



ESTUDO DE REVISÃO

Do exame à arquitetura avaliativa: uma revisão crítica sobre o papel do OSCE na educação médica contemporânea

Haraldo Cesar Saletti Filho – cesar.saletti@gmail.com

Resumo

O *Objective Structured Clinical Examination* (OSCE) consolidou-se, ao longo das últimas cinco décadas, como um dos métodos centrais de avaliação da competência clínica. Contudo, sua adoção global tem sido acompanhada por debates que ultrapassam a dimensão técnica, alcançando questões históricas, pedagógicas, éticas e epistemológicas. Este artigo apresenta uma revisão crítica que integra diferentes correntes teóricas e empíricas, examinando o desenvolvimento do OSCE desde sua formulação por Harden e Gleeson até seu reposicionamento em ecossistemas avaliativos contemporâneos. A revisão, de natureza narrativa e interpretativa, organiza-se em cinco eixos: (1) gênese histórica e problemas iniciais do exame clínico tradicional; (2) padronização e difusão do modelo OSCE; (3) críticas pedagógicas e sociológicas que problematizam sua estrutura performativa; (4) vieses e desafios de justiça avaliativa; e (5) perspectivas futuras, incluindo avaliação programática, inteligência artificial e novas métricas de qualidade. Os achados indicam que, embora o OSCE permaneça valioso como fonte estruturada de evidências, sua centralidade como exame definitivo é progressivamente substituída por abordagens sistêmicas baseadas em triangulação de métodos, julgamento profissional qualificado e governança avaliativa. Conclui-se que o OSCE deve ser compreendido como componente de arquiteturas avaliativas complexas, cuja efetividade depende menos da eliminação da subjetividade e mais de sua integração responsável, transparente e pedagógica.

Palavras-chave: Exame Clínico Estruturado Objetivo, Educação Médica, Competência Clínica, Avaliação Programática, Inteligência Artificial

Summary

The Objective Structured Clinical Examination (OSCE) has become one of the most influential methods for assessing clinical competence in medical education. However, the historical evolution of the OSCE reveals tensions that go beyond psychometric refinement, involving pedagogical, sociocultural, ethical, and epistemological dimensions. This article provides a critical narrative review that traces the OSCE from its origins—developed in response to the limitations of traditional clinical examinations—to its contemporary repositioning within complex assessment systems. Five analytical axes guide the review: (1) the historical rationale and structural problems that motivated the creation of the OSCE; (2) the model's consolidation and dissemination as a standardized approach to clinical assessment; (3) pedagogical and sociological critiques that highlight its performative nature; (4) issues of bias and fairness that challenge its promise of objectivity; and (5) future perspectives, including programmatic assessment, artificial intelligence, and advanced quality metrics. The analysis indicates that while the OSCE remains valuable as a structured sampling tool, its role as a standalone high-stakes examination is increasingly replaced by system-level approaches that integrate multiple sources of evidence and qualified professional judgment. The OSCE is thus best understood as a component within broader assessment architectures, whose effectiveness relies on transparency, triangulation, and responsible incorporation of subjectivity.

Key words: Objective Structured Clinical Examination, Medical Education, Clinical Competence, Programmatic Assessment, Artificial Intelligence

Introdução

O *Objective Structured Clinical Examination (OSCE)*, formulado por Harden e colaboradores no final do século XX, consolidou-se como uma das estratégias mais influentes para a avaliação da competência clínica na educação médica (1,2). Sua adoção global decorreu da necessidade de instrumentos capazes de reduzir a variabilidade entre avaliadores, aumentar a transparência dos processos avaliativos e oferecer maior consistência na tomada de decisões educacionais (3). O modelo introduziu um padrão de exame estruturado, baseado em estações curtas, tarefas delimitadas e critérios explícitos de desempenho, diferenciando-se significativamente do exame clínico tradicional.

Com o tempo, entretanto, a literatura passou a documentar tensões inerentes à padronização e à estruturação que caracterizam o OSCE. Pesquisas demonstram que a busca por maior controle metodológico não elimina influências contextuais, socioculturais e interacionais que moldam o desempenho dos estudantes e a interpretação da competência clínica (4). Questões relativas à validade, autenticidade, efeitos pedagógicos, fragmentação das tarefas e variabilidade residual entre avaliadores tornaram-se centrais no debate acadêmico, ampliando a compreensão dos limites e implicações do modelo.

Transformações recentes na educação médica — incluindo o desenvolvimento de arquiteturas avaliativas integradas, o avanço da avaliação programática e a incorporação de tecnologias digitais aplicadas ao acompanhamento longitudinal do desempenho —

também contribuíram para reposicionar o OSCE em ecossistemas avaliativos mais complexos (5). Nesse cenário, emergem discussões sobre suas potencialidades, limitações e condições de pertinência em modelos avaliativos que valorizam o julgamento profissional, a autenticidade das tarefas e a triangulação entre múltiplas fontes de evidência.

Apesar da amplitude do debate, a literatura permanece fragmentada, com análises que tendem a focalizar aspectos isolados do OSCE, como confiabilidade, validade, vieses avaliativos ou impacto pedagógico. Essa fragmentação dificulta a compreensão integrada da evolução conceitual do modelo e de suas implicações para a formação médica contemporânea.

Diante desse panorama, o presente trabalho tem como objetivo examinar criticamente a trajetória conceitual e prática do OSCE, articulando seus fundamentos históricos, críticas e perspectivas futuras no contexto da educação médica.

Método

O estudo foi conduzido por meio de uma revisão crítica de base conceitual, orientada pela análise interpretativa de textos clássicos e contemporâneos sobre o OSCE. Essa abordagem permite examinar a evolução histórica do modelo, incluindo suas fundamentações pedagógicas e psicométricas, bem como as críticas e perspectivas que emergiram em diferentes períodos de sua consolidação.

A construção analítica ocorreu em três movimentos. Primeiramente, identificaram-se os principais eixos temáticos da literatura especializada, abrangendo desde as contribuições fundacionais de Harden até debates recentes sobre autenticidade, vieses avaliativos, justiça e integração do OSCE em sistemas programáticos de avaliação. Em seguida, tais eixos foram organizados em blocos interpretativos, de modo a preservar a coerência interna de cada dimensão analisada e a oferecer uma leitura longitudinal da evolução conceitual do método. Por fim, produziu-se uma conclusão no formato de síntese integrativa que articula as convergências, tensões e desdobramentos presentes nos diferentes blocos, permitindo compreender o papel contemporâneo do OSCE em ecossistemas avaliativos complexos.

A escolha por essa abordagem fundamenta-se em uma justificativa epistemológica alinhada à natureza complexa do OSCE. Por não se tratar apenas de um instrumento técnico, mas de um artefato educacional carregado de pressupostos normativos, pedagógicos e psicométricos, o OSCE demanda métodos interpretativos capazes de captar suas dimensões conceituais e históricas. Abordagens exclusivamente descritivas ou centradas na síntese de achados empíricos tendem a difundir uma ilusão de estabilidade conceitual, quando, de fato, conceitos como validade, autenticidade, competência clínica e justiça avaliativa são construções teóricas sujeitas a disputas e reformulações ao longo do tempo.

A revisão crítica, nesse sentido, oferece um arcabouço adequado para analisar a trajetória do OSCE e suas implicações. Essa abordagem permite interpretar o modelo como um objeto situado, influenciado por contextos institucionais, culturais e

epistemológicos, e não apenas como um procedimento técnico isolado. Ao privilegiar a reconstrução argumentativa, a comparação entre tradições teóricas e a identificação de tensões estruturais, a revisão crítica possibilita compreender o OSCE em sua materialidade operacional e em suas bases interpretativas, alinhando-se à literatura que defende abordagens analíticas ampliadas para a avaliação da prática clínica.

Por não se tratar de revisão sistemática, não foram aplicados critérios rígidos de busca, seleção ou extração de dados. O foco da investigação reside na análise conceitual e interpretativa, orientada pela relevância teórica dos textos e pela capacidade de cada fonte contribuir para a compreensão da evolução e das implicações do modelo.

Resultados e Discussão

Bloco 1 - Emergência histórica do OSCE e problemas estruturais iniciais

O desenvolvimento do OSCE no final da década de 1970 deve ser compreendido no contexto de profundas transformações epistemológicas e políticas da educação médica. Harden e Gleeson formularam a proposta bem no momento em que crescia a pressão por mecanismos de avaliação que demonstrassem rigor, transparência e alinhamento a padrões regulatórios emergentes (1,2). No ambiente anglo-saxão, marcado pela expansão de sistemas de *accountability* e pela busca por instrumentos comparáveis entre instituições, o OSCE respondeu a demandas que extrapolavam a esfera pedagógica, atravessando dimensões institucionais, administrativas e sociopolíticas.

Antes do surgimento do OSCE, os exames clínicos tradicionais baseavam-se em interações longas com um único paciente, conduzidas por examinadores que operavam com grande liberdade interpretativa. Essa configuração produzia três problemas amplamente reconhecidos na literatura: variabilidade entre avaliadores, baixa amostragem de competências e imprevisibilidade estrutural. A ausência de critérios explícitos e a dependência do caso clínico disponível no dia do exame geravam um cenário em que estudantes distintos eram avaliados em condições incomparáveis, dificultando análises de validade e confiabilidade.

A proposta de Harden buscou resolver essa instabilidade por meio de uma lógica de estruturação, que incorporava princípios psicométricos derivados do paradigma da mensuração educacional. A introdução de estações curtas, tarefas delimitadas, tempo fixado e *checklists* operacionalizava um modelo no qual o desempenho seria decomposto em unidades discretas de observação. Essa decomposição permitia mapear com maior precisão quais domínios estavam sendo avaliados, aumentando a reprodutibilidade tanto do processo quanto dos resultados, contemplando, respectivamente, confiabilidade e validade (3).

No entanto, mesmo em sua formulação inicial, já se percebiam tensões inerentes ao método e que acompanhariam o OSCE por décadas. A primeira delas refere-se à fragmentação. A quebra do desempenho clínico em micro tarefas objetiváveis conferia clareza operacional, mas, simultaneamente, reduzia a complexidade da prática clínica

real. Atos clínicos — naturalmente contínuos, interativos e contextualmente modulados — eram convertidos em unidades discretas, avaliáveis e, sobretudo, mensuráveis. Essa escolha metodológica introduziu ganhos psicométricos, mas produziu perdas significativas no domínio da autenticidade, inaugurando um debate que persiste até hoje.

Outro problema estrutural inicial reside na relação entre objetividade e controle. O OSCE foi frequentemente apresentado como um instrumento capaz de minimizar a subjetividade do avaliador, um argumento alinhado ao espírito de sua época. Entretanto, análises posteriores demonstraram que a objetividade no OSCE é, em grande medida, um efeito institucional, e não uma característica intrínseca do exame. O controle das tarefas, a padronização dos *scripts* de pacientes simulados e o uso de *checklists* não eliminam a interpretação humana; apenas deslocam seu lócus. A subjetividade deixa de se manifestar principalmente na interação avaliador-estudante e passa a atuar na concepção das estações, na escolha dos critérios e na construção dos padrões de resposta.

Além disso, o OSCE emergiu como parte de um movimento de racionalização do currículo. O exame oferecia às escolas médicas um modelo de avaliação passível de ser escalado, replicado e, sobretudo, defendido perante instâncias externas. Essa capacidade de demonstrar rigor metodológico conferiu ao OSCE um papel estratégico nas respostas institucionais à crescente demanda por responsabilização social. Assim, desde sua origem, o OSCE não deve ser compreendido apenas como inovação pedagógica, mas como instrumento de gestão acadêmica.

Em síntese, o período de emergência do OSCE revela um método que nasce simultaneamente como solução e como problema. Se, por um lado, introduziu uma linguagem comum para avaliação clínica, ampliando a comparabilidade e a consistência dos julgamentos, por outro inaugurou um conjunto de tensões que moldariam as críticas subsequentes: fragmentação, artificialidade, deslocamento da subjetividade e redefinição do que conta como competência clínica observável.

Bloco 2 - Padronização, consolidação e expansão institucional do OSCE

A consolidação do OSCE ao longo das décadas de 1980 e 1990 constituiu um movimento complexo de institucionalização que extrapolou os limites de uma mera inovação técnica. Seu crescimento internacional deve ser entendido como resposta a um conjunto de pressões estruturais — regulatórias, psicométricas e organizacionais — que buscavam dar maior legitimidade às avaliações clínicas em um período marcado pelo fortalecimento das agendas de *accountability* e pela demanda social por maior transparência nos processos de formação médica (4–6). Nesse cenário, a padronização assumiu papel central, oferecendo uma alternativa tecnicamente mais controlável aos exames clínicos tradicionais. Ressalta-se ainda que, a partir disso, acabou redefinindo a própria compreensão do que seria avaliado, delimitando operacionalmente e validando culturalmente o escopo da competência clínica esperada.

Os primeiros estudos que sustentaram a expansão do OSCE enfatizaram a capacidade do modelo de reduzir a variabilidade entre avaliadores e mitigar a dependência dos casos fortuitos que caracterizavam o exame clínico convencional (1–3). A

padronização dos cenários, dos roteiros dos pacientes simulados e das listas de verificação parecia oferecer uma resposta metodologicamente robusta às críticas dirigidas às avaliações anteriores, especialmente no que se refere à confiabilidade e à consistência de decisões de alta consequência. A replicabilidade interinstitucional reforçou esse movimento, tornando o OSCE um instrumento privilegiado para escolas que buscavam atender exigências de agências acreditadoras e órgãos reguladores, fortalecendo sua legitimidade institucional.

No entanto, a promessa de objetividade sustentada pela padronização logo se mostrou tensionada pela própria dinâmica do método. Pesquisas subsequentes evidenciaram que, embora a estruturação das tarefas aumentasse indicadores psicométricos tradicionais, não eliminava, e por vezes redistribuía, fontes de subjetividade presentes no processo avaliativo (7–9). A variabilidade entre avaliadores, diferenças no comportamento de pacientes padronizados e interpretações divergentes de critérios considerados explícitos mostraram que o controle pretendido pelo OSCE não derivava de uma neutralidade intrínseca, mas da construção de um ambiente altamente regulado que ainda dependia, em grande medida, de julgamentos humanos. A objetividade, portanto, revelou-se menos uma propriedade ontológica do instrumento e mais um efeito institucional produzido por práticas específicas de organização e condução do exame.

Esse processo de expansão também evidenciou que a padronização funcionava como uma tecnologia de governança educacional (10–12). Ao permitir maior previsibilidade, controle e auditabilidade, o OSCE passou a orientar tanto a avaliação quanto o currículo. As competências selecionadas para compor as estações, os comportamentos considerados observáveis e os critérios de desempenho explicitados nas rubricas tornaram-se referências normativas para o ensino. Sendo assim, o modelo transcendia a função avaliativa: ele prescrevia expectativas, moldava comportamentos e influenciava a cultura pedagógica das instituições. O exame adquiria, portanto, uma dimensão regulatória, contribuindo para a homogeneização de práticas avaliativas entre escolas e países, o que consolidou sua posição como padrão internacional.

Ao mesmo tempo, essa institucionalização acentuou uma contradição inerente ao modelo. Para viabilizar a padronização e a comparabilidade entre candidatos, tornou-se necessário simplificar o constructo de competência clínica, reduzindo-o a componentes observáveis e mensuráveis. Essa redução operativa facilitou a quantificação do desempenho, mas distanciou o exame das complexidades da prática clínica nos ambientes de assistência à saúde, que envolve continuidade, julgamento contextual e ações integradas que dificilmente se enquadram em tarefas discretas de curta duração. Essa simplificação sustentou a expansão do OSCE, mas, ao mesmo tempo, preparou o terreno para críticas pedagógicas posteriores, que questionaram sua autenticidade e seus efeitos formativos. A força institucional do exame, portanto, emergiu simultaneamente de sua capacidade de organizar o campo e de sua limitação em representar, de modo integral, a natureza dinâmica e situada da prática clínica.

3. Crítica pedagógica: fragmentação, autenticidade e impacto formativo (texto corrido e aprofundado)

À medida que o OSCE se consolidava como método hegemônico de avaliação clínica, emergiu um conjunto crescente de críticas pedagógicas que problematizavam os efeitos

formativos e epistemológicos de sua estrutura. A literatura passou a apontar que, embora o exame representasse um avanço em relação à inconsistência dos métodos tradicionais, sua lógica de decomposição do desempenho em unidades discretas acarretava efeitos colaterais significativos, especialmente relacionados à **fragmentação da competência**, à **redução da autenticidade** e à **indução de estratégias performáticas** por parte dos estudantes (7–9).

A fragmentação constitui uma das críticas centrais. Ao dividir o desempenho clínico em microtarefas observáveis, o OSCE busca aumentar a confiabilidade do julgamento, mas, ao fazê-lo, desarticula elementos que, na prática real, são constitutivamente integrados — como raciocínio clínico, comunicação, capacidades relacionais, tomada de decisão e manejo da incerteza. Essa decomposição permite avaliar habilidades específicas com maior precisão técnica, mas o custo é a perda de continuidade e de coerência entre componentes que, em contextos reais, operam de forma simultânea e interdependente. Diversos autores têm argumentado que o exame, ao priorizar a avaliabilidade, produz representações empobrecidas da prática clínica, privilegiando comportamentos visíveis em detrimento de processos cognitivos complexos que sustentam o desempenho profissional (7).

Outro eixo recorrente de crítica refere-se à **autenticidade**. Apesar dos esforços para aproximar o ambiente de exame de cenários clínicos, a artificialidade inerente ao formato — tempo limitado, tarefas altamente previsíveis, pacientes padronizados treinados para respostas específicas — compromete a validade ecológica do OSCE. A literatura mostra que estudantes frequentemente reconhecem padrões de estações e desenvolvem repertórios performáticos orientados para cumprir checklists, e não para resolver problemas clínicos de forma contextualizada. Tal dinâmica pode gerar uma dissociação entre desempenho no exame e desempenho na prática real, uma vez que o OSCE tende a privilegiar comportamentos estereotipados que se encaixam nas expectativas avaliativas, e não necessariamente a expressão genuína de competência clínica integrada (8).

A crítica à autenticidade está estreitamente associada ao fenômeno de **aprendizagem estratégica**. A alta previsibilidade estrutural do OSCE induz estudantes a direcionarem seus esforços de estudo para a otimização da performance nas estações, e não para o aprofundamento substantivo das competências clínicas. Como consequência, o exame, ao invés de atuar como dispositivo formativo, pode reforçar concepções instrumentais e reducionistas da prática médica. Em alguns contextos, essa dinâmica gerou a percepção de que o OSCE se tornou um exame “ensaiável”, em que estudantes são treinados para reproduzir sequências de comportamentos padronizados que atendem aos critérios avaliativos, mas não necessariamente refletem raciocínio clínico aprofundado ou sensibilidade ao contexto do paciente (9).

O impacto emocional associado ao OSCE também foi objeto de investigação. Por ser frequentemente utilizado como avaliação de alta consequência, o exame produz níveis elevados de ansiedade, que podem interferir significativamente no desempenho. Estudos demonstram que a pressão temporal, a presença simultânea de múltiplos avaliadores e a consciência da artificialidade do ambiente contribuem para um clima de tensão que não reproduz adequadamente as condições reais da prática (8). Além disso, o efeito psicológico do exame pode reforçar desigualdades estruturais, favorecendo estudantes com perfis comunicacionais mais performáticos ou com maior familiaridade com

ambientes competitivos, mesmo quando tais características não são indicadores diretos de competência clínica.

Do ponto de vista curricular, a centralidade atribuída ao OSCE contribuiu, em alguns contextos, para um processo de **reorientação do ensino em direção à avaliabilidade**. A estrutura das estações, suas rubricas e seus critérios se tornaram referências tácitas para o que os estudantes interpretam como relevante aprender. Assim, quando o OSCE assume posição hegemônica, corre-se o risco de que o currículo seja moldado por aquilo que é mais facilmente mensurável, e não pelo que é pedagogicamente mais significativo. Essa dinâmica desloca o foco do julgamento profissional integrado para a aquisição de microcompetências observáveis, dificultando a formação de clínicos capazes de manejar incertezas, integrar informações e agir de forma contextualizada — dimensões cujo valor pedagógico excede a capacidade de mensuração do OSCE.

Por fim, a crítica pedagógica expõe uma dimensão epistemológica mais profunda: a suposição de que a competência clínica pode ser decomposta em componentes discretos avaliáveis, quando, na realidade, trata-se de um constructo complexo, situado e relacional. O OSCE, ao priorizar a clareza técnica da medição, tende a representar a prática médica como soma de partes, e não como atividade interpretativa e contínua. Essa visão fragmentada, ainda que útil para fins operacionais, cria um descompasso entre o que é ensinado, o que é avaliado e o que a prática real exige do profissional em formação.

4. Vieses e justiça avaliativa: limites da padronização (texto corrido e aprofundado)

À medida que o OSCE se consolidou como método dominante na avaliação clínica, a literatura passou a incorporar análises mais sensíveis às dimensões éticas, socioculturais e relacionais que atravessam o processo avaliativo. Embora a padronização tenha sido concebida como mecanismo para aumentar a equidade e minimizar distorções subjetivas, sucessivas investigações demonstraram que o modelo permanece permeável a múltiplos tipos de vieses, muitos dos quais operam de modo sutil e estrutural (10–12). O debate deslocou-se, portanto, de uma visão centrada na confiabilidade técnica para uma compreensão mais ampla da avaliação como prática social, situada e suscetível a condicionamentos institucionais, identitários e cognitivos.

Os estudos que examinam a variabilidade entre avaliadores revelam que, mesmo em ambientes altamente controlados, persiste uma heterogeneidade significativa nos julgamentos de desempenho. Essa variabilidade não decorre apenas de diferenças individuais, mas de **padrões cognitivos sistemáticos**, como o uso inconsistente de escalas, a tendência à severidade ou à leniência e a influência de experiências prévias com o estudante ou com o tipo de estação. Esses efeitos, conhecidos como *hawk-dove effect*, *frame-of-reference effect* e *contrast effect*, continuam presentes mesmo quando listas de verificação detalhadas são utilizadas, indicando que a padronização dos critérios não garante uniformidade na interpretação (10). Assim, a subjetividade não desaparece; apenas migra de um locus explícito para camadas menos visíveis do processo avaliativo.

A presença de **vieses identitários** representa outro campo de preocupação crescente. Pesquisas apontam que características como gênero, raça, sotaque, expressões culturais, modos de comunicação e até traços comportamentais percebidos como “profissionais” influenciam a forma como avaliadores interpretam o desempenho dos estudantes. Em

alguns casos, essas diferenças emergem por meio de micropreconceitos não intencionais, manifestados em avaliações mais rigorosas ou em expectativas distintas quanto ao comportamento considerado adequado (11). A padronização do OSCE parte da premissa de que o comportamento observado deve ser avaliado de forma neutra, mas ignora que tanto a observação quanto a interpretação são mediadas por matrizes culturais que produzem desigualdades. Assim, exames concebidos para promover equidade podem, inadvertidamente, reproduzir assimetrias históricas.

O papel dos **pacientes padronizados** também tem sido alvo de investigações críticas. Embora treinados para oferecer respostas consistentes, esses atores variam em expressividade, estilo comunicacional, afetividade e percepção subjetiva do estudante. Estudos mostram que diferenças sutis na postura corporal, no tom de voz ou na responsividade emocional podem alterar o nível de dificuldade de uma estação, criando heterogeneidade não reconhecida entre candidatos (10,11). Além disso, pacientes padronizados, como qualquer ator social, também carregam vieses implícitos que influenciam interações e avaliações, especialmente quando são convidados a atribuir escores de comunicação ou empatia.

A **ordem das estações** constitui outra fonte de viés estrutural. Efeitos como fadiga do avaliador, saturação cognitiva do estudante e impacto emocional decorrente de estações anteriores podem alterar significativamente o desempenho subsequente. Esse fenômeno produz uma cadeia de causalidade indireta: o resultado obtido em uma estação pode depender não apenas da competência do estudante, mas também de fatores situacionais acumulativos. Da mesma forma, diferenças entre turnos (manhã versus tarde) ou entre circuitos distintos de estações podem introduzir variações que afetam a equidade do exame, especialmente em contextos de alta consequência.

Outro ponto crítico refere-se à suposição de **neutralidade das tarefas**. A padronização parte da ideia de que todos os estudantes enfrentam o “mesmo” desafio, mas a literatura demonstra que certos tipos de estação favorecem estilos cognitivos específicos, repertórios comunicacionais particulares ou experiências clínicas prévias desigualmente distribuídas. Uma estação altamente estruturada pode beneficiar estudantes treinados para responder de forma objetiva e direta, enquanto prejudicar aqueles que adotam estratégias reflexivas mais lentas, ainda que competentes. Em situações inversas, estações abertas podem favorecer perfis comunicacionais mais expansivos, penalizando estudantes que apresentam timidez ou variações culturais de interação. Assim, a estrutura da tarefa atua como mediadora de desigualdades, e não como mecanismo neutro de avaliação.

No plano ético-político, essas evidências deslocam o debate da esfera técnica para o campo da **justiça avaliativa**, em que se reconhece que equidade não se alcança apenas com padronização, mas com reflexão crítica contínua sobre práticas e pressupostos. A literatura enfatiza que avaliações clínicas exigem formas robustas de governança interpretativa: monitoramento sistemático de padrões avaliativos, formação de avaliadores orientada para redução de vieses, triangulação de dados e disposição institucional para revisar profundamente as condições de produção do desempenho (12). Em vez de buscar a eliminação da subjetividade — objetivo impossível e epistemologicamente equivocado —, o foco desloca-se para o reconhecimento, a regulação e a negociação coletiva de julgamentos.

Essas discussões revelam que a promessa inicial de objetividade do OSCE precisa ser reinterpretada à luz da sua vulnerabilidade a condicionamentos estruturais, institucionais e identitários. Longe de representar um ambiente totalmente neutro, o exame constitui um espaço relacional em que julgamentos humanos são inevitáveis e, portanto, demandam abordagens avaliativas que combinem rigor técnico com sensibilidade ética e consciência crítica das desigualdades que marcam processos formativos.

5. Perspectivas emergentes e reposicionamento do OSCE em ecossistemas avaliativos complexos (texto corrido e aprofundado)

As discussões contemporâneas sobre o OSCE situam o exame em um momento de inflexão conceitual. Após décadas de hegemonia como principal método de avaliação da competência clínica, a literatura recente aponta que seu papel está deixando de ser o de instrumento central e totalizante para se tornar parte de **ecossistemas avaliativos complexos**, orientados por princípios de longitudinalidade, triangulação de evidências, foco formativo e integração de múltiplas perspectivas avaliativas (13–15). Esse reposicionamento não implica obsolescência do OSCE, mas sim uma reinterpretação de sua função pedagógica e regulatória.

Um dos movimentos mais significativos que influenciam essa transição é o avanço da **avaliação programática**, que propõe a utilização contínua e integrada de múltiplas fontes de informação sobre o desempenho, combinando observações diretas, narrativas avaliativas, feedback formativo e instrumentos de alta e baixa consequência. Nesse contexto, o OSCE deixa de carregar a responsabilidade de sintetizar, em um único momento, a complexidade da competência clínica. Em vez disso, ele passa a oferecer **evidências específicas** — principalmente relacionadas a comportamentos observáveis em cenários estruturados — que serão posteriormente trianguladas com outros dados provenientes de ambientes reais de prática, como avaliações no local de trabalho, portfólios e supervisões clínicas longitudinais.

Esse movimento desloca a ênfase da medição para a **interpretação**, reconhecendo que o julgamento profissional constitui elemento indispensável para compreender o desempenho clínico em sua totalidade. Em arquiteturas programáticas, o OSCE contribui para a análise de padrões e lacunas, mas sua leitura passa a depender de contextos mais amplos de aprendizagem, em que a progressão do estudante é avaliada ao longo do tempo, e não apenas com base em momentos isolados de alta consequência. Essa mudança reduz o peso injustamente atribuído ao OSCE como exame “decisório” e o reposiciona como ferramenta complementar dentro de sistemas avaliativos mais sensíveis à complexidade do desenvolvimento profissional.

Outro eixo emergente refere-se ao uso de **tecnologias digitais, análise de dados em larga escala e inteligência artificial** no aprimoramento dos processos avaliativos. Ferramentas computacionais têm sido exploradas para melhorar a consistência de julgamentos, identificar padrões de viés, apoiar avaliadores na tomada de decisão e oferecer feedbacks individualizados. No caso do OSCE, essas tecnologias possibilitam monitorar de forma sistemática variações entre avaliadores, discrepâncias entre turnos e diferenças no comportamento de pacientes padronizados. Além disso, sistemas de análise automatizada de linguagem, expressões e padrões comunicacionais começam a ser investigados como complementos às avaliações humanas, abrindo a possibilidade de

modelos híbridos que combinem objetividade instrumental com a interpretação contextualizada de avaliadores experientes.

A adoção dessas tecnologias, porém, não é isenta de desafios. Debates éticos emergem sobre o risco de **algoritmização de vieses existentes**, especialmente quando sistemas de IA são treinados com dados historicamente marcados por desigualdades. O problema deixa de ser apenas como aperfeiçoar o OSCE, mas como repensar o conceito de equidade em ecossistemas avaliativos que combinam agentes humanos e não humanos. Nesse cenário, a justiça avaliativa passa a exigir novas formas de governança, voltadas não apenas para a calibração de avaliadores, mas também para a supervisão, validação e auditabilidade de algoritmos.

Paralelamente, observa-se um ressurgimento do debate sobre **autenticidade e integração curricular**. À medida que currículos baseados em competências enfatizam situações de prática real e aprendizado centrado no paciente, cresce o interesse por formatos avaliativos capazes de captar o desempenho em condições menos artificiais do que aquelas tipicamente oferecidas pelo OSCE. Isso não reduz a utilidade do exame, mas implica repensar sua participação em um repertório avaliativo mais diversificado, que inclui observações diretas no local de prática (*workplace-based assessments*), discussões clínicas, mini-CEX, registros reflexivos e avaliações narrativas. O OSCE, nesse sentido, torna-se um instrumento especializado em avaliar **consistência de execuções técnicas, comunicação em cenários controlados e comportamentos padronizáveis**, enquanto outras práticas avaliativas se encarregam de dimensões de maior variabilidade contextual.

Finalmente, a literatura recente destaca uma tendência de deslocamento do discurso de “objetividade” para o discurso de **transparência e justiça interpretativa**. Em vez de apresentar o OSCE como solução para eliminar a subjetividade, novas análises argumentam que a subjetividade precisa ser reconhecida, monitorada e deliberadamente articulada ao processo de avaliação. A padronização, portanto, não desaparece, mas deixa de ser vista como garantia absoluta de equidade. O papel do OSCE torna-se contribuir com evidências claras e comparáveis, sem assumir que tais evidências representam, sozinhas, o constructo completo da competência clínica.

Nesse cenário de transformações, o OSCE permanece relevante, mas não mais como eixo central das decisões avaliativas. Seu valor deriva daquilo que **ele efetivamente pode oferecer** — estrutura, comparabilidade, amostragem eficiente de comportamentos observáveis — e de sua inserção estratégica em sistemas avaliativos que reconhecem a natureza complexa, situada e interpretativa do desempenho profissional. O futuro do OSCE, portanto, não está em reforçar sua soberania, mas em assumir seu papel como componente especializado dentro de **ecologias avaliativas amplas**, sensíveis à diversidade de métodos, à importância do feedback e à inevitável necessidade de julgamento profissional informado.

Conclusão

A análise crítica da trajetória do Objective Structured Clinical Examination (OSCE) evidencia que sua influência na educação médica decorre tanto de suas inovações técnicas quanto de sua capacidade de responder a exigências institucionais de padronização, transparência e responsabilização social. Desde sua formulação inicial, o

exame ofereceu uma alternativa mais robusta aos métodos clínicos tradicionais, contribuindo para a redução da variabilidade avaliativa e para a construção de modelos curriculares mais explícitos e estruturados. Entretanto, essa consolidação ocorreu ao custo de tensões persistentes, especialmente relacionadas à fragmentação do desempenho, à redução da autenticidade das tarefas e à simplificação do constructo de competência clínica.

A revisão mostra que a padronização, embora necessária para aumentar a consistência dos julgamentos, não elimina a subjetividade, mas redistribui seus efeitos para diferentes camadas do processo avaliativo. A presença de vieses identitários, cognitivos e organizacionais evidencia que a equidade não pode ser garantida apenas por meios técnicos e requer formas mais amplas de governança interpretativa. As críticas pedagógicas também revelam que a centralidade atribuída ao OSCE, em determinados contextos, tende a reorientar práticas de ensino para a “avaliabilidade”, em detrimento de abordagens que promovam pensamento clínico integrado, manejo da incerteza e sensibilidade contextual.

Diante das transformações contemporâneas nos currículos médicos e do avanço de arquiteturas avaliativas programáticas, o papel do OSCE passa a ser reinterpretado. Em vez de instrumento totalizante, o exame se fortalece como componente especializado dentro de ecossistemas que combinam múltiplas fontes de evidência e valorizam a longitudinalidade do desenvolvimento profissional. Tecnologias digitais e sistemas de análise avançada oferecem novas oportunidades para monitoramento de padrões, detecção de vieses e aprimoramento da qualidade das decisões avaliativas, embora introduzam desafios éticos que exigem supervisão institucional rigorosa.

Assim, o futuro do OSCE depende menos do refinamento isolado de suas propriedades psicométricas e mais de sua inserção coerente em arranjos avaliativos que reconheçam a complexidade e a natureza situada da competência clínica. Sua contribuição permanece relevante quando alinhada a práticas formativas que enfatizam julgamento profissional, feedback qualificado, triangulação de informações e compromisso explícito com a justiça avaliativa. O reposicionamento do OSCE como parte de um sistema integrado representa, portanto, não uma perda de centralidade, mas uma oportunidade para ampliar sua utilidade pedagógica e fortalecer seu papel na formação médica contemporânea.

Referências

1. Harden RM. Revisiting “Assessment of clinical competence using an objective structured clinical examination (OSCE)”. *Med Educ.* 2016;50(4):376–9.
2. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ.* 1979;13(1):41–54.
3. Patrício MF, Julião M, Fareleira F, Vaz Carneiro A. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? *Med Teach.* 2013;35(6):503–14.
4. Hodges B. OSCE! Variations on a theme by Harden. *Med Educ.* 2003;37(12):1134–40.

5. Troncon LEA. Clinical skills assessment: limitations to the introduction of an “OSCE” (Objective Structured Clinical Examination) in a traditional Brazilian medical school. *Sao Paulo Med J.* 2004;122(1):12–7.
6. Malau-Aduli BS, Jones K, Saad S, Richmond C. Has the OSCE met its final demise? Rebalancing clinical assessment approaches in the peri-pandemic world. *Front Med.* 2022;9:825502.
7. Chong L, Taylor S, Haywood M, Adelstein B-A, Shulruf B. The sights and insights of examiners in objective structured clinical examinations. *J Educ Eval Health Prof.* 2017;14:34.
8. Rodrigues MCC. Objective Structured Clinical Examination (OSCE): níveis de estresse, ansiedade e percepções de estudantes de medicina relacionados ao desempenho – um estudo de métodos mistos [tese]. Uberlândia: Universidade Federal de Uberlândia; 2021.
9. Khan KZ, Gaunt K, Ramachandran S, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: Organisation & Administration. *Med Teach.* 2013;35(9):e1447–63.
10. Boursicot K, Kemp S, Wilkinson T, Findyartini A, Canning C, Cilliers F, Fuller R. Performance assessment: Consensus statement and recommendations from the 2020 Ottawa Conference. *Med Teach.* 2020;42(12):1378–86.
11. Kane MT. Validating the interpretations and uses of test scores. *J Educ Meas.* 2013;50(1):1–73.
12. Barrows HS, Abrahamson S. The programmed patient: a technique for appraising student performance in clinical neurology. *J Med Educ.* 1964;39:802–5.
13. Goffman E. *The presentation of self in everyday life.* New York: Anchor Books; 1959.
14. van der Vleuten CPM, Schuwirth LWT, Driessen EW, Dijkstra J, Tigelaar D, Baartman LKJ, et al. A model for programmatic assessment fit for purpose. *Med Teach.* 2012;34(3):205–14.
15. Norcini JJ, Boursicot K, Roberts T, Jolly B, McKinley D. Workplace-based assessment: theory into practice. *Med Teach.* 2020;42(1):8–16.