

**UNIVERSIDADE PROFESSOR EDSON ANTÔNIO VELANO - UNIFENAS**  
**CLÁUDIA ÁVILA FERREIRA DRUMOND AMÉRICO**

**FATORES ASSOCIADOS A CONFIABILIDADE DE PROVA EM AMBIENTE  
SIMULADO: UM ESTUDO EXPERIMENTAL SOBRE O TREINAMENTO  
PRESENCIAL**

**Belo Horizonte**  
**2024**

**CLÁUDIA ÁVILA FERREIRA DRUMOND AMÉRICO**

**FATORES ASSOCIADOS A CONFIABILIDADE DE PROVA EM AMBIENTE  
SIMULADO: UM ESTUDO EXPERIMENTAL SOBRE O TREINAMENTO  
PRESENCIAL**

Dissertação apresentada à Universidade Prof. Edson Antônio Velano - UNIFENAS como parte das exigências do curso de Mestrado Profissional em Ensino em Saúde para obtenção do título de Mestre em Ensino em Saúde.

Orientador: Prof. Antonio Carlos de Castro Toledo Jr.

**Belo Horizonte**

**2024**

Dados Internacionais de Catalogação na Publicação (CIP)  
Biblioteca Unifenas BH Itapoã

Américo, Cláudia Ávila Ferreira Drumond.

Fatores associados a confiabilidade de prova em ambiente simulado: um estudo experimental sobre o treinamento presencial. [Manuscrito] / Cláudia Ávila Ferreira Drumond Américo. – Belo Horizonte, 2024. 57 f.

Orientador: Antônio Carlos de Castro Toledo Jr.

Dissertação (Mestrado) – Universidade Professor Edson Antônio Velano, Programa de Pós-Graduação Stricto Sensu em Ensino em Saúde, 2024.

1. Educação Médica. 2. Avaliação educacional. 3. Dispositivos de treinamento simulado. I. Américo, Cláudia Ávila Ferreira Drumond. II. Universidade Professor Edson Antônio Velano. III. Título.

CDU: 61:378

Bibliotecária responsável: Gisele da Silva Rodrigues CRB6 - 2404



Presidente da Fundação Mantenedora - FETA

Larissa Araújo Velano Dozza

Reitora

Maria do Rosário Velano

Vice-Reitora

Viviane Araújo Velano Cassis

Pró-Reitor Acadêmico Dannel Coelho

Pró-Reitora Administrativo-Financeira

Larissa Araújo Velano Dozza

Pró-Reitora de Planejamento e Desenvolvimento

Viviane Araújo Velano Cassis

Diretora de Pesquisa e Pós-graduação Bruno Cesar Correa Salles

Coordenador do Curso de Mestrado Profissional em Ensino em Saúde

Aloísio Cardoso Jr.

Coordenadora Adjunta do Curso de Mestrado Profissional em Ensino em Saúde

Maria Aparecida Turci

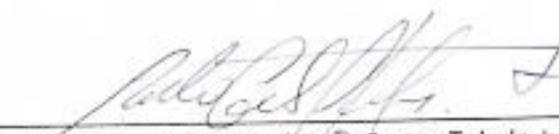
# Certificado de Aprovação

**FATORES ASSOCIADOS A CONFIABILIDADE DE PROVA EM AMBIENTE  
SIMULADO: UM ESTUDO EXPERIMENTAL SOBRE O TREINAMENTO  
PRESENCIAL**

**AUTOR:** Cláudia Ávila Ferreira Drumond Américo

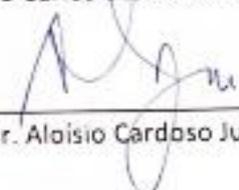
**ORIENTADOR:** Prof. Dr. Antônio Carlos de Castro Toledo Junior

Aprovada como parte das exigências para obtenção do Título de Mestre, no Programa de Pós-graduação Profissional de Mestrado em Ensino em Saúde pela Comissão Examinadora.



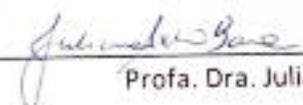
---

Prof. Dr. Antônio Carlos de Castro Toledo Junior



---

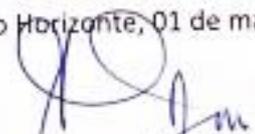
Prof. Dr. Aloisio Cardoso Junior



---

Profa. Dra. Juliana Silva Barra

Belo Horizonte, 01 de março de 2024.



**Prof. Dr. Aloisio Cardoso Junior**  
Coordenador do Mestrado Profissional  
Em Ensino em Saúde  
UNIFENAS

Dedico esta jornada de conquista e aprendizado a duas presenças que moldaram o meu caminho de maneiras únicas: aos meus pais, cujo amor, apoio e sabedoria guiaram meus passos, e a meu amado marido, cuja dedicação à medicina é uma inspiração constante.

A você, meu pai, embora não esteja mais fisicamente ao meu lado, sua influência perdura e seu espírito é uma força motriz. A cada passo deste percurso, senti sua presença, encorajando-me a persistir e a alcançar meus objetivos.

A minha mãe, por ser a base de carinho e amor que sempre nos sustentou. E por estar sempre na torcida, confiante no nosso sucesso.

Ao meu marido, Luiz Renato, cuja presença e apoio constantes são essenciais em minha vida.

A minha amada filha Luiza, que está seguindo meus passos na área médica, espero que esta jornada acadêmica inspire você a continuar a explorar e aprimorar o ensino em saúde.

Ao meu filho Felipe, mesmo que ainda jovem, você ilumina meus dias e me dá força para continuar. Saiba que estou fazendo isso por você e para um futuro melhor.

Este é um momento de celebração, amor e gratidão, e dedico esta conquista a todos vocês.

## **AGRADECIMENTOS**

Gostaria de expressar minha sincera gratidão a todos que tornaram esta jornada possível.

Primeiramente, meu mais profundo agradecimento ao meu orientador, o Prof. Antonio Toledo Jr. Sua orientação, apoio inabalável e profundo conhecimento nesta área foram essenciais para o sucesso desta pesquisa e para o meu percurso e crescimento acadêmico. Suas palavras de encorajamento e orientações sábias me impulsionaram a alcançar este marco.

Aos meus colegas e amigos que generosamente dedicaram seu tempo e conhecimento contribuindo para a minha pesquisa, não tenho palavras suficientes para expressar minha gratidão. Suas valiosas contribuições foram fundamentais para enriquecer este trabalho e torná-lo mais significativo.

Estendo meu agradecimento a todos que, de alguma forma, apoiaram, incentivaram e acreditaram em mim ao longo desta jornada. Minha família, em especial, merece uma menção à parte. Meus pais, cujo apoio incondicional moldou minha trajetória acadêmica, meu amado marido, cujo apoio incansável e compreensão tornaram tudo possível, e meus filhos, Luiza e Felipe, que me inspiram diariamente com seu amor e alegria.

*“O ato de ensinar é o ato de tocar a vida para sempre.”*

Albert Einstein

## RESUMO

No ensino médico ao longo dos séculos, observa-se a transformação da avaliação, inicialmente voltada para diferenciar competências individuais, evoluindo para uma abordagem mais ampla, buscando confiabilidade, validade e um papel mais integrado na aprendizagem. A ênfase se desloca da mera mensuração de conhecimento para a promoção de independência e pensamento crítico, com maior participação dos alunos no processo e um feedback mais significativo. O presente estudo busca confirmar a importância do treinamento dos avaliadores para minimizar variações e erros na avaliação em ambientes simulados. **Material e métodos:** experimento randomizado que avaliou como o treinamento de médicos pouco experientes afeta a confiabilidade da avaliação de habilidades clínicas em um laboratório de simulação. A população incluiu médicos voluntários sem experiência docente, que foram divididos em dois grupos. O treinamento foi realizado em um dos grupos (intervenção), e o outro permaneceu sem treinamento (controle). Esses médicos foram solicitados a avaliar alunos durante aferição da pressão arterial e exame torácico, utilizando um questionário online. A análise estatística incluiu medidas de confiabilidade entre os grupos, notas das avaliações e outros desfechos secundários. **Resultados:** não houve diferenças estatísticas entre os grupos de avaliadores, indicando eficácia na randomização. No desempenho dos alunos, não houve diferença significativa entre os grupos de avaliadores, mas o desempenho dos alunos foi abaixo do esperado em várias áreas. A análise multivariada mostrou que várias variáveis, como sexo do aluno, sexo e formação do avaliador, idade e tempo de formado, influenciaram as notas atribuídas. A confiabilidade da avaliação variou entre as áreas consideradas, sendo excelente em dados objetivos e menos consistente em aspectos subjetivos. Em relação à eficiência e eficácia, os resultados indicaram que as notas de eficiência foram estatisticamente maiores do que as de eficácia em diversas situações avaliadas. A inclusão de dados objetivos na avaliação, como a pressão arterial e achados da ausculta cardíaca, aumentou a confiabilidade, diferenciando claramente eficiência (fazer corretamente) de eficácia (obter o resultado correto). É importante ressaltar a constante necessidade de pesquisas e aprimoramento de métodos de avaliação que sejam mais eficientes, capazes de avaliar habilidades e competências de maneira confiável e uniforme.

**Palavras-chave:** educação médica; avaliação educacional; treinamento por simulação; confiabilidade do resultado; eficácia; eficiência.

## ABSTRACT

Throughout centuries of medical education, the evolution of assessment can be observed, initially aimed at differentiating individual competencies, evolving towards a broader approach seeking reliability, validity and a more integrated role in learning. The emphasis shifts from mere knowledge measurement to fostering independence and critical thinking, with increased student involvement in the process and more meaningful feedback. This study aims to confirm the importance of evaluator training to minimize variations and errors in assessment within simulated environments. **Materials and methods:** A randomized experiment assessed how training inexperienced physicians affects the reliability of clinical skills assessment in a simulation laboratory. The population included volunteer doctors without teaching experience, divided into two groups. Training was conducted for one group (intervention), while the other remained untrained (control). These doctors were tasked with evaluating students during blood pressure measurement and thoracic examination using an online questionnaire. Statistical analysis encompassed reliability measures between groups, assessment scores, and other secondary outcomes. **Results:** There were no statistical differences between the evaluator groups, indicating effectiveness in randomization. Regarding student performance, there was no significant difference between evaluator groups, but student performance was below expectations in several areas. Multivariate analysis showed that various variables, such as student sex, evaluator sex and background, age and years since graduation, influenced the assigned scores. Assessment reliability varied across evaluated areas, being excellent in objective data and less consistent in subjective aspects. Concerning efficiency and effectiveness, results indicated that efficiency scores were statistically higher than effectiveness in various evaluated situations. Including objective data in the assessment, such as blood pressure and cardiac auscultation findings, increased reliability, clearly differentiating efficiency (doing correctly) from effectiveness (achieving the correct result). It is important to emphasize the constant need for research and improvement in assessment methods that are more efficient, capable of reliably and uniformly evaluating skills and competencies.

**Keywords:** medical education; educational measurement; simulation training; reproducibility of results; efficacy; efficiency.

## LISTA DE ILUSTRAÇÕES

<b>Figura 1</b> -	Página de acesso às avaliações dos voluntários de 1 a 10 no Ambiente Virtual de Aprendizagem .....	27
<b>Figura 2</b> -	Escala de avaliação global .....	27
<b>Quadro 1</b> -	Parâmetros de interpretação do teste alfa de Cronbach .....	29

## LISTA DE TABELAS

<b>Tabela 1</b> - Análise comparativa dos dados pessoais e profissionais dos 12 voluntários de acordo com o grupo de alocação (Grupo 1 - controle e Grupo 2 - intervenção) .....	30
<b>Tabela 2</b> - Comparação entre as notas médias atribuídas a cada um dos itens da avaliação de acordo com o grupo de avaliadores (Grupo 1 - controle e Grupo 2 - intervenção) .....	31
<b>Tabela 3</b> - Comparação entre o percentual de acerto de cada um dos itens da avaliação de acordo com o grupo de avaliadores (Grupo 1 - controle e Grupo 2 - intervenção) .....	32
<b>Tabela 4</b> - Comparação das notas médias de acordo com variáveis de confusão .....	33
<b>Tabela 5</b> - Análise multivariada de variáveis de confusão selecionadas de acordo com o tipo de nota .....	33
<b>Tabela 6</b> - Análise consistência interna (alfa de Cronbach) de acordo com o grupo de avaliadores (Grupo 1 - controle e Grupo 2 - intervenção) .....	34
<b>Tabela 7</b> - Comparação entre a eficiência (fazer do modo correto) e a eficácia (obter o resultado correto) considerando a aferição da pressão arterial, a ausculta cardíaca e a nota final .....	35

## LISTA DE ABREVIATURAS, SÍMBOLOS E SIGLAS

a.C.	Antes de Cristo
ACV	Aparelho Cardiovascular
AVA	Ambiente virtual de aprendizado
CONEP	Conselho Nacional de Ética em Pesquisa
CRMMG	Conselho Regional de Medicina de Minas Gerais
d.C.	Depois de Cristo
d.p.	Desvio padrão
mmHg	Milímetros de mercúrio
MOODLE	<i>Modular Object-Oriented Dynamic Learning Environment</i>
OSCE	<i>Objective structured clinical examination</i> ou Exame clínico objetivo estruturado
PA	Pressão arterial
TCLE	Termo de Consentimento Livre e Esclarecido
UNIFENAS	Universidade Prof. Edson Antônio Velano
UVV	Universidade Vila Velha

## SUMÁRIO

1	INTRODUÇÃO.....	16
1.1	Aprendizagem baseada em simulação.....	18
1.2	Avaliação de habilidades e competências.....	19
2	JUSTIFICATIVA.....	23
3	OBJETIVOS.....	24
3.1	Objetivo geral.....	24
3.2	Objetivos específicos.....	24
4	MATERIAIS E MÉTODOS.....	25
4.1	Desenho do estudo.....	25
4.2	População, amostra e amostragem.....	25
4.3	Procedimentos.....	26
4.3.1	<i>Prova</i> .....	26
4.3.2	<i>Treinamento dos avaliadores</i> .....	27
4.3.3	<i>Avaliação da prova</i> .....	27
4.4	Análise estatística.....	29
4.5	Aspectos éticos.....	30
5	RESULTADOS.....	31
5.1	Descrição dos avaliadores.....	31
5.2	Desempenho dos alunos.....	32
5.3	Análise de confiabilidade.....	35
5.4	Eficiência <i>versus</i> Eficácia.....	36
6	DISCUSSÃO.....	37
6.1	Avaliadores.....	37
6.2	Treinamento dos avaliadores.....	37
6.3	Desempenho dos alunos.....	38
6.4	Eficiência <i>versus</i> eficácia.....	39
7	CONCLUSÕES.....	42
	REFERÊNCIAS.....	43
	APÊNDICES.....	46
	APÊNDICE A - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO.....	46

APÊNDICE B - FORMULÁRIO DE RECRUTAMENTO DE VOLUNTÁRIOS.....	50
APÊNDICE C - QUESTIONÁRIO DE AVALIAÇÃO DA PROVA .....	56
ANEXOS .....	58
ANEXO A - PARECER CONSUBSTANCIADO DO CEP SOBRE EMENTA DO ATUAL PROJETO .....	58

## 1 INTRODUÇÃO

A avaliação no processo ensino-aprendizado tem sido parte das sociedades por mais de 2000 anos (GIPPS, 1999). A mensuração do conhecimento e de habilidades tem perseverado durante o passar dos anos como forma de avaliação do saber. Os primeiros registros de avaliações datam da China, durante a Dinastia Han, onde candidatos eram submetidos a testes para servir ao governo (206 a.C. até 220 d.C.). A prática da medicina na era medieval requeria testes de competência e, no século XVII, há relatos de provas para adentrar as escolas jesuítas (NORCINI *et al.*, 2011).

Com relação às escolas médicas, as primeiras avaliações formais datam de 1788, quando, para ser selecionado para o internato nas escolas austríacas e francesas, o candidato deveria ser submetido a exames escritos e orais. Posteriormente, as avaliações entre os estudantes foram introduzidas nas escolas médicas britânicas em 1850, nas universidades de Oxford e de Cambridge. Em 1861, esse procedimento foi definido como rotina pelo Conselho Geral de Medicina, estabelecido na Inglaterra em 1858 — prática esta que foi introduzida nas demais escolas médicas europeias, no final do século XIX (NORCINI *et al.*, 2011).

Nos EUA, a situação foi diferente. Durante o século XVIII, houve grande proliferação de escolas médicas públicas e privadas, nas quais a forma de ensinar, treinar habilidades clínicas e cirúrgicas e avaliar todo esse processo era muito variável, como nos mostra o estudo de Abraham Flexner (1910), que revolucionou a educação médica nos Estados Unidos. A partir de 1930, as escolas médicas passaram a oferecer treinamento baseado em laboratório e atendimento hospitalar, com exames de final de curso para validar o aprendizado (STARR, 1982).

A maneira como a qualidade da avaliação tem sido percebida e assegurada mudou consideravelmente nas últimas cinco décadas. Originalmente, a avaliação era vista com o objetivo de diferenciar os indivíduos competentes dos não competentes. Com o tempo, maior valor foi dado à reprodutibilidade ou confiabilidade e validade do construto das avaliações, de modo que o papel do julgamento humano foi minimizado. Atualmente, procura-se tratar a avaliação de forma completa, na qual a competência é avaliada de uma perspectiva integrada e não reducionista. A pesquisa atual, portanto, se concentra em como apoiar e melhorar o julgamento humano, como triangular informações de avaliação de forma significativa e como

construir justiça, credibilidade e defensibilidade a partir de uma perspectiva de sistemas (SCHUWIRTH; VAN DER VLEUTEN, 2020).

A avaliação na educação médica na década de 1960 visava produzir uma análise padronizada e “objetiva”, devido à insatisfação com a prática predominante, muitas vezes vista como subjetiva, não confiável e tendenciosa (SCHUWIRTH; VAN DER VLEUTEN, 2020). Já na década de 1990, observou-se uma inovação no pensamento de como realizar uma avaliação, que também deveria promover independência e pensamento crítico, não se concentrando apenas na mensuração de conhecimento (BOUD, 2006). Os alunos deveriam ser incluídos como partes ativas e responsáveis por todo o processo, devendo receber feedback significativo (BOUD, 1995) — diferente da forma inicial, na qual a aprendizagem era avaliada por meio de mecanismos behavioristas, de reforço e punição. No entanto, a forma como a avaliação impulsiona a aprendizagem é mais complexa do que simplesmente punição e recompensa, sendo influenciada pela forma como os alunos constroem o significado da avaliação (CILLIERS *et al.*, 2012).

A avaliação pode ser definida como procedimento de obtenção de informações, visando à tomada de medidas para o aprimoramento do processo educacional e à interferência positiva no aprendizado do estudante. Deve incidir sobre aspectos relevantes e ser conduzida com métodos objetivos, em condições padronizadas, em circunstâncias viáveis e de aceitação indiscutível por todos os envolvidos na avaliação (TRONCON, 1996). Uma boa avaliação tem como objetivo melhorar o aprendizado dos estudantes, a qualidade dos professores e do ensino e, como produto final, aprimorar o cuidado com o paciente. Por isso, as instituições devem incentivar a pesquisa na qualidade das avaliações no sistema de educação médica (NORCINI *et al.*, 2011).

A avaliação, sendo um processo multifacetado, tem implicações para todos os envolvidos. A validade e a confiabilidade são essenciais para garantir a utilidade das avaliações. Existem limitações pragmáticas na escolha de métodos de avaliação, sendo alguns universalmente reconhecidos, como exames escritos, enquanto áreas emergentes, como simulação e avaliações de habilidades não técnicas, estão sendo ativamente pesquisadas (GOSMAN, 2016). É importante que a avaliação seja criteriosa e confiável. A escolha do critério mais adequado para uma boa avaliação vai depender de vários fatores, como o público que será avaliado, quem fará a avaliação e o objetivo final do processo de avaliação, entre outros (CHESSER, 2009).

Os critérios para uma boa avaliação podem ser aplicados para uma única avaliação, ou para uma sequência de avaliações com foco em um mesmo objetivo final (NORCINI *et al.*, 2011).

Particular importância deve ser dada à(ao):

- coerência - conjunto de evidências que levam o uso do resultado da avaliação para um propósito particular;
- reprodutibilidade, consistência ou confiabilidade - os resultados daquela avaliação serão sempre semelhantes se reproduzidos em circunstâncias similares;
- equivalência - os resultados são sempre equivalentes se administrados por diferentes instituições ou em diferentes ciclos de testes;
- viabilidade - a avaliação é prática e realista dentro das circunstâncias e do contexto em que é empregada;
- efeito educacional - a avaliação tem efeito motivador para aqueles que participam do processo e recebem o feedback do resultado;
- efeito catalítico - a avaliação, seus resultados e feedback criam chances para o desenvolvimento e aprimoramento de todo o processo educacional.
- aceitabilidade - os envolvidos (estudantes e professores) acreditam no processo avaliativo e têm confiança nos resultados.

Nos últimos 50 anos, muitas ferramentas foram desenvolvidas para avaliação dos graduandos de medicina. Ferramentas estas direcionadas para diferentes dimensões da competência médica e para novas técnicas de ensino e aprendizado. Observou-se o uso da tecnologia (computadores) como parte integrante dos processos de avaliação e ensino, além da possibilidade da aplicação de ferramentas individuais para avaliação de resultados (NORCINI *et al.*, 2011).

### **1.1 Aprendizagem baseada em simulação**

A Aprendizagem Baseada em Simulação (ABS) é uma estratégia pedagógica ativa amplamente empregada na educação médica contemporânea. Ela pode ser definida como uma técnica de ensino que se fundamenta nos princípios do ensino orientado para tarefas em um modelo artificial denominado simulador (PAZIN FILHO; SCARPELINI, 2007). O termo "simulador" refere-se a um objeto físico que representa, total ou parcialmente, uma tarefa a ser executada, enquanto "simulação" diz respeito à aplicação desse simulador para fins educacionais ou treinamento (COOPER; TAQUERI, 2004). Diversas modalidades de simulação são empregadas, incluindo simuladores de baixa tecnologia, pacientes-padrão, simuladores de

computador, simuladores de tarefas complexas e simuladores de pacientes (PAZIN FILHO; SCARPELINI, 2007).

A prática da simulação na área da saúde remonta ao século XVIII, quando Angélique Marguerite Le Boursier du Coudray, na França, confeccionou manequins de pano para o treinamento de parteiras e médicos, com o objetivo de reduzir a morbidade e mortalidade nos partos (MORAN, 2010). Na década de 1960, Asmund Laerdal, incentivado pelos Dr. Bjorn Lind e Dr. Peter Safar, desenvolveu o pioneiro manequim para reanimação cardiopulmonar, desempenhando um papel crucial no surgimento da simulação médica moderna (COOPER; TAQUETI, 2004). Desde então, a simulação ganhou destaque crescente no processo de ensino e aprendizagem.

A simulação pode ser empregada para desenvolver e substituir práticas vivenciais, mantendo aspectos substanciais do mundo real em um ambiente que pode ser controlado e, por isso, é mais seguro. As aplicações da simulação na área da saúde podem ser categorizadas em onze dimensões, abrangendo objetivos, unidade de participação, nível de experiência dos participantes, domínio da saúde, disciplina profissional, tipo de conhecimento, habilidades, atitudes ou comportamentos, idade simulada do paciente, tecnologia aplicável, local de simulação, extensão da participação direta e método de feedback (GABA, 2007).

Os principais propósitos da ABS incluem o desenvolvimento de habilidades e competências, a capacitação prévia do futuro médico antes de procedimentos em pacientes reais e a promoção da segurança tanto para o paciente quanto para o estudante. Além do ensino, a simulação possibilita a avaliação da competência adquirida pelo aprendiz (PAZIN FILHO; SCARPELINI, 2007).

## **1.2 Avaliação de habilidades e competências**

A avaliação de habilidades e competências, como a realização de anamnese e do exame físico ou a confirmação de um diagnóstico, habilidades fundamentais do médico, é um processo desafiador (NOVACK *et al*, 1993). Ao desenvolver simulações, técnica que se baseia em princípios fundamentados em tarefas realizadas em um modelo artificial (PAZIN FILHO; SCARPELINI, 2007), é crucial refletir sobre o propósito, seja no ensino ou na aprendizagem. No caso do ensino, a ênfase deve ser nos critérios de desempenho, avaliando a eficácia do

método. Se a simulação visa à aprendizagem de ordem superior, a avaliação deve medir várias habilidades integradas.

Diversas ferramentas de avaliação estão disponíveis para medir habilidades e competências, sendo crucial validá-las, uma vez que a confiabilidade dessas ferramentas são essenciais, garantindo que avaliem efetivamente o que se propõem a medir. Na prática, a prestação de cuidados de saúde real exige uma combinação de habilidades técnicas, analíticas e interpessoais. Portanto, as ferramentas devem ser abrangentes e capazes de avaliar de forma consistente e reproduzível. A validação dessas ferramentas na educação em simulação de saúde assegura que os alunos possam integrar conhecimentos e habilidades de maneira realista.

O exame clínico objetivo estruturado (OSCE), criado por Harden na década de 1970, foi uma resposta à essa necessidade. O OSCE é organizado em estações que simulam a realidade clínica e os estudantes são avaliados em tarefas específicas, como anamnese, exame físico, comunicação com paciente, procedimentos clínicos não invasivos, entre outras. Assim, é realizado o rodízio de alunos e um ou dois examinadores, que permanecem fixos nas estações, pontuam o desempenho dos alunos em cada tarefa, com critérios e tempo previamente determinados. O OSCE pode facilitar a execução de avaliação confiável de habilidades e competências clínicas (COMERT *et al*, 2016; NORMAN, 2002).

Há poucos estudos publicados sobre a confiabilidade de avaliações realizadas em ambientes simulados durante a graduação. Uma revisão sistemática realizada por Lee, Brain e Martin (2017) identificou três estudos. Em dois deles, os graduandos foram avaliados por avaliadores com habilidades básicas em aplicação de provas (GOVAERTS *et al.*, 2011, 2013) e o terceiro avaliou o resultado mais do que o processo cognitivo da avaliação em si. O julgamento do avaliador depende muito do contexto e da interação avaliador-estudante. “Traduzir” o processo de avaliação em escalas é uma das chaves para a correção desse viés na avaliação do estudante (GAUTIER, ST-ONGE, TAVARES, 2016). A redução da confiabilidade no avaliador pode ser explicada pela dificuldade de alinhamento dos objetivos entre os avaliadores, conceitos próprios e influências externas, que podem levar ao erro e à variação das notas obtidas (GINGERICH, REGEHR, EVA, 2011). Outros fatores podem influenciar na confiabilidade, como sexo do avaliador e do avaliado, a idade e o tempo de experiência clínica do avaliador e o fato de o avaliador conhecer ou não o aluno (CHESSER, 2009).

A variação na precisão dos julgamentos do examinador é uma fonte de erro de medição em testes baseados em desempenho, o que pode reduzir a confiabilidade da avaliação (VAN DER VLEUTEN, 1989). O treinamento dos avaliadores pode ser uma estratégia para minimizar esse problema: durante o treinamento, que pode se resumir a alguns encontros para estudo da situação (estação do OSCE) que servirá de base à simulação em que consiste a prova, devem ser definidos os escores que serão aplicados aos estudantes, com base nos itens verificados (CHESSER, 2009; DANIELS; PUGH, 2018). O avaliador pode receber orientações sobre o nível do aluno, o objetivo da prova e o grau de interação permitido entre examinador e aluno. Com isso, a construção de um modelo mental de desempenho desejado pode elevar à confiabilidade da pontuação (DANIELS; PUGH, 2018).

Um dos grandes obstáculos no campo do ensino em saúde reside na avaliação dos estudantes. Avaliar de maneira consistente aspectos desejados, como a habilidade do aluno na resolução de problemas, enfrenta desafios provenientes de fatores tanto intrínsecos quanto extrínsecos. A avaliação deve minimizar possíveis fontes de erro, assegurando uma pontuação que verdadeiramente reflita o desempenho do aluno. Avaliações de competências clínicas estão suscetíveis a erros em sua concepção, execução e mensuração, com o avaliador representando uma possível fonte de viés. Quanto mais complexas as habilidades avaliadas, maior o desafio na criação de uma avaliação robusta (OLIVEIRA, 2022).

Os estudos sobre o impacto do treinamento de avaliadores na educação médica têm resultados controversos. Alguns estudos sugerem que o treinamento produz melhora marginal ou nenhuma melhora na precisão dos julgamentos do examinador (VAN DER VLEUTEN, 1989) e outros indicam que um corpo docente treinado é mais rigoroso e tem menor variação em algumas das classificações (WEITZ, 2014). Segundo Newble, Hoare e Sheldrake (1980), o treinamento de avaliadores não reduz a variação do examinador na pontuação. Já Van der Vleuten *et al.* observaram que o treinamento melhora a precisão de avaliadores sem experiência em exames clínicos, sendo o efeito moderado em avaliadores experientes (VAN DER VLEUTEN *et al.*, 1989).

Em estudo prévio, realizado no Curso de Mestrado em Ensino em Saúde da UNIFENAS, observou-se que o treinamento on-line de avaliadores experientes não teve efeito na confiabilidade da avaliação em um ambiente simulado (OLIVEIRA, 2022). Assim sendo, o

objetivo deste estudo é avaliar se o treinamento presencial de avaliadores sem experiência com avaliação tem efeito na confiabilidade de provas em ambiente simulado.

## 2 JUSTIFICATIVA

Dado o complexo processo de formação dos estudantes, a avaliação desempenha papel crucial e deve ser de alta qualidade. A confiabilidade da avaliação está vinculada à precisão, consistência e reprodutibilidade, sendo influenciada por fatores como amostragem, padronização da avaliação, instrumentos de avaliação bem elaborados e critérios de correção previamente acordados.

Sabe-se que a avaliação é importante no processo de ensino e aprendizado. A avaliação de habilidades e competências é um desafio em especial, pois é mais dependente do avaliador e, por isso, possui menor confiabilidade do que as avaliações cognitivas.

A confiabilidade de avaliações em ambiente simulado ainda é um tema pouco estudado em educação médica, de modo que poucas pesquisas analisaram o impacto do treinamento dos avaliadores sobre a confiabilidade nessa situação. O maior conhecimento desse tema, feito por meio de maior número de estudos, permitirá o desenvolvimento de estratégias que poderão aprimorar o processo de avaliação.

### **3 OBJETIVOS**

#### **3.1 Objetivo geral**

Avaliar fatores que influenciam na confiabilidade de prova prática em ambiente simulado entre avaliadores pouco experientes.

#### **3.2 Objetivos específicos**

Os objetivos específicos deste estudo são:

- avaliar o efeito do treinamento na confiabilidade de uma prova em ambiente simulado;
- avaliar o efeito de outros fatores, como sexo, idade, tempo de formado, na confiabilidade de uma prova em ambiente simulado.

## **4 MATERIAIS E MÉTODOS**

### **4.1 Desenho do estudo**

Este é um estudo experimental e randomizado para avaliar o impacto do treinamento de avaliadores pouco experientes no que diz respeito à confiabilidade de avaliação de habilidades clínicas em laboratório de simulação. Ressalta-se que este trabalho é a continuação de pesquisa anterior, que avaliou o impacto do treinamento entre avaliadores experientes (OLIVEIRA, 2022).

### **4.2 População, amostra e amostragem**

Utilizou-se amostra estratificada dividida em dois grupos: alunos e avaliadores. Os alunos foram os mesmos 30 do estudo anterior (OLIVEIRA, 2022). Definiu-se como meta deste estudo a seleção de 20 avaliadores (10 em cada grupo). A unidade amostral final era a avaliação, que foi analisada de forma independente por cada avaliador. Dessa forma, a amostra final foi o produto do número de avaliadores e número de avaliações (alunos), que representa 600 avaliações.

Os avaliadores foram recrutados entre médicos de ambos os sexos, sem experiência como docentes de cursos de graduação e pós-graduação nos últimos 10 anos. Foram incluídos médicos que preencheram os seguintes critérios:

- com registro ativo no CRMMG;
- residentes ou atuantes em Belo Horizonte ou região metropolitana;
- sem experiência docente nos últimos 10 anos;
- que assinaram o Termo de Consentimento Livre e Esclarecido (TCLE - APÊNDICE A).

Critérios de exclusão de voluntários:

- não realizar todas as atividades programadas no estudo, incluindo o treinamento e avaliações;
- ser randomizado incorretamente;
- não assinar o TCLE.

Utilizou-se amostra de conveniência por meio da técnica *snowball*. Os potenciais voluntários foram contatados por e-mail e WhatsApp, entre grupos de trabalho dos pesquisadores, dando-se preferência para ortopedistas, oftalmologistas e otorrinolaringologistas. Os voluntários potenciais preencheram formulário de pré-recrutamento pelo Google Forms (APÊNDICE B). As pessoas que preenchiam os critérios de inclusão foram contatadas individualmente por um pesquisador, que explicou os procedimentos da pesquisa. Os que concordaram em participar foram randomizados em dois grupos através de uma sequência aleatória de número em blocos de cinco. O Grupo 1 (controle) não foi treinado para o uso do instrumento de avaliação e o Grupo 2 (intervenção) recebeu treinamento presencial, no qual foram contempladas instruções para a aferição correta da pressão arterial e exame torácico cardiológico. Os participantes do Grupo 1 receberam o TCLE por e-mail e só foram incluídos após a devolução do termo assinado. Os voluntários do Grupo 2 assinaram o TCLE no dia do treinamento.

### **4.3 Procedimentos**

#### **4.3.1 Prova**

Para avaliação, utilizou-se a mesma prova gravada do estudo anterior, realizada por 30 alunos voluntários do Curso de Medicina do 8º período da Universidade de Vila Velha, Espírito Santo. A prova foi gravada no Laboratório de Simulação, utilizando-se duas câmeras, uma frontal e uma do alto, sobre o manequim SimMan® Laerdal, que foi programado com pressão arterial de 130 x 80 mmHg e bloqueio atrioventricular de 2º grau com murmúrio sistólico.

Foi apresentado aos alunos o seguinte cenário: “Você é chamado para avaliar um paciente masculino, 60 anos, diabético e hipertenso, que procurou a unidade de saúde devido a mal-estar e desconforto precordial de início há 1 hora. O paciente se encontra deitado na maca da sala de emergência. Você já cumprimentou, se identificou, e explicou o procedimento de exame ao mesmo.

**Tarefa 1:** Faça a aferição da pressão arterial do paciente. Verbalize em voz alta os valores aferidos em mmHg.

**Tarefa 2:** Faça o exame físico do tórax (inspeção, palpação e ausculta cardíaca) do paciente. Verbalize os achados do seu exame físico.”

### ***4.3.2 Treinamento dos avaliadores***

Os médicos voluntários do Grupo 2 receberam o treinamento presencial em uma aula ministrada por médico clínico, na qual foram abordados os conceitos básicos sobre simulação e a técnica de aferição correta da pressão arterial (PA) e exame cardiológico do tórax. Após a exposição teórica, o professor fez a demonstração da aferição da PA em um voluntário e, em seguida, eles treinaram a técnica entre si. Na última fase do treinamento, os voluntários foram levados ao Laboratório de Simulação Realística da UNIFENAS, Campus de Belo Horizonte, onde tiveram a oportunidade de ter contato com o manequim SimMan®, regulado com os mesmos parâmetros do utilizado para a avaliação dos alunos. Foi demonstrado pelo professor como os alunos deveriam agir diante do cenário. A seguir discutiu-se os parâmetros de correção e cada voluntário realizou o teste, recebendo feedback dos colegas e do professor. Por fim, foram discutidas e esclarecidas as dúvidas técnicas dos voluntários.

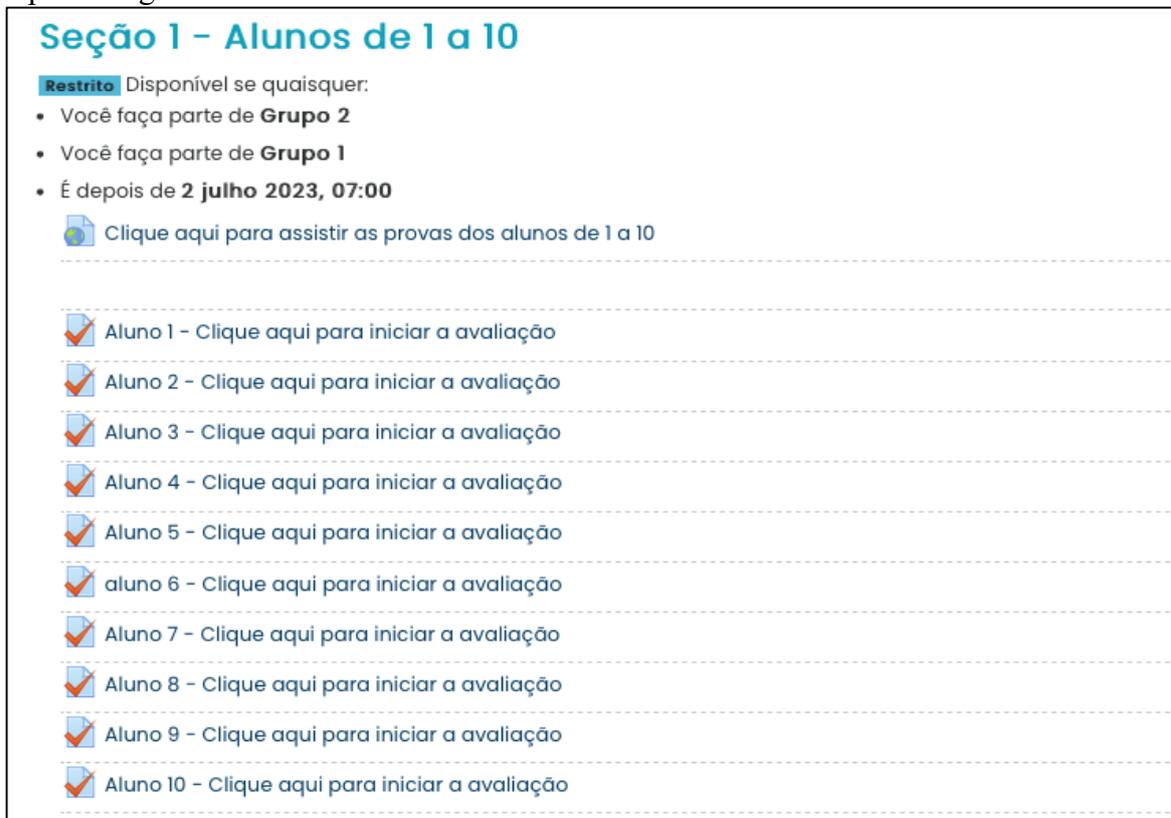
### ***4.3.3 Avaliação da prova***

A avaliação foi realizada on-line por todos os voluntários, através do ambiente virtual de aprendizado (AVA) da UNIFENAS, baseado no Moodle (FIG. 1). O acesso foi feito por senhas individuais. Os dois grupos tinham acesso a uma apresentação sobre aprendizagem baseada em simulação e as instruções para realização da avaliação. As 30 provas foram divididas em três blocos de 10 provas e os voluntários foram orientados a assistir a uma prova de cada vez e preencher a avaliação do aluno logo em seguida. Também foram orientados a não rever o vídeo, ou pará-lo no meio da prova. Elaborou-se um questionário para cada aluno, para evitar o preenchimento errado do número do aluno, uma vez que cada avaliador só podia responder a avaliação de cada aluno uma única vez.

Os voluntários preenchiam um questionário eletrônico no próprio AVA, com as seguintes perguntas:

1. Qual sua avaliação da técnica da medida da pressão arterial?
2. Qual o valor da pressão **SISTÓLICA** relatado pelo aluno (em mmHg)?
3. Qual o valor da pressão **DIASTÓLICA** relatado pelo aluno (em mmHg)?
4. Qual sua avaliação da técnica do exame cardiológico do tórax?
5. O aluno relatou alguma alteração do ritmo cardíaco?
6. O aluno relatou a presença de algum sopro?

**Figura 1** - Página de acesso às avaliações dos voluntários de 1 a 10 no Ambiente Virtual de Aprendizagem



Para as perguntas 1 e 4 utilizou-se escala de classificação tipo Likert (FIG. 2) e o avaliador indicava o valor correspondente a sua avaliação; nas perguntas 2 e 3, o avaliador inseria o valor relatado pelos alunos; e as perguntas 5 e 6 eram do tipo sim ou não, não sendo avaliado o tipo de alteração, apenas sua presença (APÊNDICE C).

**Figura 2** - Escala de classificação global

Não realizou	Realizou				
	Insatisfatório			Satisfatório	
	Péssimo	Ruim	Limítrofe	Bom	Excelente
Zero	1	2	3	4	5

Fonte: Elaborada pela autora.

Para as perguntas 1 e 4 utilizou-se o valor indicado pelo avaliador como valor da nota. Para a aferição da PA, considerou-se correta a pressão sistólica de  $130 \pm 5$  mmHg (2,5 pontos) e a pressão diastólica de  $80 \pm 5$  mmHg (2,5 pontos). A resposta correta para as perguntas 5 e 6 era

“sim” e cada acerto correspondeu a 2,5 pontos. A soma das questões 1 e 4 representavam a técnica de realização do exame (máximo de 10 pontos) e a demais questões representavam a eficácia do aluno, que também somavam 10 pontos.

#### 4.4 Análise estatística

O desfecho principal do estudo foi confiabilidade das notas, medida pelo alfa de Cronbach, entre os dois grupos. Os desfechos secundários foram a nota da técnica de realização do exame físico, nota da eficácia e nota total. O fator de exposição foi o treinamento. Ademais, foram consideradas as seguintes variáveis de confusão: especialidade do avaliador (clínica ou cirúrgica), sexo do avaliador e do aluno, idade e tempo de formado do avaliador. A averiguação foi realizada de forma cega em relação ao grupo de alocação dos avaliadores.

A análise descritiva incluiu as medidas de tendência central e dispersão para as variáveis contínuas, e a distribuição de frequência para as variáveis categóricas. A comparação entre os grupos foi realizada utilizando o teste t de *Student* para variáveis contínuas e o teste Qui-quadrado ( $\chi^2$ ) ou o teste de Fisher para variáveis categóricas. A confiabilidade dos resultados nos dois grupos foi avaliada com base no teste alfa de Cronbach.

A seleção do método estatístico para a avaliação da confiabilidade deve ser guiada pelo que se pretende avaliar, como destacado por Souza, Alexandre e Guirardello (2017) e Troncon (1996). O alfa de Cronbach foi desenvolvido com o propósito de oferecer uma medida de consistência interna de um teste ou escala, apresentando-se como um valor numérico entre 0 e 1 (TAVAKOL; DENNICK, 2011). Desde a década de 1950, o alfa de Cronbach tem sido amplamente utilizado como método de avaliação de confiabilidade, refletindo o grau de covariância entre os itens de uma escala de avaliação. Algumas pesquisas consideram valores acima de 0,7 como satisfatórios (SOUZA; ALEXANDRE; GUIRARDELLO, 2017; BRANNICK; EROL-KORKMAZ; PREWETT, 2011). Neste estudo, o alfa de Cronbach foi interpretado conforme o Quadro 1.

A análise multivariada foi realizada por meio de regressão linear para ajustar o valor de p. O nível de significância estatística adotado foi de  $p < 0,05$ . O banco de dados foi formatado no Microsoft® Excel® (Microsoft, USA) e a análise estatística foi feita com o software estatístico JASP 0.17.2.1 (*University of Amsterdam*, Holanda).

**Quadro 1** - Parâmetros de interpretação do teste alfa de Cronbach

Valor observado	Consistência interna
> 0,9	Excelente
> 0,8 e ≤ 0,9	Boa
> 0,7 e ≤ 0,8	Aceitável
> 0,6 e ≤ 0,7	Questionável
> 0,5 e ≤ 0,6	Fraca
< 0,5	Inaceitável

Fonte: SOUZA; ALEXANDRE; GUIRARDELLO, 2017.

#### 4.5 Aspectos éticos

Este estudo foi aprovado pelos Comitês de Ética em Pesquisa da UNIFENAS (Parecer 4.594.758 e 5.978.220 - ANEXO A) e da UVV (Parecer 4.788.957), e está em acordo com a Declaração de Helsinque e a Resolução 466/2012 do CONEP. Todos os voluntários tiveram a oportunidade de esclarecer suas dúvidas e assinaram o TCLE antes de sua inclusão no estudo.

Nenhum voluntário recebeu pagamento ou outro tipo de benefício pessoal para participar do estudo. Os voluntários receberam um lanche no dia da avaliação ou no dia do treinamento. Não houve risco físico para alunos e avaliadores, pois não foi realizado nenhum procedimento invasivo. No entanto, os alunos podem ter sentido algum grau de pressão psicológica devido ao fato de estarem sendo filmados e de serem avaliados, além de algum cansaço físico e/ou mental durante a realização do projeto, devido ao tempo de confinamento. Já os avaliadores podem ter sentido algum cansaço físico e/ou mental durante a realização do projeto, pelo tempo de avaliação da prova. É possível que tenha ocorrido algum constrangimento durante a realização da prova ou sua avaliação, apesar de não ter havido manifestação de nenhum dos voluntários em relação a isso. O risco de quebra de sigilo foi minimizado pelo arquivamento do TCLE em local diferente dos questionários e vídeos, e pela censura feita por meio de desfoque do rosto dos alunos nos vídeos. Além disso, todos os questionários foram identificados por números e somente os pesquisadores tiveram acesso às informações e ao banco de dados.

## 5 RESULTADOS

No período de 18 de maio a 11 de julho de 2023 foram recrutados 132 possíveis voluntários, sendo que 27 confirmaram a participação e foram randomizados após a assinatura do TCLE. Deste total, 12 desistiram após a randomização, 2 foram randomizados errados e 1 não completou a avaliação. A amostra final foi composta por 12 voluntários (360 avaliações), 7 alocados no Grupo 1 (210 avaliações) e 5 no Grupo 2 (150 avaliações). O treinamento do Grupo 2 ocorreu em 01/07/2023. O questionário para avaliação dos alunos ficou disponível para preenchimento pelos dois grupos de 05 de julho a 27 de agosto de 2023.

### 5.1 Descrição dos avaliadores

A Tabela 1 apresenta os dados pessoais e profissionais dos voluntários. A maior parte deles era do sexo feminino (66,7%), com formação em uma especialidade cirúrgica (75,0%), e não possuía mestrado ou doutorado (83,3%). As especialidades cirúrgicas envolvidas foram ginecologia e obstetrícia, oftalmologia, otorrinolaringologia e cirurgia geral. Metade dos voluntários afirmou realizar o exame físico do aparelho cardiovascular com frequência e 58,7% relataram que a habilidade para realização deste exame era ruim. A média de idade foi de 52,5 ( $\pm 9,6$ ) anos e o tempo de formado de 28,7 ( $\pm 8,7$ ) anos. Não houve diferença estatisticamente significativa entre os dois grupos, sugerindo que a randomização foi eficaz. É importante destacar que, no Grupo 1, o percentual maior de voluntários relatou realizar o exame do aparelho cardiovascular com frequência, apesar de não haver significância estatística.

**Tabela 1** - Análise comparativa dos dados pessoais e profissionais dos 12 voluntários de acordo com o grupo de alocação (Grupo 1 - controle e Grupo 2 - intervenção) [continua...]

Variável		Grupo 1		Grupo 2		Total		p*
		N	%	N	%	N	%	
Sexo	Feminino	6	75,0	2	25,0	8	66,7	0,151
	Masculino	1	25,0	3	75,3	4	33,3	
Especialidade	Ginecologia e obstetrícia	2	28,6	0	0,0	2	16,7	0,151
	Oftalmologia	0	0,0	2	40,0	2	16,7	
	Otorrinolaringologia	2	28,6	0	0,0	2	16,7	
	Cirurgião geral	0	0,0	1	20,0	1	8,3	
	Dermatologista	1	14,3	0	0,0	1	8,3	
	Endocrinologista	1	14,3	0	0,0	1	8,3	
	Geriatría	1	14,3	0	0,0	1	8,3	
	Medicina de Família	0	0,0	1	20,0	1	8,3	
	Ortopedia	0	0,0	1	20,0	1	8,3	

Tabela 1 - Análise comparativa dos dados pessoais e profissionais dos 12 voluntários de acordo com o grupo de alocação (Grupo 1 - controle e Grupo 2 - intervenção) [... conclusão]

Variável		Grupo 1		Grupo 2		Total		p*
		N	%	N	%	N	%	
Especialidade	Cirúrgica	5	55,6	4	44,4	9	75,0	0,636
	Clínica	2	66,7	1	33,3	3	25,0	
Mestrado ou	Não	5	50,0	5	50,0	10	83,3	0,318
Doutorado	Sim	2	100,0	0	0,0	2	16,7	
Realiza exame ACV	Raramente/às vezes	2	33,3	4	66,7	6	50,0	0,121
	Com frequência	5	83,3	1	16,7	6	50,0	
Habilidade de realizar o exame do ACV	Ruim	4	57,1	3	42,9	7	58,3	0,689
	Boa/muito boa	3	60,0	2	40,0	5	41,7	
Variável		Média	d.p.	Média	d.p.	Média	d.p.	p**
Idade		50,1	11,3	56,0	6,2	52,5	9,6	0,320
Tempo de formado		26,3	10,5	31,4	4,9	28,4	8,7	0,340

\* teste de Fischer unicaudal; \*\* teste T de *Student* (variáveis apresentaram distribuição normal); ACV - aparelho cardiovascular; d.p. – desvio padrão

Fonte: Elaborada pela autora.

## 5.2 Desempenho dos alunos

A Tabela 2 apresenta a média das notas atribuídas aos alunos para cada um dos itens da avaliação. Cada um dos primeiros quatro itens da tabela tinha valor máximo de 5,0 pontos. A nota da técnica correta corresponde à soma da técnica de aferição da PA e da técnica do exame cardíaco e a nota referente ao resultado correto é a soma das outras duas notas (acerto da medida da PA e da ausculta cardíaca alterada). Estas duas notas tinham valor máximo possível de 10,0 pontos cada. A nota total corresponde à soma das quatro primeiras notas. Não houve diferença entre o Grupo 1 (controle) e o Grupo 2 (treinamento).

Tabela 2 - Comparação entre as notas médias atribuídas a cada um dos itens da avaliação de acordo com o grupo de avaliadores (Grupo 1 - controle e Grupo 2 - intervenção)

Variável	Grupo 1 (n = 210)		Grupo 2 (n = 150)		Total (n = 360)		p*
	Média	d.p.	Média	d.p.	Média	d.p.	
1 Técnica de aferição da PA	3,71	0,88	3,71	1,13	3,71	0,99	0,440
2 Técnica de exame cardíaco	3,48	0,93	3,51	1,12	3,49	1,01	0,504
3 Nota acerto da medida da PA	1,91	1,38	1,90	1,38	1,90	1,38	0,980
4 Nota acerto de ausculta alterada	2,89	2,05	2,67	2,06	2,80	2,05	0,301
5 Nota da técnica correta (1 + 2)	7,19	1,66	7,21	2,02	7,20	1,82	0,499
6 Nota do resultado correto (3 + 4)	4,80	2,79	4,57	2,65	4,70	2,73	0,414
7 Nota total (5 + 6)	11,98	3,47	11,78	3,34	11,90	3,41	0,407

PA - pressão arterial. \* Teste de Mann-Whitney, uma vez que a distribuição não era normal (teste de Shapiro-Wilk < 0,05).

Fonte: Elaborada pela autora.

A Tabela 3 permite identificar que o maior percentual de erro ocorreu na medida da pressão arterial, principalmente da pressão sistólica (16,1% de acerto). Apenas 6,1% dos alunos acertaram a medida dos dois componentes da pressão arterial. O desempenho em relação à identificação da presença de arritmia cardíaca e de sopro foi melhor, 51,1% e 60,0% respectivamente, mas ainda abaixo do esperado para o grupo de alunos avaliados. Também não houve diferença entre os grupos de avaliadores.

**Tabela 3** - Comparação entre o percentual de acerto de cada um dos itens da avaliação de acordo com o grupo de avaliadores (Grupo 1- controle e Grupo 2 - intervenção)

Variável		Grupo 1 (n = 210)		Grupo 2 (n = 150)		Total (n = 360)		p*
		N	%	N	%	N	%	
Medida pressão sistólica	Incorreta	176	58,3	126	41,7	302	83,9	0,961
	Correta	34	58,6	24	41,4	58	16,1	
Medida pressão diastólica	Incorreta	84	58,3	60	41,7	144	40,0	1,000
	Correta	126	58,3	90	41,7	216	60,0	
Medida pressão arterial	Incorreta	197	58,3	141	41,7	338	93,9	0,941
	Correta	13	59,1	9	40,9	22	6,1	
Identificou arritmia na ausculta	Não	92	52,3	84	47,7	176	48,9	<b>0,023</b>
	Sim	118	64,1	66	35,9	184	51,1	
Identificou sopro na ausculta	Não	85	60,3	56	39,7	141	39,1	0,547
	Sim	125	57,1	94	42,9	219	60,8	

Fonte: Elaborada pela autora.

A Tabela 4 compara as notas médias de acordo com diferentes variáveis de confusão. Observa-se que há diferenças significativas em várias notas médias com base nessas variáveis. Estas diferenças sugerem que o sexo do discente, o sexo do avaliador e o tipo de formação do avaliador podem influenciar as notas atribuídas no ambiente simulado, sendo as notas médias das alunas significativamente mais altas do que as dos alunos. Os avaliadores do sexo masculino deram notas mais altas do que as avaliadoras do sexo feminino e os avaliadores com formação cirúrgica atribuíram notas mais altas aos alunos. Quanto mais velho o avaliador, menor a nota. Quanto maior o tempo de formado, maior a nota conferida. A análise utilizou o Teste de Mann-Whitney devido à não normalidade na distribuição das variáveis (teste de Shapiro-Wilk < 0,05).

**Tabela 4** - Comparação das notas médias de acordo com variáveis de confusão

Sexo discente	Feminino		Masculino		Total		p*
	Média	d.p.	Média	d.p.	Média	d.p.	
Nota da técnica correta	7,06	1,75	7,29	1,86	7,20	1,82	0,143
Nota do resultado correto	5,73	2,74	4,02	2,50	4,70	2,73	< 0,001
Nota total	12,79	3,18	11,30	3,44	11,90	3,41	< 0,001
Sexo avaliador	Média	d.p.	Média	d.p.	Média	d.p.	p*
Nota da técnica correta	6,80	1,80	8,00	1,57	7,20	1,82	< 0,001
Nota do resultado correto	4,76	2,79	4,58	2,62	4,70	2,73	0,480
Nota total	11,56	3,58	12,58	2,96	11,90	3,41	0,026
Tipo de formação avaliador	Cirúrgica		Clínica		Total		p*
	Média	d.p.	Média	d.p.	Média	d.p.	
Nota da técnica correta	7,44	1,64	6,46	2,11	7,20	1,82	< 0,001
Nota do resultado correto	4,64	2,68	4,89	2,88	4,70	2,73	0,444
Nota total	12,08	3,28	11,34	3,76	11,90	3,41	0,126

\* Teste de Mann-Whitney, uma vez que a distribuição não era normal (teste de Shapiro-Wilk < 0,05)

Fonte: Elaborada pela autora.

A Tabela 5 exhibe os resultados da análise multivariada (regressão linear) que explora as relações entre variáveis independentes, como o sexo do aluno, o sexo do avaliador, a idade do avaliador, o tipo de formação do avaliador e o tempo de formado, e variáveis dependentes, que representam notas em vários aspectos da avaliação.

Considerando-se a nota da técnica semiológica correta, que é a mais subjetiva, os avaliadores do sexo masculino e com mais tempo de formado atribuíram notas mais elevadas, enquanto os avaliadores de formação clínica e os mais velhos deram notas mais baixas. Em relação à nota do resultado correto (objetiva), apenas o sexo do aluno influenciou substancialmente a nota, com valores significativamente mais baixos. A nota final (soma das duas anteriores) foi influenciada pelo sexo do aluno (os alunos do sexo masculino receberam nota mais baixa) e pelo sexo do avaliador (os avaliadores do sexo masculino deram notas mais altas).

**Tabela 5** - Análise multivariada de variáveis de confusão selecionadas de acordo com o tipo de nota [continua...]

Nota da técnica correta	Beta não padronizado	p ajustado
Aluno do sexo masculino	0,225	0,222
Avaliador do sexo masculino	1,148	< 0,001
Idade do avaliador	-0,180	0,017
Avaliador com formação clínica	-0,575	0,013
Tempo formado	0,204	0,015

**Tabela 5** - Análise multivariada de variáveis de confusão selecionadas de acordo com o tipo de nota [... conclusão]

<b>Nota do resultado correto</b>	<b>Beta não padronizado</b>	<b>p ajustado</b>
Aluno do sexo masculino	-1,713	< <b>0,001</b>
Avaliador do sexo masculino	-0,111	0,741
Idade do avaliador	2,247×10 <sup>-4</sup>	0,998
Avaliador com formação clínica	0,198	0,575
Tempo formado	-0,003	0,981
<b>Nota final</b>	<b>Beta não padronizado</b>	<b>p ajustado</b>
Aluno do sexo masculino	-1,488	< <b>0,001</b>
Avaliador do sexo masculino	1,037	<b>0,015</b>
Idade do avaliador	-0,180	0,221
Avaliador com formação clínica	-0,377	0,400
Tempo formado	0,201	0,216

Fonte: Elaborada pela autora.

### 5.3 Análise de confiabilidade

A Tabela 6 exhibe os resultados da análise de confiabilidade de acordo com o alfa de Cronbach.

**Tabela 6** - Análise consistência interna (alfa de Cronbach) de acordo com o grupo de avaliadores (Grupo 1 - controle e Grupo 2 - intervenção)

<b>Variável</b>	<b>Grupo 1</b>	<b>Grupo 2</b>
1 Técnica de aferição da pressão arterial	0,742	0,405
2 Técnica de exame cardíaco	0,746	0,637
3 Nota acerto da medida da PA	0,993	0,966
4 Nota acerto de ausculta alterada	0,952	0,969
5 Nota da técnica correta (1 + 2)	0,789	0,599
6 Nota do resultado correto (3 + 4)	0,975	0,967
7 Nota total (5 + 6)	0,963	0,907

Fonte: Elaborada pela autora.

Nas avaliações da técnica semiológica, a confiabilidade foi aceitável no Grupo 1 (controle) e questionável ou fraca no Grupo 2 (treinamento), sugerindo que a intervenção pode ter piorado a confiabilidade. Já nos dados objetivos, a confiabilidade foi excelente nos dois grupos. A nota

final apresentou confiabilidade elevada nos dois grupos, independentemente da baixa confiabilidade nas subjetivas (técnica semiológica).

#### 5.4 Eficiência *versus* Eficácia

Pode-se definir que as notas das técnicas semiológicas representam a eficiência (fazer do modo correto), enquanto as notas da aferição da PA e da identificação das alterações na ausculta representam a eficácia (obter o resultado correto). A Tabela 7 compara a eficácia e a eficiência em relação à aferição da PA (máximo de 5 pontos), ao exame cardíaco (máximo de 5 pontos) e à nota final (máximo de 10 pontos). De modo geral, a nota da eficiência foi estatisticamente maior que a da eficácia nas três situações. A maior diferença foi observada na PA e a menor diferença na ausculta cardíaca.

**Tabela 7** - Comparação entre a eficiência (fazer do modo correto) e a eficácia (obter o resultado correto) considerando a aferição da pressão arterial, a ausculta cardíaca e a nota final

Variável	Média	d.p.	p*
Técnica de aferição da pressão arterial (eficiência)	3,71	0,99	< 0,001
Nota acerto da medida da PA (eficácia)	1,90	1,38	
Técnica de exame cardíaco (eficiência)	3,49	1,01	< 0,001
Nota acerto de ausculta alterada (eficácia)	2,80	2,05	
Nota da técnica correta (eficiência)	7,20	1,82	< 0,001
Nota do resultado correto (eficácia)	4,70	2,73	

d.p. – desvio padrão. \* Teste de Mann-Whitney  
Fonte: Elaborada pela autora.

## **6 DISCUSSÃO**

O objetivo geral deste estudo era avaliar fatores que influenciam na nota de prova prática em ambiente simulado. O objetivo principal era avaliar a influência do treinamento de avaliadores pouco experientes sobre a confiabilidade da nota de prova prática em ambiente simulado. Desse modo, trata-se de estudo experimental randomizado.

### **6.1 Avaliadores**

O estudo foi iniciado com um promissor grupo de 132 possíveis voluntários dispostos a contribuir com a pesquisa. No entanto, ao longo do processo, apareceram desafios significativos. Dois erros de randomização que poderiam comprometer a integridade dos dados foram excluídos. Além disso, ocorreu uma quantidade considerável de desistências por parte dos participantes, um aspecto que deixou clara a complexidade de realizar pesquisas que dependem exclusivamente da participação voluntária. Esses obstáculos reforçaram a noção de que conduzir estudos dessa natureza não é tarefa simples, destacando a importância de abordar com cautela as nuances da pesquisa com participantes voluntários.

### **6.2 Treinamento dos avaliadores**

No presente estudo não houve diferença estatisticamente significativa com relação ao resultado final (nota), entre os Grupos 1 e 2. A confiabilidade foi pior no Grupo 2, com relação à técnica de realização do exame (dado subjetivo). Isso também foi observado por Oliveira (2022), que analisou o efeito do treinamento em avaliadores experientes. De acordo com a literatura, o treinamento do avaliador não tem influência mensurável na precisão. Porém, avaliadores treinados tendem a ser mais rigorosos do que os avaliadores não treinados, o que não foi confirmado no estudo atual (WEITZ, 2014).

Dois possíveis explicações estão relacionadas ao senso comum de que a tarefa seria de execução e avaliação pouco complexas, e o treinamento pode ter confundido os avaliadores em relação ao seu conhecimento prévio das tarefas avaliadas. Outra possibilidade é que o treinamento não seja necessário para a avaliação deste tipo de tarefa. A maioria dos voluntários do Grupo 1 relatou realizar o exame do aparelho cardiovascular com frequência em sua prática diária, além de possuir uma boa habilidade em realizar este exame. Apesar de a diferença não

ser estatisticamente significativa, devido ao pequeno número de avaliadores, ela pode explicar a maior confiabilidade nas notas deste Grupo.

Segundo a literatura, o esforço de implementar formação de avaliadores para melhorar a imparcialidade dos exames pode não ser eficaz (WEITZ, 2014). O treinamento varia em eficácia em função da experiência médica, além de alguns trabalhos mostrarem que leigos treinados podem ser utilizados como examinadores. Para docentes e estudantes de medicina, o treinamento também influencia a natureza dos erros cometidos, reduzindo o número de erros de comissão em testes baseados em desempenho (VAN DER VLEUTEN, 1989; CHESSER, 2009). De acordo com Schleicher (2017), a inconsistência frequentemente observada nas notas de provas pode ser minimizada através do treinamento prévio dos avaliadores, mas nunca será completamente eliminada.

Deve-se destacar que a utilização de dados objetivos (acerto da medida da pressão arterial e alterações na ausculta cardíaca) aumentou significativamente a confiabilidade entre os dois grupos, eliminando totalmente a diferença observada com os dados subjetivos (avaliação da técnica do exame), como observado por Oliveira (2022) e por Garcia (2019).

### **6.3 Desempenho dos alunos**

Diferente do observado por Holmboe (2004), que em seus estudos não verificou qualquer influência no sexo do aluno no resultado final das avaliações, observou-se que o sexo do discente, o sexo do avaliador e o tipo de formação do avaliador podem influenciar as notas atribuídas no ambiente simulado, de modo que as notas médias das discentes (sexo feminino) foram significativamente mais altas do que as dos discentes (sexo masculino) ( $p < 0,001$ ). Os avaliadores do sexo masculino dão notas mais altas e as avaliadoras do sexo feminino são mais exigentes. Os avaliadores com formação cirúrgica conferem notas mais altas aos alunos ( $p < 0,001$ ), o que pode significar que são mais diretos em suas avaliações e menos detalhistas no “como fazer”. Quanto mais velho o avaliador, menor a nota. Quanto maior o tempo de formado, maior a nota conferida. Como a prova não valida nota, pode ser que os alunos tenham sido mais displicentes que as alunas, que normalmente são mais compromissadas.

A literatura mostra que o tempo de formado e, conseqüentemente, a experiência clínica não implicam necessariamente uma avaliação mais confiável ou justa por parte do examinador.

(SCHLEICHER, 2017). Considerando que a confiabilidade de uma prova é a reprodutibilidade do seu resultado por avaliadores diferentes, observa-se que avaliadores diferentes deram diferentes resultados, o que reduz a confiabilidade da prova.

Conforme evidenciado na literatura, reafirma-se a tendência de avaliadores do sexo masculino atribuírem notas mais elevadas, especialmente a estudantes do sexo feminino, enquanto avaliadoras do sexo feminino tendem a ser geralmente mais exigentes (SCHLEICHER, 2017). Os fenômenos conhecidos como efeito "pomba" (leniência) e efeito "gavião" (rigor), junto com um viés de gênero, contribuem para que professores do sexo masculino deem avaliações mais favoráveis a alunas do sexo feminino, ao passo que avaliadoras do sexo feminino avaliam de maneira mais rigorosa todos os alunos. Dada a valorização crescente da mulher como profissional na sociedade contemporânea, é plausível justificar as notas mais elevadas concedidas a estudantes do sexo feminino. A atual responsabilidade associada ao papel da mulher na sociedade também pode explicar a postura mais rigorosa adotada por avaliadoras femininas. Nota-se, ainda, uma transformação nos perfis das atuais escolas médicas, refletindo um aumento no número de alunas. Este cenário reflete a crescente valorização e reconhecimento da importância da mulher como profissional na sociedade atual.

Já Oliveira (2022) observou efeito inverso. As avaliadoras concederam pontuações mais elevadas em comparação aos seus colegas do sexo masculino. Além disso, constatou-se que avaliadores com maior tempo de docência e idade tenderam a atribuir pontuações mais baixas, enquanto aqueles com menos tempo de formação pontuam mais.

Os resultados apresentam dados potencialmente conflitantes, como avaliadores mais velhos atribuírem notas mais baixas e avaliadores graduados há mais tempo atribuírem notas mais altas. Essas são duas variáveis interrelacionadas, uma vez que os avaliadores mais velhos, em sua maioria, possuem mais tempo de graduação. Assim, é possível que essa aparente discrepância esteja relacionada a uma associação estatística espúria.

#### **6.4 Eficiência *versus* eficácia**

A inclusão de dados objetivos na avaliação, como o valor da pressão arterial e a presença de sopro ou arritmia cardíaca na ausculta, teve impacto positivo na confiabilidade dos resultados, reduzindo o viés de confusão que poderia surgir da interpretação do “como fazer” (eficiência).

Chama a atenção a diferença estatisticamente significativa entre as notas das técnicas semiológicas (eficiência — fazer do modo correto), estas bem superiores, e as notas do acerto da alteração (eficácia — obter o resultado correto). Por não se tratar de uma avaliação real, pode ter ocorrido certa displicência dos alunos em relação aos valores da PA e às alterações da ausculta.

No entanto, outros fatores devem ser considerados. Na aprendizagem em simulação, pode ocorrer o condicionamento dos alunos a repetirem mecanicamente os comportamentos e manobras consideradas “corretas”, de modo a garantir a nota na avaliação final. Este condicionamento reduz a interpretação do aluno em relação dos resultados, pois seu foco está apenas em fazer corretamente (CHESSER, 2009).

### **6.5 Limitações e vantagens**

Uma limitação deste estudo foi, sem dúvida, a dificuldade de angariar voluntários que pudessem participar efetivamente do treinamento e se comprometerem a dedicar seu tempo para a avaliação das provas gravadas. Muitos candidatos desistiram quando souberam que poderia ser necessário fazer o treinamento presencial em um sábado pela manhã, de acordo com a randomização. Outros, que concordaram em participar e foram randomizados, não compareceram ao treinamento. Além disso, um voluntário que participou do treinamento desistiu após o início das avaliações. Isso pode estar relacionado à falta de tempo generalizada de todas as pessoas no mundo atual, no qual o tempo para outras atividades que não estejam relacionadas a trabalho remunerado é escasso. A desistência de voluntários em pesquisa científica é um fenômeno comum e pode ser resultado de vários fatores, incluindo falta de interesse, mudanças nas circunstâncias pessoais dos participantes, desconforto com os procedimentos da pesquisa, entre outros motivos. Apesar disso, a amostra final não pode ser considerada pequena. Mesmo assim, pode ser que a amostra final não tenha sido suficiente para identificar o efeito do treinamento.

A incorporação de vídeos se revelou uma ferramenta valiosa na avaliação de testes, possibilitando a ampliação de dados estatísticos ao envolver um número maior de avaliadores para um mesmo contingente de examinandos e cenários. No entanto, não é possível mensurar

o nível de engajamento dos avaliadores durante a execução da atividade, nem acompanhar o cumprimento das regras, constituindo-se uma restrição potencial do estudo.

Os resultados corroboraram os achados do estudo anterior (OLIVEIRA, 2022). A falta de efeito do treinamento pode estar relacionada à prova em si, que pode ser considerada relativamente simples, sugerindo que, nessas situações, o treinamento não seria indicado. Outro ponto a se destacar é o aumento de confiabilidade ao se adicionar dados objetivos a tarefas, como observado por Oliveira (2022) e Garcia (2019).

## 7 CONCLUSÕES

O treinamento de avaliadores pouco experientes não interferiu na nota e na confiabilidade de avaliação em ambiente simulado. Os avaliadores do sexo masculino atribuíram notas mais altas e os alunos do sexo masculino receberam notas mais baixas. O acréscimo de dados objetivos parece aumentar a confiabilidade da avaliação. Os resultados sugerem que o treinamento não é necessário no caso da avaliação de tarefas mais simples (habilidades), como aferição da pressão arterial, mas não permite concluir sobre o efeito do treinamento de avaliadores em relação a tarefas mais complexas (competências).

## REFERÊNCIAS

- BOUD, D. Assessment and the promotion of academic values. **Studies in higher education**, [S.l.], v. 15, n. 1, pp. 101-111, 2006.
- BOUD, D. Assessment and learning: contradictory or complementary? In KNIGHT, P. **Assessment for Learning in Higher Education**. Londres: Routledge, 1995, pp. 35-48.
- CHESSER, A. *et al.* Sources of variation in performance on a shared OSCE station across four UK medical schools. **Medical Education**, v. 43, pp. 526-532, 2009.
- CILLIERS, J. F. *et al.* A model of the pre-assessment learning effects of summative assessment in medical education. **Advances in Health Sciences Education**, [S.l.], v. 17, pp. 39-53, 2012.
- COMERT, M. *et al.* Assessing communication skills of medical students in objective structured clinical examinations (OSCE) - A systematic review of rating scales. **PLOS ONE**, [S.l.], v. 11, n. 3, p. e0152717, 2016.
- COOPER, J. B.; TAQUERI, V. R. A brief history of the development of mannequin simulators for clinical education and training. *Quality and Safety in Health care*, [S. l.], v. 13, n. 1, pp. 11-18, 2004.
- DANIELS, V. J.; PUGH, D. Twelve tips for developing an OSCE that measures what you want. **Medical Teacher**, [S.l.], v. 40, pp. 1208-1213, 2018.
- FLEXNER, A. **Medical education in the United States and Canada: a report to the Carnegie Foundation for the Advancement of Teaching**. New York, 1910.
- GABA, D. M. The future vision of simulation in healthcare. **Simulation in Healthcare**, [S. l.], v. 2, n. 2, pp. 126-135, 2007.
- GARCIA, L. M. **Comparação da confiabilidade entre três escalas para avaliação de habilidades clínicas em Laboratório de Simulação e três tipos diferentes de avaliadores**. 2019, 54f. Dissertação Mestrado – Universidade José do Rosário Velano, Belo Horizonte, MG, 2019.
- GAUTIER, G.; ST-ONGE, C.; TAVARES, W. Rater cognition: Review and integration of research findings. **Medical Education**, [S.l.], v. 50, pp. 511-522, 2016.
- GINGERICH, A.; REGEHR, G., EVA, K. Rater-based assessments as social judgments: Rethinking the etiology of rater errors. **Academic Medicine**, [S.l.], v. 86, p. S1-7, 2011.
- GIPPS, C. Sociocultural aspects of assessment. **Review of Educational Research**, [S.l.], v. 24, pp. 355-392, 1999.
- GOSMAN, A. *et al.* Implementing Assessment Methods in Plastic Surgery. **Plastic and Reconstructive Surgery**, [S.l.], v. 137, n. 3, pp. 617e-623e, 2016.

GOVAERTS, M. *et al.* Workplace-based assessment: raters' performance theories and constructs. **Advances in Health Sciences Education**, [S.l.], v. 18, pp. 375-396, 2013.

GOVAERTS, M. *et al.* Workplace-based assessment: Effects of rater expertise. **Advances in Health Sciences Education**, [S.l.], v.16, pp. 151-165, 2011.

HOLMBOE, E. S. Faculty and the observation of trainees' clinical skills: problems and opportunities. **Academic Medicine**, [S.l.], v. 79, n. 5, pp. 16-22, 2004.

LEE, V.; BRAIN, K.; MARTIN, J. Factors influencing mini-CEX rater judgments and their practical implications. **Academic Medicine**, [S.l.], v. 92, n. 6, pp. 880-887, 2017.

MORAN, M. D. Enlightenment via Simulation: "Crone-ology's" First Woman. **Journal of Endourology**, [S. l.], v. 24, n. 1, pp. 5-8, jan. 2010.

NEWBLE, D. I.; HOARE, J.; SHELDRAKE, P. F. The selection and training of examiners for clinical examinations. **Medical Education**, [S.l.], v. 14, pp. 345-349, 1980.

NORCINI, J. *et al.* Criteria of good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. **Medical Teacher**, [S.l.], v. 33, pp. 206-214, 2011.

NORMAN, G. Research in medical education: Three decades of progress. **BMJ**, [S.l.], v. 324, pp. 1560-1562, 2002.

NOVACK, D. *et al.* Medical interviewing and interpersonal skills teaching in US Medical schools. **JAMA**, [S.l.], v. 269, pp. 2101-5, 1993.

OLIVEIRA, R. **Impacto do treinamento de avaliadores sobre a confiabilidade de prova em ambiente simulado**. 2022, 34f. Dissertação Mestrado – Universidade José do Rosário Velano, Belo Horizonte, MG, 2022.

PAZIN FILHO, A.; SCARPELINI, S. Simulação: definição. **Medicina**, Ribeirão Preto, v. 40, n. 2, pp. 162-166, abr/jun. 2007.

SCHLEICHER, I. *et al.* Examiner effect on the objective structured clinical exam-a study at five medical schools. **BMC Medical Education**, [S.l.], n. 17, v. 71, 2017.

SCHUWIRTH, L. W. T.; VAN DER VLEUTEN, C. P. M. A history of assessment in medical education. **Advances in Health Sciences Education**, [S.l.], v. 25, n. 5, pp. 1045-1056, 2020.

SOUZA, A. C; ALEXANDRE, N. M. C; GUIRARDELLO, E. D. Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. **Epidemiologia e Serviços de Saúde**, Brasília, v. 26, n. 3, pp. 649-659, 2017.

STARR, P. **The social transformation of American medicine**. Basic Books, New York, 1982.

TAVAKOL, M. DENNICK, R. Making Sense of Cronbach's Alpha. **International Journal of Medical Education**, [S.l.], v. 2, pp. 53-55, 2011.

TRONCON, L. E. A. Avaliação do Estudante de Medicina. **Medicina**, Ribeirão Preto, v. 29, pp. 429-439, out./dez. 1996.

VAN DER VLEUTEN, C. P M. *et al.* Training and experience of examiners. **Medical Education**, [S.l.], v. 23, pp. 290-296, 1989.

WEITZ, G. *et al.* Effects of a rater training on rating accuracy in a physical examination skills assessment. **GMS Zeitschrift für Medizinische Ausbildung**, [S.l.], v. 31, n. 4, pp. 1-17, 2014.

## APÊNDICES

### APÊNDICE A - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

#### 1 DADOS DA PESQUISA

**TÍTULO DA PESQUISA: IMPACTO DO TREINAMENTO RESENCIAL DE AVALIADORES POUCO EXPERIENTES SOBRE A CONFIABILIDADE DE PROVA EM AMBIENTE SIMULADO**

**PESQUISADORA:** Dra. Cláudia Ávila Ferreira Drumond Américo

**PESQUISADORES PARTICIPANTES:** Prof. Antonio Carlos C. Toledo Jr.

**ENDEREÇO:** Avenida do Contorno, 4747/403. Serra. Belo Horizonte – MG. CEP: 30110-921

**TELEFONE DE CONTATO:** (31) 999507453/ (31) 32644867

**E-MAIL:** clauamerico@yahoo.com.br ou draclaudiaavila.otorrino@gmail.com ou

claudia.americo@aluno.unifenas.br

**PATROCINADORES:** projeto próprio

Você está sendo convidado para participar, como voluntário, de uma pesquisa científica. Pesquisa é um conjunto de procedimentos que procura criar ou aumentar o conhecimento sobre um assunto. Estas descobertas embora frequentemente não tragam benefícios diretos ao participante da pesquisa, podem no futuro ser úteis para muitas pessoas.

Para decidir se aceita ou não participar desta pesquisa, o(a) senhor(a) precisa entender o suficiente sobre os riscos e benefícios, para que possa fazer um julgamento consciente. Sua participação não é obrigatória, e, a qualquer momento, você poderá desistir de participar e retirar seu consentimento. Sua recusa não trará nenhum prejuízo em sua relação com o pesquisador(a) ou com a instituição.

Explicaremos as razões da pesquisa. A seguir, forneceremos um termo de consentimento livre e esclarecido (TCLE), documento que contém informações sobre a pesquisa, para que leia e discuta com familiares ou outras pessoas de sua confiança. Caso seja necessário, alguém lerá e gravará a leitura para o(a) senhor(a). Uma vez compreendido o objetivo da pesquisa e havendo seu interesse em participar, será solicitada a sua rubrica em todas as páginas do TCLE e sua assinatura na última página. Uma via assinada deste termo deverá ser retida pelo senhor(a) ou por seu representante legal e uma cópia será arquivada pelo pesquisador responsável.

#### 2 INFORMAÇÕES DA PESQUISA

**2.1 Justificativa:** a simulação ou treinamento de habilidades é recurso importante na educação médica. Ela permite que o aluno aprenda e desenvolva habilidades e competências antes do contato com o paciente real, o que garante segurança para as duas partes. Como em várias áreas da educação, a avaliação é um dos grandes desafios do ensino em simulação. O exame clínico objetivo estruturado (OSCE) foi criado na década de 1970 e hoje é considerado o método de referência para avaliação de habilidades e competências em ambiente simulado.

No entanto, alguns fatores podem fazer com que a avaliação seja imprecisa, ou não seja confiável. A confiabilidade de uma avaliação se refere à sua capacidade de reproduzir os resultados, ou seja, alunos com desempenho semelhante têm notas semelhantes quando avaliados pelo mesmo avaliador ou determinado aluno tem notas semelhantes quando avaliado por avaliadores diferentes. A diferença entre avaliadores pode reduzir a confiabilidade de provas práticas. A literatura médica tem sugerido que o treinamento dos avaliadores pode ser uma estratégia para reduzir essas variações e aumentar a confiabilidade da prova, mas os resultados dos estudos ainda são conflitantes.

**2.2 Objetivo:** o objetivo geral é avaliar os fatores associados aos examinadores que podem interferir na confiabilidade do OSCE, entre eles o treinamento dos avaliadores.

**2.3 Metodologia:** caso aceite participar dessa pesquisa como avaliador, você se somará a outros 25 médicos que participarão como avaliadores independentes. O grupo será randomizado e metade do grupo receberá um treinamento presencial sobre a avaliação em ambiente simulado e a outra metade não será treinada. Você deverá avaliar 30 alunos que realizarão uma prova tipo OSCE. A prova será gravada e os vídeos disponibilizados para você, juntamente com as instruções para utilizar o instrumento de avaliação. Você terá uma semana para realizar a avaliação.

Você também receberá um questionário para coleta de informações sobre sexo, idade, tempo de formado, se possui experiência docente ou com aplicação de OSCE, e a especialidade médica. Os formulários serão identificados por números e todo esforço será feito para manter a sua identidade em sigilo.

**2.4 Riscos e Desconfortos:** caso aceite participar do trabalho, informamos que não haverá risco físico, pois não será realizado nenhum procedimento invasivo. No entanto, podem experimentar algum cansaço físico e/ou mental durante a realização do projeto devido, e tempo de avaliação da prova. Eventualmente pode ocorrer algum constrangimento, caso identifique algum aluno que você conheça. Todos os esforços serão feitos para que o cansaço seja minimizado e para se evitar eventuais constrangimentos, como tempo adequado para a realização do treinamento e da avaliação. Para prevenir a quebra de sigilo, o TCLE e o banco de dados serão arquivados em locais separados, seu nome não será utilizado no banco de dados e apenas os pesquisadores responsáveis terão acesso ao banco de dados. Todos os questionários serão identificados apenas por números.

**2.5 Benefícios:** não há nenhum benefício direto para você, mas sua participação ajudará a identificar fatores que influenciam a confiabilidade de provas práticas em ambiente simulado, o que será uma importante contribuição na literatura visto que no momento há um debate acadêmico sem conclusão sobre o tema.

**2.6 Forma de acompanhamento:** não haverá necessidade de acompanhamento a longo prazo.

**2.7 Alternativas de tratamento:** não se trata de pesquisa com tratamento.

**2.8 Privacidade e Confidencialidade:** os seus dados serão analisados em conjunto com outros participantes, não sendo divulgada a identificação de nenhum participante sob qualquer circunstância. Solicitamos sua autorização para que os dados obtidos nesta pesquisa sejam utilizados em uma publicação científica, meio pelos quais os resultados de uma pesquisa são divulgados e compartilhados com a comunidade científica. Todos os dados da pesquisa serão armazenados em local seguro por 5 anos. Todos os esforços serão realizados para manter a confidencialidade das informações.

**2.9 Acesso aos resultados:** você tem direito de acesso atualizado aos resultados da pesquisa, ainda que os mesmos possam afetar sua vontade em continuar participando da mesma.

### **3 Liberdade de recusar-se e retirar-se do estudo**

A escolha de entrar ou não nesse estudo é inteiramente sua. Caso o(a) senhor(a) se recuse a participar deste estudo, o(a) senhor(a) receberá o tratamento habitual, sem qualquer tipo de prejuízo ou represália. O(A) senhor(a) também tem o direito de retirar-se deste estudo a qualquer momento e, se isso acontecer, não haverá qualquer prejuízo ao tratamento ou represália.

### **4 Garantia de Ressarcimento**

O(A) senhor(a) não poderá ter compensações financeiras para participar da pesquisa, exceto como forma de ressarcimento de custos. Tampouco, o(a) senhor(a) não terá qualquer custo, pois o custo desta pesquisa será de responsabilidade do orçamento da pesquisa. O(A) senhor(a) tem direito a ressarcimento em caso de despesas decorrentes da sua participação na pesquisa.

### **5 Garantia de indenização**

Se ocorrer qualquer problema ou dano pessoal durante ou após os procedimentos aos quais o Sr.(a.) será submetido(a), lhe será garantido o direito a tratamento imediato e gratuito na Instituição, não excluindo a possibilidade de indenização determinada por lei, se o dano for decorrente da pesquisa.

### **6 Acesso ao pesquisador:**

Você tem garantido o acesso, em qualquer etapa da pesquisa, aos profissionais responsáveis pela mesma, para esclarecimento de eventuais dúvidas acerca de procedimentos, riscos, benefícios etc., através dos contatos abaixo:

**Pesquisador:** Dra. Cláudia Ávila Ferreira Drumond Américo

Telefone: (31) 999507453 \ (31) 32644867

Endereço: Avenida do Contorno 4747 \ 403. Serra, Belo Horizonte – MG. CEP: 30110-921

E-mail: clauamerico@yahoo.com.br ou draclaudiaavila.otorrno@gmail.com

**7 Acesso a instituição:**

Você tem garantido o acesso, em qualquer etapa da pesquisa, à instituição responsável pela mesma, para esclarecimento de eventuais dúvidas acerca dos procedimentos éticos, através do contato abaixo:

**Comitê de Ética - UNIFENAS:**

Rodovia MG 179, Km 0, Alfenas – MG

Telefone: (35) 3299-3137

E-mail: comitedeetica@unifenas.br

Segunda à sexta-feira das 14:00h às 16:00h

**8 Consentimento do participante**

Eu, abaixo assinado, declaro que concordo em participar desse estudo como voluntário(a) de pesquisa. Ficaram claros para mim quais são os objetivos do estudo, os procedimentos a serem realizados, os desconfortos e riscos, as garantias de confidencialidade e de esclarecimentos permanentes. Ficou claro também que a minha participação é isenta de despesas e que tenho garantia do acesso aos pesquisadores e à instituição de ensino. Foi-me garantido que eu posso me recusar a participar e retirar meu consentimento a qualquer momento, sem que isto me cause qualquer prejuízo, penalidade ou responsabilidade. A minha assinatura neste Termo de Consentimento Livre e Esclarecido dará autorização aos pesquisadores, ao patrocinador do estudo e ao Comitê de Ética em Pesquisa da Universidade José do Rosário Vellano, de utilizarem os dados obtidos quando se fizer necessário, incluindo a divulgação dos mesmos, sempre preservando minha identidade.

Assino o presente documento em duas vias de igual teor e forma, ficando uma em minha posse.

NOME: \_\_\_\_\_  
 RG: \_\_\_\_\_ SEXO: ( ) Masculino ( ) Feminino ( ) Prefiro não informar  
 DATA DE NASCIMENTO: \_\_/\_\_/\_\_\_\_\_  
 ENDEREÇO: \_\_\_\_\_  
 BAIRRO: \_\_\_\_\_  
 CIDADE: \_\_\_\_\_ ESTADO: \_\_\_\_\_ CEP: \_\_\_\_\_  
 TELEFONE: \_\_\_\_\_ - E-MAIL: \_\_\_\_\_

**9 Declaração do pesquisador**

Declaro que obtive de forma apropriada e voluntária, o Consentimentos Livre e Esclarecido deste participante (ou representante legal) para a participação neste estudo. Declaro ainda que me comprometo a cumprir todos os termos aqui descritos.

Belo Horizonte, \_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_

\_\_\_\_\_  
 Voluntário

\_\_\_\_\_  
 Pesquisador Responsável

\_\_\_\_\_  
 Voluntário

TESTEMUNHA

NOME: \_\_\_\_\_

ASSINATURA: \_\_\_\_\_

RG: \_\_\_\_\_



## APÊNDICE B - FORMULÁRIO DE RECRUTAMENTO DE VOLUNTÁRIOS

### Recrutamento de Voluntários

Cara Colega e Caro Colega,

Venho convidá-lo a participar do meu projeto de Mestrado intitulado "**O impacto de treinamento de avaliadores na confiabilidade de prova em ambiente simulado**". Esse questionário tem por objetivo avaliar se você preenche os critérios de inclusão e exclusão do estudo. **Seu preenchimento não implica em concordância em participar dele.** Eu, Dra. Cláudia Américo, farei contato com você para detalhar a sua possível participação e encaminhar o Termo de Consentimento Livre e Esclarecido. De forma similar a um ensaio clínico, os voluntários serão divididos aleatoriamente em dois grupos: controle e intervenção.

A simulação é um recurso utilizado cada vez mais na educação médica. Ela permite que o aluno treine, em um ambiente seguro, as técnicas de entrevista e exame físico e a realização de procedimentos, como entubação, punções, etc. A avaliação das habilidades e competências nesse ambiente é um grande desafio, pois pode haver muita variação entre os avaliadores. A literatura em educação médica não é clara sobre o impacto do treinamento sobre a diminuição dessa variabilidade (ou aumento da confiabilidade), que é o objetivo principal do estudo.

Mais uma vez, agradeço seu interesse e disponibilidade. O preenchimento deste questionário dura cerca de 5 minutos.

Atenciosamente,

Dra. Cláudia Américo  
Mestranda em Ensino em Saúde - UNIFENAS-BH

Prof. Antonio Toledo  
Orientador

*\* Indica uma pergunta obrigatória*

1. E-mail \*

\_\_\_\_\_

2. Você é formado em Medicina? \*

*Marcar apenas uma oval.*

Sim

Não *Pular para a seção 8 (Não preenche critérios)*

### Registro ativo CRM

3. Você possui registro ativo no CRM? \*

*Marcar apenas uma oval.*

Sim

Não *Pular para a seção 8 (Não preenche critérios)*

### Local de residência e trabalho

4. Você mora ou trabalha na região metropolitana de Belo Horizonte? \*

*Marcar apenas uma oval.*

Sim

Não *Pular para a seção 8 (Não preenche critérios)*

### Experiência docente

5. Você atua ou atuou como professor para graduação em medicina ou outro curso superior nos últimos 10 anos?

*Marcar apenas uma oval.*

Sim *Pular para a seção 8 (Não preenche critérios)*

Não

### Experiência supervisão

6. Você atua ou atuou como preceptor de residência médica e/ou internato médico, ou como professor de pós-graduação *latu senso* ou *strictu senso* nos últimos 10 anos?

*Marcar apenas uma oval.*

- SIM *Pular para a seção 8 (Não preenche critérios)*  
 NÃO

### **Identificação**

7. Qual seu nome completo? \*

\_\_\_\_\_

8. Por favor, indique a forma mais fácil entrar em contato com você (e-mail e/ou número de telefone)?

\_\_\_\_\_

9. Qual seu sexo? \*

*Marcar apenas uma oval.*

- Feminino  
 Masculino

10. Qual sua idade? (em anos completos) \*

\_\_\_\_\_

11. Em que ano você formou em Medicina? (use 4 dígitos) \*

\_\_\_\_\_

12. Você fez residência médica ou especialização? \*

*Marcar apenas uma oval.*

Sim

Não

13. Qual a sua especialidade ou área de atuação? (se tiver mais de uma, indique a que pratica com maior frequência)

\_\_\_\_\_

14. Possui mestrado ou doutorado já concluído? \*

*Marcar apenas uma oval.*

Sim

Não

15. Normalmente você realiza o exame do aparelho respiratório na sua prática clínica? \*

*Marcar apenas uma oval.*

Raramente

Algumas vezes

Com frequência

16. Como você avalia sua habilidade de realizar o exame do aparelho respiratório? (**essa pergunta tem 5 opções de resposta, se não estiver enxergando todas, coloque o seu telefone deitado**)

*Marcar apenas uma oval.*

1 2 3 4 5

\_\_\_\_\_  Ruim      Muito boa \_\_\_\_\_

17. Normalmente você realiza o exame do aparelho cardiovascular na sua prática clínica? \*

*Marcar apenas uma oval.*

- Raramente  
 Algumas vezes  
 Com frequência

18. Como você avalia sua habilidade de realizar o exame do aparelho cardiovascular? (**essa pergunta tem 5 opções de resposta, se não estiver enxergando todas, coloque o seu telefone deitado**)

*Marcar apenas uma oval.*

- 1 2 3 4 5  
Ruin      Muito boa

19. Normalmente você realiza o exame do aparelho digestório na sua prática clínica? \*

*Marcar apenas uma oval.*

- Raramente  
 Algumas vezes  
 Com frequência

20. Como você avalia sua habilidade de realizar o exame do aparelho digestório? (**essa pergunta tem 5 opções de resposta, se não estiver enxergando todas, coloque o seu telefone deitado**)

*Marcar apenas uma oval.*

- 1 2 3 4 5  
Ruin      Muito boa

**Preenche critério**

Muito obrigado por seu tempo. Em breve farei contato.

Dra. Cláudia Américo  
Mestranda em Ensino em Saúde - UNIFENAS-BH

**Não preenche critérios**

Infelizmente, você não preenche os critérios de inclusão do estudo nesse momento. Agradeço seu interesse em participar do estudo.

Dra. Cláudia Américo  
Mestranda em Ensino em Saúde - UNIFENAS-BH

---

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários

## APÊNDICE C - QUESTIONÁRIO DE AVALIAÇÃO DA PROVA

### Questão 1

Ainda não respondida

Não avaliada

Qual sua avaliação da técnica da medida da pressão arterial?  
(use o quadro a seguir com o referência para a nota)

ESCALA DE CLASSIFICAÇÃO GLOBAL

Não realizou	Realizou				
	Insatisfatório		Satisfatório		
	Péssimo	Ruim	Limítrofe	Bom	Excelente
Zero	1	2	3	4	5

Escolha uma opção:

- 0 - Não realizou
- 1- Insatisfatório - péssimo
- 2 - Insatisfatório - limítrofe
- 3 - Satisfatório - limítrofe
- 4 - Satisfatório - bom
- 5 - Satisfatório - excelente

### Questão 2

Ainda não respondida

Não avaliada

Qual o valor da pressão **SISTÓLICA** relatado pelo aluno (em mm Hg)

Resposta:

### Questão 3

Ainda não respondida

Não avaliada

Qual o valor da pressão **DIASTÓLICA** relatado pelo aluno (em mm Hg)

Resposta:

### Questão 4

Ainda não respondida

Não avaliada

Qual sua avaliação da técnica do exame cardiológico do tórax?

(use o quadro a seguir com o referência para a nota)

ESCALA DE CLASSIFICAÇÃO GLOBAL

Não realizou	Realizou				
	Insatisfatório		Satisfatório		
	Péssimo	Ruim	Limítrofe	Bom	Excelente
Zero	1	2	3	4	5

Escolha uma opção:

- 0 - Não realizou
- 1- Insatisfatório - péssimo

- 2 - Insatisfatório - limitado
- 3 - Satisfatório - limítrofe
- 4 - Satisfatório - bom
- 5 - Satisfatório - excelente

**Questão 5**

Ainda não  
respondida  
Não avaliada

O aluno relatou alguma alteração do ritmo cardíaco?

(não é necessário citar ou descrever alteração, apenas citar a alteração, com o arritmia, ritmo de galope, presença de B3, ritmo em 3 tempos, etc. Ritmo em dois tempos ou bulhas normorrítmicas são as referências da normalidade)

Escolha uma opção:

- Não
- Sim

**Questão 6**

Ainda não  
respondida  
Não avaliada

O aluno relatou a presença de algum sopro?

(não é necessário citar ou descrever o sopro, apenas se identificou a existência de sopro)

Escolha uma opção:

- Não
- Sim

**Manter contato**

! [ned@unifenas.br](mailto:ned@unifenas.br)

[Resumo de retenção de dados](#)

" Obter o aplicativo para dispositivos móveis

## ANEXOS

## ANEXO A - PARECER CONSUBSTANCIADO DO CEP SOBRE EMENTA DO ATUAL PROJETO



Continuação do Parecer: 5.978.220

Somente os pesquisadores terão acesso às informações e ao banco de dados. Não haverá prejuízo de desempenho curricular dos alunos, pois a avaliação não fará parte da grade curricular avaliativa.

**Benefícios:**

Não há nenhum benefício direto para você, mas sua participação ajudará a identificar fatores que influenciam a confiabilidade de provas práticas em ambiente simulado, o que será uma importante contribuição na literatura visto que no momento há um debate acadêmico sem conclusão sobre o tema.

**Comentários e Considerações sobre a Pesquisa:**

Não foi encontrado nenhum óbice ético.

**Considerações sobre os Termos de apresentação obrigatória:**

Não foi encontrado nenhum óbice ético.

**Recomendações:**

Recomendações:

Introdução é um apanhado do assunto a ser desenvolvido no trabalho. Deve conter apenas o essencial para introduzir o assunto. Não se trata de uma revisão de literatura. a introdução deve conter no máximo duas páginas e poucas referências.

**Conclusões ou Pendências e Lista de Inadequações:**

Não foi encontrado nenhum óbice ético.

**Considerações Finais a critério do CEP:**

**Este parecer foi elaborado baseado nos documentos abaixo relacionados:**

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_2052659_E1.pdf	04/03/2023 08:36:32		Aceito
Projeto Detalhado / Brochura Investigador	brochura_pesquisador_treinamento_v3.pdf	04/03/2023 08:33:43	Antonio Toledo Jr	Aceito
Folha de Rosto	folhaDeRosto.pdf	04/03/2023 08:32:52	Antonio Toledo Jr	Aceito
Outros	justificativa_ementa_modificacoes.	04/03/2023	Antonio Toledo Jr	Aceito

**Endereço:** Rodovia MG 179 km 0, BLOCO VI SALA 602 1º ANDAR  
**Bairro:** Campus Universitário **CEP:** 37.130-000  
**UF:** MG **Município:** ALFENAS  
**Telefone:** (35)3299-3137 **Fax:** (35)3299-3137 **E-mail:** comitedeetica@unifenas.br



UNIVERSIDADE JOSÉ  
ROSÁRIO VELLANO -



Continuação do Parecer: 5.978.220

Outros	pdf	08:28:54	Antonio Toledo Jr	Aceito
Parecer Anterior	PB_PARECER_CONSUBSTANCIADO_CEP_4594758.pdf	04/03/2023 08:07:44	Antonio Toledo Jr	Aceito
Outros	questionario_do_avaliador.pdf	04/03/2023 08:02:42	Antonio Toledo Jr	Aceito
Outros	prova_simulacao_treinamento.pdf	04/03/2023 08:01:47	Antonio Toledo Jr	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	tcle_aluno_treinamento_medico_v2.pdf	04/03/2023 07:59:57	Antonio Toledo Jr	Aceito
Outros	conhecimento_UVV.pdf	30/10/2020 10:27:38	Antonio Toledo Jr	Aceito
Declaração de concordância	Declaracao_Conhecimento_treinamento_docente_Unifenas.pdf	23/10/2020 11:29:11	Antonio Toledo Jr	Aceito
Outros	modelo_prova_treinamento_docente.pdf	23/10/2020 11:14:49	Antonio Toledo Jr	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	tcle_aluno_treinamento_docente.pdf	23/10/2020 11:14:00	Antonio Toledo Jr	Aceito

**Situação do Parecer:**

Aprovado

**Necessita Avaliação da CONEP:**

Não

ALFENAS, 31 de Março de 2023

Assinado por:  
**MARCELO REIS DA COSTA**  
(Coordenador(a))

**Endereço:** Rodovia MG 179 km 0, BLOCO VI SALA 602 1º ANDAR  
**Bairro:** Campus Universitário **CEP:** 37.130-000  
**UF:** MG **Município:** ALFENAS  
**Telefone:** (35)3299-3137 **Fax:** (35)3299-3137 **E-mail:** comitedeetica@unifenas.br