: clinical performance examination, multimedia case-based assessment, and modified essay question. *J Educ Eval Health Prof.* 2019;16:10.
78. Kotwal S, Klimpl D, Tackett S, Kauffman R, Wright S. Documentation of clinical reasoning in admission notes of hospitalists: validation of the CRANAPL assessment rubric. *J Hosp Med.* 2019;14(12):746-53.
79. Kün-Darbois JD, Annweiler C, Lerolle N, Lebdaï S. Script concordance test acceptability and utility for assessing medical students' clinical reasoning: a user's survey and an institutional prospective evaluation of students' scores. *BMC Med Educ.* 2022;22(1):277.
80. Kunina-Habenicht O, Hautz WE, Knigge M, Spies C, Ahlers O. Assessing clinical reasoning (ASCLIRE): instrument development and validation. *Adv Health Sci Educ.* 2015;20(5):1205-24.
81. Lai JH, Cheng KH, Wu YJ, Lin CC. Assessing clinical reasoning ability in fourth-year medical students via an integrative group history-taking with an individual reasoning activity. *BMC Med Educ.* 2022;22(1):573.
82. Lambert C, Gagnon R, Nguyen D, Charlin B. The script concordance test in radiation oncology: validation study of a new tool to assess clinical reasoning. *Radiat Oncol.* 2009;4(1):7.
83. Lineberry M, Hornos E, Pleguezuelos E, Mella J, Brailovsky C, Bordage G. Experts' responses in script concordance tests: a response process validity investigation. *Med Educ.* 2019;53(7):710-22.
84. Loh KW, Rotgans JI, Tan K, Tan NCK. Predictors of clinical reasoning in neurological localisation: a study in internal medicine residents. *Asia Pac Sch.* 2020;5(3):54-61.
85. Lukas RV, Blood A, Park YS, Brorson JR. Assessment of neurological clinical management reasoning in medical students. *J Clin Neurosci.* 2014;21(6):919-22.
86. Talvard M, Olives JP, Mas E. Évaluation des étudiants en médecine lors de leur stage en gastro-entérologie pédiatrique par un test de concordance de script. *Arch Pédiatrie.* 2014;21(4):372-6.
87. Marie I, Sibert L, Roussel F, Hellot MF, Lechevallier J, Weber J. Le test de concordance de script: un nouvel outil d'évaluation du raisonnement et de la compétence clinique en médecine interne? *Rev Médecine Interne.* 2005;26(6):501-7.
88. Mathieu S, Couderc M, Glace B, Tournadre A, Malochet-Guinamand S, Pereira B, et al. Construction and utilization of a script concordance test as an assessment tool for dcem3 (5th year) medical students in rheumatology. *BMC Med Educ.* 2013;13(1):166.
89. Meterissian S, Zabolotny B, Gagnon R, Charlin B. Is the script concordance test a valid instrument for assessment of intraoperative decision-making skills? *Am J Surg.* 2007;193(2):248-51.
90. Monajemi A, Soltani A, Hadadgar A, Hadizadeh F, Adibi P, Akbari R, et al. A comprehensive test of clinical reasoning for medical students: an olympiad experience in Iran. *J Educ Health Promot.* 2012;1(1):10.
91. Mzoughi K, Zairi I, Kedous MA, Mhamdi SE, Dhiab MB, Mghaieth F. Le test de concordance de script comme outil d'évaluation sanctionnante en cardiologie script concordance test as a sanctionnal evaluation in cardiology.
92. Nazim S, Talati J, Pinjani S, Biyabani SR, Ather M, Norcini J. Assessing clinical reasoning skills using Script Concordance Test (SCT) and extended matching questions (EMQs): a pilot for urology trainees. *J Adv Med Educ Prof.* 2019;7(1). doi: <https://doi.org/10.30476/jamp.2019.41038>
93. Nomura O, Itoh T, Mori T, Ihara T, Tsuji S, Inoue N, et al. Creating clinical reasoning assessment tools in different languages: adaptation of the pediatric emergency medicine Script Concordance Test to Japanese. *Front Med.* 2021;8:765489.
94. Nouh T, Boutros M, Gagnon R, Reid S, Leslie K, Pace D, et al. The script concordance test as a measure of clinical reasoning: a national validation study. *Am J Surg.* 2012;203(4):530-4.
95. Omega A, Wijaya Ramlan AA, Soenarto RF, Heriwardito A, Sugiarto A. Assessing clinical reasoning in airway related cases among anesthesiology fellow residents using Script Concordance Test (SCT). *Med Educ Online.* 2022;27(1):2135421.
96. Park WB, Kang SH, Myung SJ, Lee YS. Does Objective Structured Clinical Examinations Score reflect the clinical reasoning ability of medical students? *Am J Med Sci.* 2015;350(1):64-7.
97. Peterson BD, Magee CD, Martindale JR, Dreicer JJ, Mutter MK, Young G, et al. REACT: Rapid Evaluation Assessment of Clinical Reasoning Tool. *J Gen Intern Med.* 2022;37(9):2224-9.
98. Petrucci AM, Nouh T, Boutros M, Gagnon R, Meterissian SH. Assessing clinical judgment using the Script Concordance Test: the importance of using specialty-specific experts to develop the scoring key. *Am J Surg.* 2013;205(2):137-40.
99. Peyrony O, Hutin A, Truchot J, Borie R, Calvet D, Albaladejo A, et al. Impact of panelists' experience on script concordance test scores of medical students. *BMC Med Educ.* 2020;20(1):313.
100. Piovezan RD, Custódio O, Cendoroglo MS, Batista NA, Lubarsky S, Charlin B. Assessment of undergraduate clinical reasoning in geriatric medicine: application of a Script Concordance Test. *J Am Geriatr Soc.* 2012;60(10):1946-50.

101. Pottier P, Castillo JM, Boet S, Le Pabic E, Hardouin JB. Validation d'une grille d'évaluation des compétences cliniques utilisée par des patients standardisés. *Rev Médecine Interne*. 2016;37(12):802-10.
102. Power A, Lemay JF, Cooke S. Justify your answer: the role of written think aloud in Script Concordance Testing. *Teach Learn Med*. 2017;29(1):59-67.
103. Rajeswaran V, Devine L, Lorens E, Robertson S, Huszti E, Panisko DM. Types of clinical reasoning in a summative clerkship oral examination. *Med Teach*. 2022;44(6):657-63.
104. Reinert A, Berlin A, Swan-Sein A, Nowygrod R, Fingeret A. Validity and reliability of a novel written examination to assess knowledge and clinical decision making skills of medical students on the surgery clerkship. *Am J Surg*. 2014;207(2):236-42.
105. Ruiz JG, Tunuguntla R, Charlin B, Ouslander JG, Symes SN, Gagnon R, et al. The Script Concordance Test as a measure of clinical reasoning skills in geriatric urinary incontinence. *J Am Geriatr Soc*. 2010;58(11):2178-84.
106. Sadeghi A, Ali Asgari A, Moulai N, Mohammadkarimi V, Delavari S, Amini M, et al. Combination of different clinical reasoning tests in a national exam. *J Adv Med Educ Prof*. 2019;7(4):230-4.
107. Schaubert SK, Hautz SC, Kämmer JE, Stroben F, Hautz WE. Do different response formats affect how test takers approach a clinical reasoning task? An experimental study on antecedents of diagnostic accuracy using a constructed response and a selected response format. *Adv Health Sci Educ*. 2021;26(4):1339-54.
108. Schaye V, Miller L, Kudlowitz D, Chun J, Burk-Rafel J, Cocks P, et al. Development of a clinical reasoning documentation assessment tool for resident and fellow admission notes: a shared mental model for feedback. *J Gen Intern Med*. 2021;37(3):507-12.
109. Schaye V, Guzman B, Burk-Rafel J, Marin M, Reinstein I, Kudlowitz D, et al. Development and validation of a Machine Learning Model for automated assessment of resident clinical reasoning documentation. *J Gen Intern Med*. 2022;37(9):2230-8.
110. See KC, Tan KL, Lim TK. The script concordance test for clinical reasoning: re-examining its utility and potential weakness. *Med Educ*. 2014;48(11):1069-77.
111. Shenoy RV, Newbern D, Cooke DW, Chia DJ, Panagiotakopoulos L, DiVall S, et al. The structured oral examination: a method to improve formative assessment of fellows in pediatric endocrinology. *Acad Pediatr*. 2022;22(7):1091-6.
112. Sibert L, Darmoni SJ, Dahamna B, Hellot MF, Weber J, Charlin B. On line clinical reasoning assessment with Script Concordance Test in urology: results of a French pilot study. *BMC Med Educ*. 2006;6(1):45.
113. Silberman EK, Ramesh S, Adler D, Sargent J, Moore T, Blanco M. Assessing residents' skills in psychiatric reasoning: the tufts test of formulation and treatment planning. *Acad Psychiatry*. 2020;44(6):701-8.
114. Sim JH, Aziz YFA, Mansor A, Vijayanathan A, Foong CC, Vadivelu J. Students' performance in the different clinical skills assessed in OSCE: what does it reveal? *Med Educ Online*. 2015;20(1):26185.
115. Min Simpkins AA, Koch B, Spear-Ellinwood K, St. John P. A developmental assessment of clinical reasoning in preclinical medical education. *Med Educ Online*. 2019;24(1):1591257.
116. Steinberg E, Cowan E, Lin M, Sielicki A, Warrington S. Assessment of emergency medicine residents' clinical reasoning: validation of a Script Concordance Test. *West J Emerg Med*. 2020;21(4) [acesso em 7 fev 2024]. Disponível em: <https://escholarship.org/uc/item/92b825mf>.
117. Subra J, Chicoulaa B, Stillmunkès A, Mesthé P, Oustric S, Rougé Bugat ME. Reliability and validity of the script concordance test for postgraduate students of general practice. *Eur J Gen Pract*. 2017;23(1):209-14.
118. Sulaiman ND, Hamdy H. Assessment of clinical competencies using clinical images and videos "CIVA". *BMC Med Educ*. 2013;13(1):78.
119. Surry LT, Torre D, Trowbridge RL, Durning SJ. A mixed-methods exploration of cognitive dispositions to respond and clinical reasoning errors with multiple choice questions. *BMC Med Educ*. 2018;18(1):277.
120. Sutherland RM, Reid KJ, Chiavaroli NG, Smallwood D, McColl GJ. assessing diagnostic reasoning using a standardized case-based discussion. *J Med Educ Curric Dev*. 2019;6:238212051984941.
121. Thammasitboon S, Sur M, Rencic JJ, Dhaliwal G, Kumar S, Sundaram S, et al. Psychometric validation of the reconstructed version of the assessment of reasoning tool. *Med Teach*. 2021;43(2):168-73.
122. Van Den Broek WES, Van Asperen MV, Ten Cate OTHJ, Custers E, Valk GD. Effects of two different instructional formats on scores and reliability of a script concordance test. *Perspect Med Educ*. 2012;1(3):119-28.
123. Volkan K, Simon SR, Baker H, Todres ID. Psychometric structure of a comprehensive Objective Structured Clinical Examination: a factor analytic approach. *Adv Health Sci Educ*. 2004;9(2):83-92.
124. Waechter J, Allen J, Lee CH, Zwaan L. Development and pilot testing of a data-rich clinical reasoning training and assessment tool. *Acad Med*. 2022;97(10):1484-8.
125. Wan MSH, Tor E, Hudson JN. Construct validity of script concordance testing: progression of scores from novices to experienced clinicians. *Int J Med Educ*. 2019;10:174-9.
126. Wan MSH, Tor E, Hudson JN. Examining response process validity of script concordance testing: a think-aloud approach. *Int J Med Educ*. 2020;11:127-35.
127. Williams RG, Klamen DL, White CB, Petrusa E, Fincher RME, Whitfield CF, et al. Tracking development of clinical reasoning ability across five medical schools using a Progress Test. *Acad Med*. 2011;86(9):1148-54.
128. Wilson AB, Pike GR, Humbert AJ. Testing for multilevel dimensionality: a higher-order factor analysis of a Script Concordance Test. *Med Sci Educ*. 2015;25(4):439-46.
129. Wilson AB, Pike GR, Humbert AJ. Analyzing Script Concordance Test scoring methods and items by difficulty and type. *Teach Learn Med*. 2014;26(2):135-45.
130. Yassin BAG, Almothaffar A, Aladhadh M, Hussein MF, Ghafour ZA, Gorial FI. Studying clinical reasoning of internal medicine residents in medical city campus using Script Concordance Test.
131. Yudkowsky R, Hyderi A, Holden J, Kiser R, Stringham R, Gangopadhyaya A, et al. Can nonclinician raters be trained to assess clinical reasoning in postencounter patient notes? *Acad Med*. 2019;94(11S):S21-7.
132. Zavaleta-Hernández S, Cerón-Rodríguez M, Olivar-López V, Espinoza R, Rizzoli-Córdoba A. Validation of the Script Concordance Test as an instrument to assess clinical reasoning of residents in pediatric emergency medicine in Mexico. *Bol Med Hosp Infant Mex*. 2011;68.
133. Zegota S, Becker T, Hagmayer Y, Raupach T. Using item response theory to appraise key feature examinations for clinical reasoning. *Med Teach*. 2022;44(11):1253-9.
134. Zia S, Obaid M, Ahmed J, Qazi S. Key-feature questions for assessment of clinical reasoning: reliable and valid or not? *RAWAL Med J*. 2019;44(1):189-191.
135. Zuo T, Sun B, Guan X, Zheng B, Qu B. Evidence of construct validity of computer-based tests for clinical reasoning: instrument validation study. *JMIR Serious Games*. 2021;9(4):e17670.
136. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. 2011;2:53-5.
137. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA 2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Syst Rev*. 2022;18(2):e1230.
138. Daniel M, Rencic J, Durning SJ, Holmboe E, Heist B, Lubarsky S, et al. Clinical reasoning assessment methods: a scoping review and practical guidance. *Acad Med*. 2019;94(6):901-12.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.