

**SALA DE AULA INVERTIDA
NO ENSINO DE
BIOESTATÍSTICA NA
GRADUAÇÃO EM
MEDICINA: ESTUDO QUASE-
EXPERIMENTAL DE
DESEMPENHO ACADÊMICO
E VARIÁVEIS DE
ENGAJAMENTO**

**FLIPPED CLASSROOM IN BIOSTATISTICS TEACHING FOR
UNDERGRADUATE MEDICAL STUDENTS: A QUASI-EXPERIMENTAL STUDY
OF ACADEMIC PERFORMANCE AND ENGAGEMENT VARIABLES**

Ciências da Saúde • 17/08/2026

REGISTRO DOI: [10.70773/revistatopicos/786933870](https://doi.org/10.70773/revistatopicos/786933870)

José Eduardo de Carvalho Lima¹

Ana Cristina Martins Sousa²

Maria Clara Santos Monte³

José Glauco Norões Xenofonte⁴

Gustavo de Alencar Tavares Norões⁵

RESUMO

A Bioestatística é competência fundante da prática baseada em evidências, mas o ensino expositivo alimenta a *estatofobia*, comprometendo a literacia estatística dos egressos. Avaliaram-se os efeitos da Sala de Aula Invertida (SAI) sobre o desempenho em Bioestatística na graduação em Medicina e, de forma exploratória, a associação de variáveis de engajamento extraclasse com o ganho. Empregou-se delineamento quase-experimental de grupo único, pré-teste/pós-teste, com 32 estudantes de uma turma regular de uma instituição brasileira, comparando-se as notas antes (AV1) e após (AV2) a intervenção. Violada a normalidade das diferenças, adotaram-se teste t pareado e Wilcoxon, d de Cohen, correlações de Spearman, Mann-Whitney, regressões e cluster *K-means*. A média subiu 0,95 ponto (IC 95% [0,44; 1,46]; $t(31)=3,780$; $p<0,001$; $d=0,668$), e a proporção de estudantes com nota inferior a 5,0 caiu de 25,0% para 9,4%. A frequência correlacionou-se negativamente com o ganho e os não usuários da biblioteca obtiveram ganho superior, padrões atribuíveis à menor proficiência inicial desses subgrupos. O modelo múltiplo explicou 30,7% da variância, com interações significativas envolvendo frequência, biblioteca e laboratório, e a análise de cluster identificou três perfis de engajamento. A SAI associou-se à melhora significativa e educacionalmente relevante, mais pronunciada entre estudantes de menor proficiência inicial; as análises de engajamento, de natureza exploratória, sugerem operação interativa e não linear que demanda abordagens diferenciadas por perfil.

Palavras-chave: Metodologias ativas; Desempenho acadêmico; Bioestatística; Educação médica.

ABSTRACT

Biostatistics is a foundational competency for evidence-based practice, but lecture-based instruction fosters statistics anxiety,

compromising the statistical literacy of graduates. The effects of the Flipped Classroom (FC) on Biostatistics performance in undergraduate medical education were evaluated, along with on an exploratory basis the association between out-of-class engagement variables and score gains. A single-group quasi-experimental pretest-posttest design was employed, with 32 students from a regular class at a Brazilian institution, comparing scores before (AV1) and after (AV2) the intervention. Because the normality of the differences was violated, paired t-test and Wilcoxon test, Cohen's d, Spearman correlations, Mann-Whitney test, regression analyses, and K-means clustering were used. The mean increased by 0.95 points (95% CI [0.44, 1.46]; $t(31) = 3.780$; $p < 0.001$; $d = 0.668$), and the proportion of students scoring below 5.0 decreased from 25.0% to 9.4%. Attendance was negatively correlated with the gain, and library non-users showed greater gains patterns attributable to the lower initial proficiency of these subgroups. The multiple regression model explained 30.7% of the variance, with significant interactions involving attendance, library use, and laboratory use; cluster analysis identified three engagement profiles. The FC was associated with a significant, educationally relevant improvement that was more pronounced among students with lower initial proficiency; the exploratory engagement analyses suggest that engagement operates interactively and nonlinearly, warranting differentiated approaches by profile.

Keywords: Active learning methodologies; Academic performance; Biostatistics; Medical education.

1. INTRODUÇÃO

A Estatística integra o núcleo formativo obrigatório dos cursos da área da saúde no Brasil, com destaque para a Medicina, sendo

reconhecida pelas Diretrizes Curriculares Nacionais (DCNs) como competência para a interpretação crítica da literatura científica e a decisão clínica baseada em evidências (Brasil, 2024; 2025). As novas DCNs de Medicina, aprovadas em 2025 (Parecer CNE/CES nº 536/2025; Resolução CNE/CES nº 3/2025), reforçam a adoção de metodologias ativas e o desenvolvimento da literacia estatística como pilares da formação médica contemporânea (Brasil, 2025; Oliveira et al., 2026).

Contraditoriamente, a disciplina figura entre os maiores desafios da formação em saúde. A literatura documenta o fenômeno da *estatofobia* combinação de ansiedade matemática, resistência epistemológica e baixo engajamento com conteúdos quantitativos que compromete o desempenho e a retenção do conhecimento estatístico (Silva et al., 2025; Barros et al., 2026). O modelo expositivo tradicional, que consome o tempo presencial com a transmissão de conteúdo acessível de forma autônoma, tem sido apontado como um dos fatores que perpetuam esse distanciamento; em contraste, a aprendizagem ativa reduz as taxas de reprovação em 1,5 vez e eleva o desempenho médio em 0,47 desvio-padrão em disciplinas STEM (Freeman et al., 2014).

Entre as metodologias ativas, a Sala de Aula Invertida (SAI) inverte a lógica instrucional convencional: transfere o primeiro contato com o conteúdo para o ambiente extraclasse, por meio de materiais digitais, e libera o tempo presencial para atividades de maior complexidade cognitiva (Bergmann; Bergmann; Sams, 2023). Sua eficácia é atribuída ao incremento da motivação intrínseca e da aprendizagem autorregulada (Abeysekera; Dawson, 2015). Meta-análises recentes confirmam efeito positivo e consistente da SAI sobre o desempenho na educação em saúde e em disciplinas

quantitativas (Naing et al., 2023; Spaic et al., 2025; Shi et al., 2025; Haziki et al., 2025).

Apesar desse avanço, permanece incipiente, sobretudo no contexto brasileiro a investigação de como variáveis de engajamento extraclasse (frequência e uso de recursos físicos e digitais) se associam ao ganho de desempenho sob a SAI. Especificamente, inexistem estudos que analisem de forma sistemática, com múltiplas variáveis de engajamento e identificação de perfis, a SAI aplicada à Bioestatística na graduação em Medicina no Brasil, lacuna que este trabalho se propõe a preencher.

O objetivo foi avaliar os efeitos da SAI sobre o desempenho acadêmico em Bioestatística, comparando as notas antes (AV1) e após (AV2) a intervenção, e investigar, em caráter exploratório, a associação de variáveis de engajamento extraclasse (frequência, uso da biblioteca, do laboratório e da plataforma digital) com o ganho de desempenho, incluindo análises de interação e perfis. Hipotetizou-se que a SAI se associaria a ganho de desempenho estatisticamente significativo, com benefício mais expressivo entre estudantes de menor proficiência inicial.

2. FUNDAMENTAÇÃO TEÓRICA OU REVISÃO DA LITERATURA

2.1. Sala de Aula Invertida: Fundamentos, Mecanismos e Evidências

A Sala de Aula Invertida (SAI), designação em português para *Flipped Classroom*, é uma das metodologias ativas mais utilizadas no ensino superior em saúde. Sua proposição central é a inversão da lógica instrucional convencional, o contato inicial com o conteúdo é transferido para o ambiente extraclasse, com suporte de materiais

digitais (videoaulas, *podcasts*, textos e questões-guia), liberando o tempo presencial para atividades de maior complexidade cognitiva (Bergmann; Bergmann; Sams, 2023). O referencial mais influente atribui sua eficácia a dois mecanismos, o incremento da motivação intrínseca, pela maior autonomia sobre o ritmo do aprendizado, e o fortalecimento da aprendizagem autorregulada, que demanda planejamento, monitoramento e avaliação das próprias estratégias de estudo (Abeysekera; Dawson, 2015).

Do ponto de vista da carga cognitiva, a antecipação do conteúdo em ambiente autocontrolado em que o estudante pausa, retoma e regula a velocidade, reduz a sobrecarga da memória de trabalho típica das exposições densas, favorecendo a consolidação na memória de longo prazo (Abeysekera; Dawson, 2015). No tempo presencial, as tarefas mobilizam os níveis superiores da Taxonomia de Bloom, analisar, avaliar e criar, em contraposição ao nível recordativo do modelo expositivo (Anderson; Krathwohl, 2001; Gopalan et al., 2024).

As evidências empíricas do último triênio consolidaram o modelo. Meta-análises de grande escala reportam efeito global favorável à SAI sobre o desempenho: Li et al. (2025), $g = 0,53$ (IC 95%: [0,46; 0,61]); Naing et al. (2023), em revisão *Campbell* com 31 estudos em ciências da saúde, $SMD = 0,57$ (IC 95%: [0,38; 0,77]); e Spaic et al. (2025), em revisão com 141 estudos e 21.171 participantes, que confirmaram a motivação intrínseca e a autorregulação como mediadores consistentes. Lo et al. (2025), em meta-análise de segunda ordem (46 revisões; 2.338 estudos), identificaram que o alinhamento entre estrutura instrucional e demandas cognitivas é o fator mais associado à magnitude do efeito.

Na educação médica, Shi et al. (2025) verificaram superioridade em desempenho (SMD = 0,481 a 0,642), satisfação e autoeficácia, e McLaughlin et al. (2014) documentaram, em redesenho curricular seminal em profissões da saúde, ganhos simultâneos de aprendizagem e engajamento. Os efeitos, contudo, não são uniformes, dependem do *design* instrucional, da percepção de utilidade das atividades e, sobretudo, da fidelidade de implementação, Olana, Bacha, Lemma, (2025) mostram declínio expressivo dos efeitos quando a adesão às atividades pré-aula é inferior a 70%.

2.2. Bioestatística, Estatofobia e o Papel das Metodologias Ativas

A Bioestatística é fundante da Medicina Baseada em Evidências (MBE), a literacia estatística capacidade de compreender, interpretar e avaliar criticamente dados e inferências quantitativas, é condição *sine qua non* para seu exercício competente (Barros et al., 2026). Menezes, Petraconi e Martins (2026) evidenciaram que grande parte dos estudantes da área da saúde apresenta dificuldades na interpretação de medidas estatísticas, intervalos de confiança e valores de p , lacunas atribuídas a modelos pedagógicos que privilegiam a memorização de procedimentos. No plano normativo, as DCNs de Medicina reformadas pela Resolução CNE/CES nº 3/2025 estabelecem entre as competências do egresso “avaliar criticamente estudos científicos, interpretar dados epidemiológicos e estatísticos e aplicar a medicina baseada em evidências” (Brasil, 2025, art. 5º), elevando a literacia estatística de competência acessória a pilar estrutural da formação (Oliveira et al., 2026).

Um dos principais obstáculos a essa competência é a *estatofobia* combinação de ansiedade matemática, resistência epistemológica e

baixo engajamento com conteúdos quantitativos, de raízes multifatoriais (experiências negativas prévias, baixa autoeficácia, percepção de irrelevância clínica e ensino centrado em fórmulas) (Silva et al., 2025). Barros et al. (2026), em análise qualitativa com 48 estudantes de Medicina de três instituições brasileiras, identificaram três dimensões interdependentes do fenômeno cognitiva, afetiva e comportamental, associadas a reprovação acima da média curricular, baixa retenção e dificuldade de aplicação clínica posterior. O modelo expositivo tradicional agrava o problema ao não integrar conceitos abstratos, aplicação em *softwares* e interpretação clínica; em contraste, a aprendizagem ativa reduz as taxas de reprovação em 1,5 vez e eleva o desempenho médio em 0,47 desvio-padrão em disciplinas STEM (Freeman et al., 2014; Naing et al., 2023).

As evidências específicas sobre SAI em disciplinas quantitativas convergem: Evaristo-Chiyong e Mattos-Vela (2019) reportaram melhora de desempenho e de atitude entre estudantes de Odontologia; Haziki et al. (2025), em meta-análise com 38 estudos no ensino superior de Matemática, encontraram $g = 0,402$ (IC 95%: [0,31; 0,49]); Ilgaz (2025) associou a SAI com gamificação ao aumento da motivação autônoma em Bioestatística; e Althubaiti e Althubaiti (2024) verificaram redução da ansiedade e aumento da autoeficácia no ensino de *softwares* estatísticos. Esses achados sustentam a hipótese de que a SAI atua sobre a estatofobia por dupla via redução da carga cognitiva e aumento da autoeficácia (Abeysekera; Dawson, 2015).

2.3. Engajamento Extraclasse, Recursos Digitais e Perfis de Aprendizagem

O engajamento estudantil, em sua dimensão comportamental, manifesta-se em condutas observáveis frequência às aulas, uso de recursos institucionais (biblioteca, laboratório e AVA) e dedicação extraclasse. Na SAI, ao tornar o contato inicial com o conteúdo uma atividade extraclasse necessária à participação ativa, o uso desses recursos deixa de ser periférico e torna-se central à aprendizagem (Abeysekera; Dawson, 2015). Khadawardi et al. (2025) identificaram o uso do AVA e a conclusão das atividades pré-aula como mediadores robustos do efeito da SAI em Medicina. A frequência, porém, guarda relação mais complexa do que a suposta pelo senso comum, estudantes menos assíduos frequentemente os de menor proficiência inicial podem compensar a menor presença com maior engajamento com os materiais pré-aula e recursos alternativos (Maryami et al., 2025). Kiljunen et al. (2024) demonstraram que universitários com histórico de dificuldades de aprendizagem desenvolvem estratégias que substituem funcionalmente a presença uso intensivo de plataformas, laboratórios e grupos de estudo, de modo que a frequência, tomada como *proxy* único, pode subestimar o efeito da SAI sobre perfis alternativos.

O AVA é a infraestrutura tecnológica central do modelo, e a literatura avançou da mera presença/ausência de uso para a intensidade aferida por *logs*, Deng, Feng e Shen (2024) demonstraram, em estudo experimental com 124 universitários, que questões embutidas nos vídeos aumentaram em 37% o tempo de engajamento com os materiais pré-aula, e Olana, Bacha, Lemma, (2025) identificaram que a taxa de conclusão das atividades pré-aula supera, em poder preditivo, a frequência às aulas. Os recursos físicos operam como espaços complementares: a biblioteca oferece acervo e bases científicas relevantes para disciplinas metodológicas, e o laboratório de informática é insubstituível para competências

práticas em *softwares* estatísticos previstas nas DCNs 2025 e associadas à redução da ansiedade quando integradas a metodologias ativas (Althubaiti; Althubaiti, 2024; Gopalan et al., 2024).

A investigação de perfis de aprendizagem tem emergido como abordagem promissora. Técnicas de *cluster*, como o *K-means*, identificam grupos naturais a partir de múltiplas variáveis de engajamento, superando análises univariadas e revelando a heterogeneidade que as médias do grupo obscurecem (Hair et al., 2019). Essa heterogeneidade é central, Kiljunen et al. (2024) evidenciaram que estudantes com dificuldades se beneficiam de forma equivalente ou superior quando há suporte adequado, e Maryami et al. (2025) encontraram benefício inversamente proporcional ao desempenho inicial, sugerindo eficácia da SAI como estratégia de recuperação e inclusão. Daí decorrem três implicações que orientam o presente estudo: (1) análises baseadas na média podem ocultar benefícios de subgrupos; (2) intervenções diferenciadas por perfil podem amplificar o efeito geral; e (3) a identificação precoce de perfis de risco, por variáveis de engajamento observáveis, pode orientar suporte proativo, em consonância com a educação inclusiva preconizada pelas DCNs 2025 (Brasil, 2025; Oliveira et al., 2026).

3. METODOLOGIA

O estudo adotou delineamento quase-experimental de grupo único, pré-teste/pós-teste, amplamente utilizado em pesquisas educacionais quando a randomização não é operacionalmente viável (Campbell; Stanley, 2015; Jannuzzi, 2023). Foi conduzido em uma turma regular de Bioestatística do curso de graduação em

Medicina de uma instituição de ensino superior brasileira, ao longo de um semestre letivo completo. A turma era inicialmente composta por 37 estudantes matriculados. Para a composição da amostra, adotaram-se critérios de inclusão: (a) matrícula ativa durante todo o semestre; (b) participação em ambas as avaliações (AV1 e AV2); (c) frequência mínima de 75% nas aulas presenciais e de exclusão: ausência em qualquer das avaliações; trancamento ou abandono após o início da intervenção; frequência inferior a 75% no período SAI. Após sua aplicação, a amostra final foi de $n = 32$ estudantes com participação integral em ambos os instrumentos. Por se tratar de uma turma completa, a amostra foi censitária entre os elegíveis, sem cálculo amostral a priori.

A intervenção seguiu o modelo canônico da SAI (Bergmann; Bergmann; Sams, 2023; Abeysekera; Dawson, 2015), com o semestre estruturado em duas fases pedagógicas. A Fase I, pré-intervenção, adotou o modelo expositivo tradicional. A Fase II operacionalizou a SAI em dois momentos: (a) etapa pré-aula, videoaulas de 12–20 min, *podcasts* de 10–14 min, leituras e questões-guia disponibilizadas com cinco dias de antecedência no AVA institucional (Deng, Feng e Shen, 2024); e (b) etapa presencial, resolução colaborativa de problemas estatísticos contextualizados em cenários clínicos e revisão mediada pelo docente (Gopalan et al., 2024). A fidelidade de implementação foi monitorada semanalmente pelos registros de acesso ao AVA e pelas questões-guia respondidas (Olana, Bacha, Lemma, 2025).

O desempenho acadêmico foi operacionalizado pelas notas (escala 0–10) em duas avaliações somativas estruturalmente equivalentes, AV1 (pré-intervenção, ao final da Fase I) e AV2 (pós-intervenção, ao final da Fase II). A equivalência de formato, distribuição de pontos e nível cognitivo, segundo a Taxonomia de Bloom revisada (Anderson;

Krathwohl, 2001), foi assegurada por rubrica padronizada. Ressalve-se que a equivalência se refere ao formato e ao nível cognitivo, e não ao conteúdo, específico de cada fase. O desfecho primário foi o ganho de desempenho (AV2 - AV1), os desfechos secundários corresponderam às associações desse ganho com as variáveis de engajamento extraclasse.

Quatro variáveis de engajamento foram registradas durante a Fase II. A frequência escolar (%), contínua, reflete o percentual de presença nas atividades presenciais e foi categorizada em Alta ($\geq 90\%$), Média (75–89%) e Baixa ($< 75\%$). As demais são dicotômicas (Sim/Não), o uso da biblioteca (recursos físicos ou digitais da biblioteca institucional), o uso do laboratório (atividades nos laboratórios de informática ou de análise de dados) e o uso da plataforma digital (acesso ao AVA institucional).

As análises foram conduzidas em Python 3.11 (NumPy, SciPy, statsmodels e scikit-learn), com nível de significância $\alpha = 0,05$, e todas as etapas foram verificadas de forma independente no Jamovi 2.3, garantindo maior confiabilidade dos resultados. A comparação AV1 vs. AV2 constituiu a análise confirmatória principal, as análises das variáveis de engajamento tiveram caráter exploratório. Para o desempenho (AV1 vs. AV2), calcularam-se estatísticas descritivas completas (média, desvio-padrão, mediana, mínimo, máximo e intervalo interquartilico), a normalidade das diferenças foi avaliada pelo teste de Shapiro-Wilk, e a comparação entre os momentos seguiu estratégia dupla, com o teste *t* pareado e o *Wilcoxon signed-rank*, estimando-se o tamanho de efeito pelo *d* de Cohen.

A frequência escolar foi examinada por correlações de *Pearson* e *Spearman* com o ganho, comparação entre categorias por ANOVA

one-way e *Kruskal-Wallis*, e regressão linear simples. Para o uso da biblioteca e do laboratório, examinaram-se as proporções de participação, compararam-se os ganhos por *Mann-Whitney* e verificou-se a associação com a melhora ($AV2 > AV1$) por qui-quadrado e por regressão logística binária, com *odds ratio* e IC 95%. O uso da plataforma digital foi descrito por estatísticas de adesão. As análises multivariadas incluíram a matriz de correlações de *Spearman* e a regressão linear múltipla (frequência, biblioteca e laboratório), com avaliação da multicolinearidade pelo VIF e modelos adicionais com termos de interação (variáveis centradas na média).

Por fim, a análise de agrupamentos (*K-means* com padronização *Z-score*) teve o número ótimo de grupos determinado pelo coeficiente de silhueta, com comparações de desempenho entre os grupos por *Kruskal-Wallis* e *Mann-Whitney* em análises par a par.

A investigação caracteriza-se como pesquisa educacional baseada em registros de desempenho já produzidos no curso regular, utilizados de forma agregada e anonimizada, o que a enquadra nas hipóteses de dispensa de registro e avaliação pelo Sistema CEP/CONEP previstas no art. 1º da Resolução CNS nº 510/2016 (incisos V e VII). Ainda que dispensado de submissão, o estudo observou integralmente os princípios de respeito à privacidade, confidencialidade e não maleficência, à luz da Resolução CNS nº 510/2016 e da Lei nº 14.874/2024.

4. RESULTADOS E DISCUSSÕES

Os resultados são apresentados em sete blocos temáticos, cada qual seguido de discussão integrada referenciada pela literatura pertinente. Essa estrutura articula, de forma imediata, os achados

empíricos às implicações teóricas e pedagógicas que deles derivam, em consonância com o delineamento adotado (Campbell; Stanley, 2015). A comparação AV1 vs. AV2 (4.1) constitui a análise confirmatória; os blocos relativos às variáveis de engajamento (4.2 a 4.7) têm caráter exploratório, dado o tamanho amostral e a multiplicidade de testes e devem ser lidos como geradores de hipóteses.

4.1. Desempenho Acadêmico: AV1 vs. AV2

A Tabela 1 apresenta as estatísticas descritivas completas para AV1, AV2 e as diferenças individuais. A nota média passou de 5,86 (DP = 1,46) na avaliação pré-intervenção para 6,81 (DP = 1,36) na pós-intervenção, com diferença média de +0,95 ponto (IC 95%: [0,44; 1,46]). A mediana da diferença (0,45) e o IQR amplo (2,63) antecipam heterogeneidade nas trajetórias individuais.

Tabela 1. Estatísticas descritivas de desempenho acadêmico (n = 32)

Estatística	AV1 (Pré-SAI)	AV2 (Pós-SAI)	Diferença (AV2-AV1)
n	32	32	32
Média	5,86	6,81	0,95
Desvio-padrão	1,46	1,36	1,42
Mediana	6,00	6,90	0,45
Mínimo	3,00	4,00	-1,00
Máximo	8,50	9,00	+3,50
Q1	4,95	5,95	-0,33
Q3	7,00	8,00	+2,30

IQR	2,05	2,05	2,63
IC 95% da diferença	—	—	[0,44; 1,46]

Fonte: Elaborado pelos autores com suporte do *software* estatístico Jamovi, 2026.

A análise das trajetórias individuais revela que 22 estudantes (68,8%) melhoraram de AV1 para AV2, enquanto 10 (31,2%) apresentaram queda. Os cinco maiores ganhos (A10: +3,5; A25: +3,4; A21: +3,0; A30: +2,8; A03, A06 e A08: +2,5) foram observados em estudantes com AV1 $\leq 5,2$, ao passo que as quedas se concentraram entre aqueles com desempenho inicial superior (AV1 $\geq 6,0$). Esse padrão diferencial de especial relevância pedagógica será retomado nas análises de frequência e cluster (seções 3.2 e 3.7). O número de estudantes com nota abaixo de 5,0 reduziu-se de 8 (25,0%) para 3 (9,4%), evidenciando um efeito de inclusão pedagógica da SAI relevante para contextos de formação em saúde, onde a reprovação em Bioestatística é historicamente elevada (Silva et al., 2025).

O teste de *Shapiro-Wilk* aplicado às diferenças individuais retornou $W = 0,912$; $p = 0,013$, indicando violação estatisticamente significativa da normalidade. As distribuições de AV1 ($W = 0,979$; $p = 0,766$) e AV2 ($W = 0,970$; $p = 0,511$) são adequadas, porém a assimetria positiva moderada das diferenças reflexo dos ganhos extremos nos quartis inferiores justifica a adoção de estratégia de análise dupla, teste *t* pareado (sustentado pelo Teorema Central do Limite para $n \geq 30$; Field, 2024) e Wilcoxon signed-rank como contraprova não-paramétrica (Grawe, 2016).

Ambos os testes confirmaram diferença estatisticamente significativa (Tabela 2). O teste *t* pareado [$t(31) = 3,780$; $p < 0,001$] e o *Wilcoxon signed-rank* [$W = 121,5$; $p = 0,008$] convergem para a rejeição da hipótese nula. O *d* de Cohen para amostras pareadas foi de 0,668, situado na faixa de efeito médio a grande (Cohen, 1988, 2013).

Tabela 2. Resultados dos testes inferenciais para comparação AV1 vs. AV2 ($n = 32$)

Teste / Medida	Estatística	gl	p-valor	Observação
Shapiro-Wilk (diferenças)	$W = 0,912$	31	$p = 0,013$	Violação de normalidade
Teste <i>t</i> pareado	$t = 3,780$	31	$p < 0,001$	IC 95%: [0,44; 1,46]
Wilcoxon signed-rank	$W = 121,5$	—	$p = 0,008$	Confirma teste <i>t</i>
<i>d</i> de Cohen (pareado)	$d = 0,668$	—	—	Efeito médio a grande

*** $\alpha = 0,05$.**

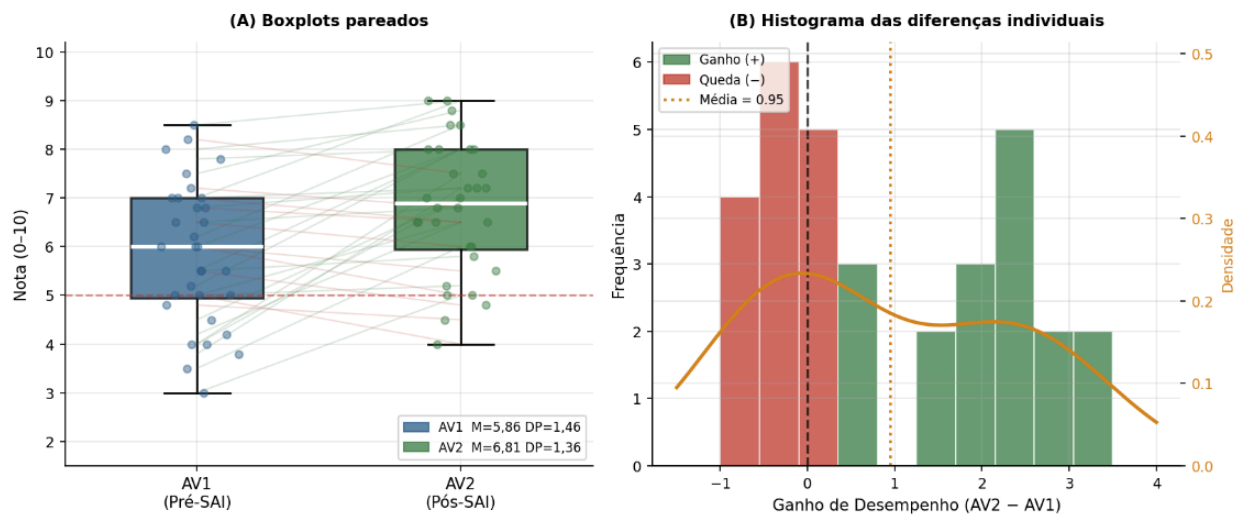
Fonte: Elaborado pelos autores com suporte do *software* estatístico Jamovi, 2026.

O valor $d = 0,668$ situa-se acima das médias reportadas pelas três principais meta-análises recentes, Naing *et al.* (2023), em revisão com 31 estudos em saúde, encontraram $SMD = 0,57$ (IC 95%: [0,38; 0,77]); Li *et al.* (2025), em meta-análise de terceiro nível com 129 estudos e 12.727 participantes, reportaram $g = 0,53$; Lo *et al.* (2025), em meta-análise de segunda ordem, identificaram efeito global $g = 0,402$. O posicionamento ligeiramente acima dessas médias é

plausível considerando o tamanho amostral reduzido e a ausência de grupo controle condições que, reconhecidamente, tendem a inflar estimativas de efeito em delineamentos pré-pós de grupo único (Naing *et al.*, 2023; Spaic *et al.*, 2025).

A convergência entre os testes paramétrico e não-paramétrico fortalece a potência dos achados. Do ponto de vista pedagógico, o efeito médio a grande é coerente com o mecanismo proposto por Abeysekera e Dawson (2015), a SAI, ao transferir o primeiro contato com o conteúdo para o espaço extraclasse, libera o tempo presencial para atividades de maior complexidade cognitiva, potencializando a aprendizagem especialmente em disciplinas que exigem aplicação e resolução de problemas, como a Bioestatística. Wang *et al.* (2025) corroboraram esse argumento ao demonstrar que a SAI aumenta significativamente a motivação intrínseca universitária ($d = 0,634$ a $1,331$ por domínio), e Khadawardi *et al.* (2025) reportaram ganhos de motivação e aprendizagem entre pares em estudantes de medicina submetidos ao modelo invertido.

Figura 1. Desempenho AV1 vs. AV2. (A) Boxplots pareados AV1 vs. AV2 com pontos individuais sobrepostos; linhas indicam trajetória individual (verde = melhora; vermelho = queda). (B) Histograma das diferenças individuais com curva de densidade. Barras verdes = ganho positivo; vermelhas = queda; linha âmbar = média (+0,95).



Fonte: Elaborado pelos autores com suporte da linguagem Python, 2026.

A Figura 1(A) dos Boxplots pareados revela o deslocamento ascendente das distribuições de AV1 para AV2. A caixa de AV2 (verde) está posicionada acima da de AV1 (azul), a mediana de AV2 (~6,9) supera o terceiro quartil de AV1 (7,0), e o Q1 de AV2 (~5,95) aproxima-se da mediana de AV1 (6,0), indicando que metade inferior da turma na AV1 migrou predominantemente para a metade superior na AV2. As linhas de trajetória individual confirmam a direção majoritária do movimento, as linhas verdes (melhora) são mais numerosas e, em sua maioria, mais inclinadas para cima entre os pontos de AV1 mais baixos, evidenciando ganhos mais expressivos nos quartis inferiores. As linhas vermelhas (queda), em menor número, partem de AV1 elevado ($\geq 6,0$) e mostram declínios modestos (até -1,0 ponto). A linha de referência em 5,0 pontos mostra que a quantidade de pontos abaixo dela reduziu visivelmente de AV1 para AV2, corroborando o efeito de inclusão pedagógica da SAI.

A Figura 1(B) dos Histogramas das diferenças individuais revela a distribuição assimétrica positiva (inclinada para a direita) das diferenças AV2 - AV1, consistente com o resultado do *Shapiro-Wilk* ($p = 0,013$). O histograma apresenta duas regiões distintas, uma zona de queda (barras vermelhas, ganho ≤ 0), com o pico de frequência ao redor de 0 (cinco estudantes), e outra de melhora (barras verdes),

com pico secundário entre +2,0 e +2,5 pontos (cinco estudantes). A cauda à direita se estende até +3,5, correspondendo ao maior ganho individual (A10). A curva de densidade KDE (linha âmbar) apresenta formato bimodal suave, com um primeiro pico próximo a zero e uma região de alta densidade entre +2,0 e +2,5, padrão que reflete a heterogeneidade da resposta à SAI e que fundamenta a escolha do Wilcoxon como contra-prova não-paramétrica. A linha tracejada vertical da média (+0,95) posiciona-se à direita do modo principal, arrastada pelos ganhos extremos da cauda, sinal clássico de assimetria positiva.

4.2. Frequência Escolar

A frequência escolar média durante a Fase II foi de 85,72% (DP = 6,49%), variando de 74% a 96%. A Tabela 3 apresenta a distribuição por categorias e os ganhos correspondentes. Destaca-se que 34,4% dos estudantes situaram-se na categoria Alta ($\geq 90\%$) e 62,5% na faixa Média (75–89%). Apenas um estudante (A25, 74%) integrou a categoria Baixa ($< 75\%$), o que impõe cautela analítica para esse grupo.

Tabela 3. Distribuição da frequência escolar e ganho de desempenho por categoria (n = 32)

Categoria	n	%	Ganho Médio	Desvio-Padrão	Mediana do Gar
Alta ($\geq 90\%$)	11	34,4%	0,527	0,978	0,30
Média (75–89%)	20	62,5%	1,055	1,531	0,95

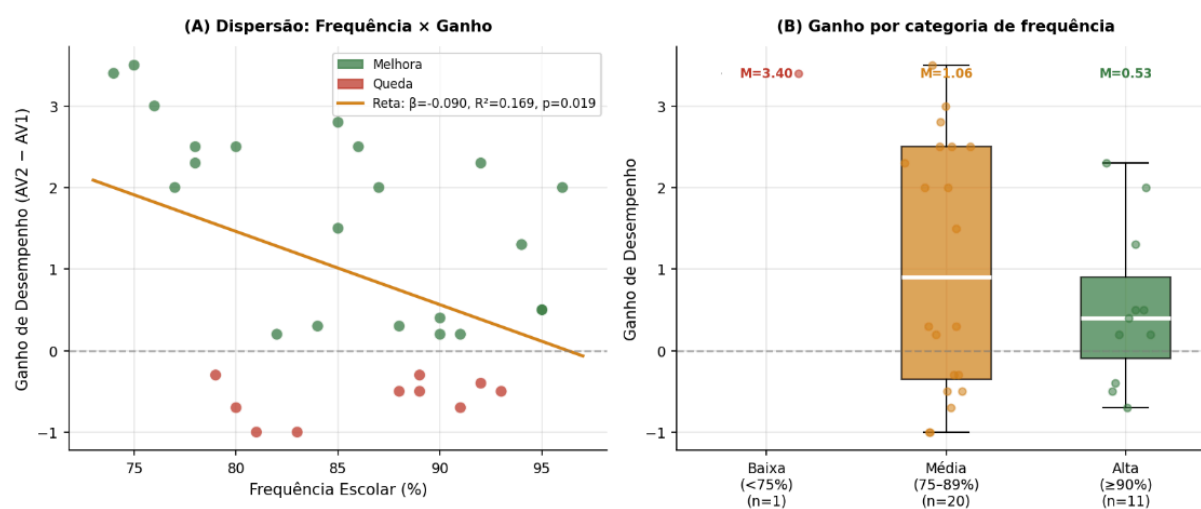
Baixa (<	1	3,7%	3,400	n/a	3,40
----------	---	------	-------	-----	------

⚠ Esta tabela possui muitas colunas e foi cortada para impressão. Para visualizá-la completa, acesse o artigo original em: <https://revistatopicos.com.br/artigos/sala-de-aula-invertida-no-ensino-de-bioestatistica-na-graduacao-em-medicina-estudo-quase-experimental-de-desempenho-academico-e-variaveis-de-engajamento?noblockage>

Nota: Com $n = 1$ na categoria Baixa, as estimativas desta linha devem ser interpretadas com extrema cautela.

Fonte: Elaborado pelos autores com suporte do software estatístico Jamovi, 2026.

Figura 2. Frequência Escolar e Ganho de Desempenho. (A) Dispersão Frequência × Ganho com reta de regressão ($\beta = -0,090$; $R^2 = 0,169$; $p = 0,019$); pontos verdes = melhora, vermelhos = queda. (B) Boxplots do ganho por categoria de frequência com jitter.



Fonte: Elaborado pelos autores com suporte da linguagem Python, 2026.

A Figura 2(A) apresenta a nuvem de pontos que evidencia a inclinação negativa da reta de regressão ($\beta = -0,090$; $R^2 = 0,169$), que desce de aproximadamente +2,1 pontos no extremo inferior da frequência (74%) até próximo de zero no extremo superior (96%).

Pontos verdes (melhora) concentram-se nitidamente na região de menor frequência (74–82%), onde ganhos de +2,0 a +3,5 pontos são predominantes.

Na faixa de frequência alta (90–96%), observa-se uma mescla de pontos verdes e vermelhos (queda), com ganhos predominantemente abaixo de +0,5, padrão visual que corrobora a correlação negativa. A dispersão em torno da reta é expressiva ($R^2 = 0,169$ explica apenas 16,9% da variância), sinalizando que a frequência é preditor parcial e que outras variáveis explicam o restante do ganho justificando os modelos multivariados.

Já a Figura 2(B) mostra gradiente visual decrescente de ganho médio à medida que a frequência aumenta, $M = 3,40$ (Baixa, $n=1$, ponto isolado, sem caixa), $M = 1,06$ (Média, $n=20$) e $M = 0,53$ (Alta, $n=11$). A caixa da categoria Média (75–89%) apresenta o maior IQR de todas (de aproximadamente $-0,8$ a $+2,5$), refletindo a alta heterogeneidade interna desse grupo, que concentra 62,5% da amostra e inclui tanto ganhos expressivos quanto quedas significativas. A caixa da categoria Alta ($\geq 90\%$) é mais compacta e deslocada para baixo, com mediana próxima a $+0,40$, e whisker inferior atingindo $-1,0$ indicando que estudantes de alta frequência, embora mais homogêneos, incluem casos de queda expressiva. O único ponto da categoria Baixa (A_{25} , $+3,4$) é o maior ganho individual de toda a amostra, posicionando-se como outlier visual que ilustra de forma extrema o padrão de composição por proficiência inicial.

A correlação de Spearman entre frequência e ganho foi $r_s = -0,328$ ($p = 0,066$), tendência próxima ao limiar de significância, confirmada pela correlação de Pearson: $r = -0,411$ ($p = 0,019$). A ANOVA *one-way*

($F(2,29) = 2,196$; $p = 0,129$) e o *Kruskal-Wallis* ($H = 2,988$; $p = 0,224$) não atingiram significância entre categorias, em parte pelo poder estatístico reduzido com $n = 32$ e apenas um caso na faixa Baixa. A regressão linear simples confirmou a relação negativa: $\text{Ganho} = 8,647 + (-0,090) \times \text{Frequência}$; $\beta = -0,090$; $R^2 = 0,169$; $p = 0,019$ (IC 95%: $[-0,164; -0,016]$), indicando que cada ponto percentual adicional de frequência se associou a redução de 0,09 ponto no ganho predito.

A correlação negativa entre frequência e ganho é, à primeira vista, contraintuitiva. Sua correta interpretação requer o reconhecimento de um efeito de composição por proficiência inicial, estudantes com maior frequência apresentavam AV1 médio mais elevado e, portanto, menor margem absoluta de crescimento. Esse mecanismo é distinto de um efeito causal da frequência sobre o ganho e seria adequadamente controlado por modelos de regressão com AV1 como covariável (ANCOVA) análise recomendada em estudos futuros.

O padrão é coerente com o modelo de Abeysekera e Dawson (2015), que atribui à SAI efeito mais pronunciado sobre motivação intrínseca e autorregulação processos especialmente mobilizados em estudantes com menor proficiência inicial e maior necessidade de adaptação estratégica. Maryami et al. (2025), em ensaio randomizado por conglomerados, corroboraram esse padrão ao documentar benefícios desproporcionalmente maiores da SAI em estudantes de enfermagem com menor desempenho inicial.

4.3. Uso da Biblioteca

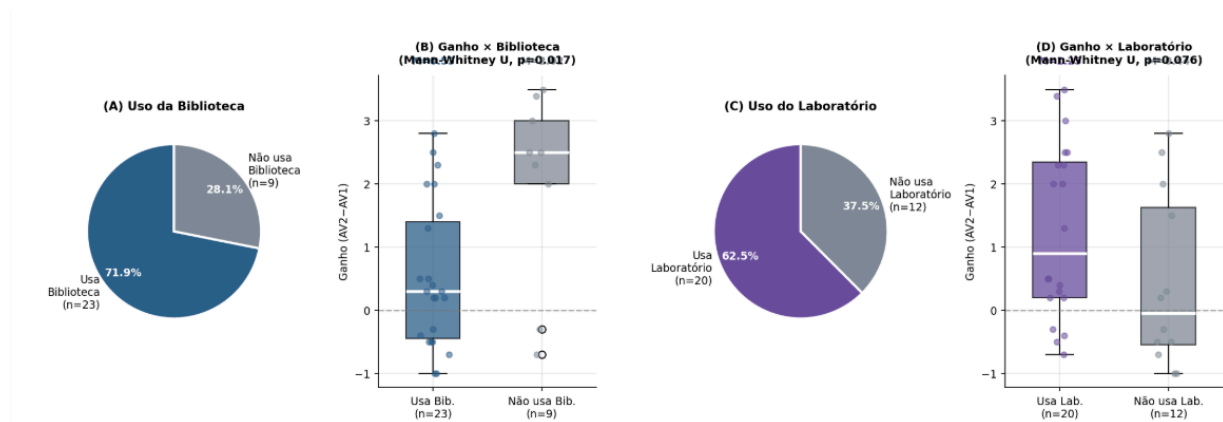
Dos 32 estudantes, 23 (71,9%) utilizaram a Biblioteca durante a Fase II e 9 (28,1%) não a utilizaram. Os não-usuários apresentaram ganho

médio substancialmente superior: $M = 2,022$ ($DP = 1,516$) vs. $M = 0,526$ ($DP = 1,156$), diferença estatisticamente significativa pelo Mann-Whitney: $U = 46,0$; $p = 0,017$, com correlação rank-biserial de $r = 0,556$ (efeito grande). O teste qui-quadrado para associação com melhora ($AV2 > AV1$) não foi significativo ($\chi^2 = 0,070$; $gl = 1$; $p = 0,791$), e a regressão logística indicou $OR = 0,536$ (IC 95%: $[0,089; 3,210]$; $p = 0,495$). Esses três resultados conjuntos revelam que a probabilidade de melhorar é similar entre os grupos, mas a magnitude do ganho difere expressivamente.

O achado aparentemente paradoxal de que não utilizar a biblioteca se associa a maior ganho tem explicação estrutural coerente. Estudantes que não usaram a biblioteca apresentavam AV1 médio de 4,14, enquanto os usuários tinham AV1 médio de 6,49. Trata-se, portanto, de um efeito de teto diferenciado, os não-usuários partiam de notas mais baixas, possuíam maior margem absoluta de crescimento e foram justamente aqueles mais beneficiados pela SAI. Adicionalmente, a análise de *cluster* demonstrará que esses estudantes tenderam a compensar a ausência do recurso biblioteca com uso mais intensivo do laboratório configurando uma estratégia alternativa de engajamento.

Kiljunen et al. (2024), em investigação com universitários com histórico de dificuldades de aprendizagem, documentaram que esse perfil frequentemente substitui recursos convencionais por recursos alternativos com igual ou maior eficácia. A interpretação deve ser acompanhada de cautela: não se recomenda, em hipótese alguma, desencorajar o uso da biblioteca; os achados refletem diferenças de ponto de partida, não impacto negativo do recurso.

Figura 3. Uso de recursos extraclasse: Biblioteca e Laboratório. (A) Proporção de uso da Biblioteca. (B) Distribuição do ganho por uso de Biblioteca (Mann-Whitney $p = 0,017^*$). (C) Proporção de uso do Laboratório. (D) Distribuição do ganho por uso de Laboratório (Mann-Whitney $p = 0,076$).



Fonte: Elaborado pelos autores com suporte da linguagem Python, 2026.

O gráfico de pizza 3(A) ilustra a assimetria na adoção do recurso, a fatia azul majoritária (71,9%, $n=23$) representa os usuários, enquanto o segmento cinza (28,1%, $n=9$) representa os não-usuários, proporção relevante para um recurso institucional. O boxplot do Gráfico 3(B) exibe separação visual notável entre os grupos, a caixa do grupo "Não usa Biblioteca" situa-se inteiramente acima da do grupo "Usa Biblioteca". A mediana dos não-usuários ($\sim 2,5$) ultrapassa o Q3 dos usuários ($\sim 1,4$), e os *whiskers* dos dois grupos apresentam sobreposição mínima configuração visual consistente com um tamanho de efeito grande ($r_{rb} = 0,556$). Os dois pontos atípicos do grupo "Não usa Biblioteca" (outliers brancos em $\sim -0,7$), embora representem exceções, não alteram o padrão de superioridade do ganho médio desse grupo. No grupo "Usa Biblioteca", a dispersão é ampla (de $-1,0$ a $+2,5$), indicando heterogeneidade interna que reflete diferenças de proficiência inicial entre seus membros.

O gráfico de pizza 3(C) mostra distribuição menos assimétrica do que a biblioteca, 62,5% (n=20) usaram o laboratório e 37,5% (n=12) não usaram. O boxplot 3(D) revela padrão oposto ao da biblioteca, porém com sobreposição intergrupar maior, a caixa do grupo "Usa Lab." (roxa) está deslocada para cima em relação à de "Não usa Lab." (cinza), com medianas visivelmente diferentes ($\sim 0,9$ vs. $\sim -0,05$). A caixa do grupo "Usa Lab." apresenta Q1 próximo de zero e Q3 em torno de +2,3 indicando que a maioria dos usuários do laboratório melhorou. Em contraste, a mediana de "Não usa Lab." está próxima de zero, com metade inferior da caixa adentrando o campo negativo, o que sinaliza que a probabilidade de queda é maior nesse grupo. A sobreposição parcial dos whiskers explica a ausência de significância formal ($p=0,076$), enquanto o OR=4,00 na regressão logística captura a direção clinicamente relevante do efeito.

4.4. Uso do Laboratório

Dos 32 estudantes, 20 (62,5%) participaram de atividades de laboratório e 12 (37,5%) não participaram. Usuários apresentaram ganho médio superior: $M = 1,250$ ($DP = 1,379$) vs. $M = 0,442$ ($DP = 1,389$). O Mann-Whitney indicou tendência: $U = 166,0$; $p = 0,076$ abaixo do limiar convencional de $\alpha = 0,05$, porém com relevância prática expressiva. O qui-quadrado não foi significativo ($\chi^2 = 1,901$; $gl = 1$; $p = 0,168$). A regressão logística revelou OR = 4,000 (IC 95%: [0,828; 19,325]; $p = 0,085$): estudantes que usaram o laboratório tinham probabilidade quatro vezes maior de apresentar melhora. O IC 95% amplo reflete diretamente a limitação do tamanho amostral com $n \geq 80$, o poder estatístico seria suficiente para detectar um OR dessa magnitude.

Tabela 4. Comparação de ganho e regressão logística por uso de recursos extraclasse (n = 32)

Grupo	n (%)	Ganho M	DP	Mediana	Test
Usa Biblioteca	23 (71,9%)	0,526	1,156	0,30	U=46 p=0,01
Não usa Biblioteca	9 (28,1%)	2,022	1,516	2,50	—
Usa Laboratório	20 (62,5%)	1,250	1,379	0,90	U=166 p=0,01
Não usa	12 (37,5%)	0,442	1,389	0,05	—

⚠ Esta tabela possui muitas colunas e foi cortada para impressão. Para visualizá-la completa, acesse o artigo original em: <https://revistatopicos.com.br/artigos/sala-de-aula-invertida-no-ensino-de-bioestatistica-na-graduacao-em-medicina-estudo-quase-experimental-de-desempenho-academico-e-variaveis-de-engajamento?noblockage>

* $p < 0,05$.

Fonte: Elaborado pelos autores com suporte do *software* estatístico Jamovi, 2026.

O OR = 4,00 para o laboratório, embora não significativo pelo critério convencional, constitui uma estimativa de relevância prática inegável no contexto pedagógico. A ausência de significância formal deve ser lida à luz das limitações de poder com $n = 32$, e não como evidência de ausência de efeito. Gopalan et al. (2024) demonstraram que atividades práticas em laboratório potencializam os efeitos da SAI em disciplinas biomédicas, especialmente para conteúdos de aplicação exatamente o foco da Fase II deste estudo.

A modelagem de interação revelará que esse efeito é moderado pela frequência escolar: o laboratório é especialmente benéfico para estudantes com menor presença nas aulas expositivas, funcionando como recurso de recuperação e aprendizagem alternativa padrão consistente com os achados sobre o Cluster C.

4.5. Plataforma Digital

Todos os 32 estudantes (100%) acessaram a plataforma digital (AVA) durante a Fase II. A adesão universal constitui, em si mesma, um resultado educacionalmente relevante no contexto brasileiro, a integração do AVA ao ensino presencial frequentemente encontra resistência e subutilização. O dado sugere que a SAI ao tornar os materiais digitais (videoaulas, podcasts, questões-guia) estruturalmente necessários para participação ativa nas aulas presenciais opera como indutor natural de engajamento com a plataforma (Deng; Gao, 2024).

A ausência de variabilidade nessa variável dicotômica impossibilita qualquer análise comparativa intergrupar. Para investigações de relação dose-resposta, a coleta de variáveis contínuas de uso, número de acessos ao AVA, tempo de permanência, módulos completados, questões-guia respondidas é indispensável em estudos futuros. A literatura recente tem demonstrado que a intensidade de uso da plataforma digital modera o efeito da SAI de forma significativa (Olana, Bacha, Lemma, 2025); Deng; Gao, 2024), tornando essa limitação analítica uma prioridade para a agenda de pesquisa.

4.6. Análises Multivariadas e Termos de Interação

A Tabela 5 apresenta as correlações de *Spearman* entre as variáveis extraclasse e o ganho de desempenho. O achado de maior implicação metodológica é a alta correlação entre Frequência e Biblioteca: $r_s = +0,780$; $p < 0,001$, indicando que, nessa amostra, os comportamentos de "estudante frequente" e "usuário da biblioteca" tendem a co-ocorrer, configurando um perfil único de engajamento convencional que variáveis isoladas não conseguem decompor adequadamente. Essa correlação tem consequência direta sobre a estimação dos modelos de regressão múltipla.

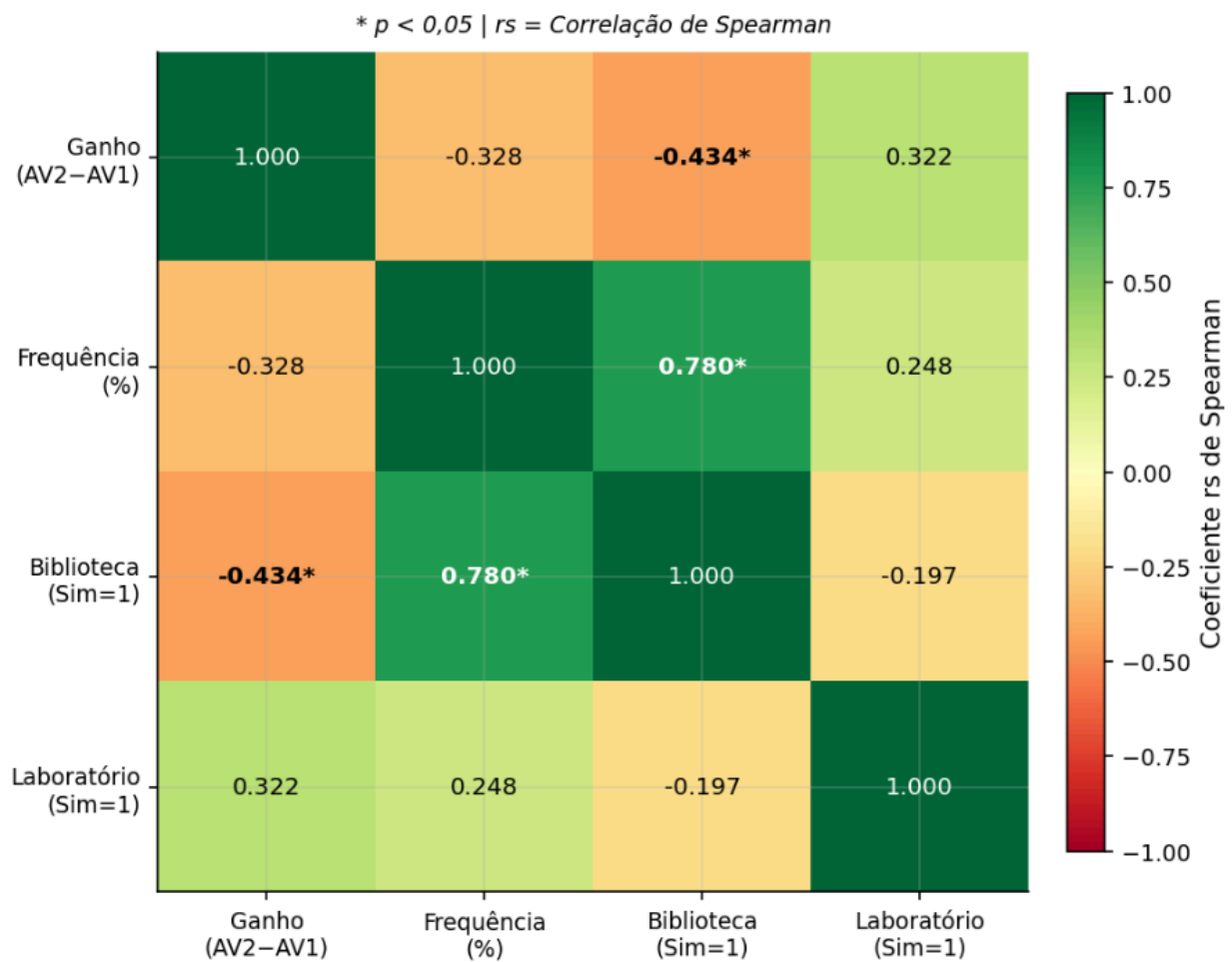
Tabela 5. Matriz de correlações de *Spearman* entre variáveis extraclasse e ganho ($n = 32$)

Variável	Ganho (AV2-AV1)	Frequência (%)	Biblioteca (0/1)	Laboratório (0/1)
Ganho (AV2-AV1)	1,000	-0,328	-0,434*	+0,322
Frequência (%)	-0,328	1,000	+0,780*	+0,248
Biblioteca (Sim=1)	-0,434*	+0,780*	1,000	-0,197
Laboratório (Sim=1)	+0,322	+0,248	-0,197	1,000

*** $p < 0,05$.**

Fonte: Elaborado pelos autores com suporte do *software* estatístico Jamovi, 2026.

Figura 4. Heatmap da matriz de correlações de Spearman (variáveis extraclasse e ganho de desempenho). Cores quentes (vermelho) = correlação positiva; frias (verde) = correlação negativa.



Fonte: Elaborado pelos autores com suporte da linguagem Python, 2026.

O modelo de regressão linear múltipla com Ganho como variável dependente foi estatisticamente significativo: $R^2 = 0,307$; R^2 ajustado = 0,232; $F(3,28) = 4,125$; $p = 0,015$ explicando 30,7% da variância do ganho de desempenho (Tabela 6).

Tabela 6. Regressão linear múltipla: Ganho ~ Frequência + Biblioteca + Laboratório (n = 32)

Preditor	β	Erro-Padrão	Estatística t	p-valor	VIF
Constante	8,462	5,562	t = 1,52	p = 0,139	—
Frequência (%)	-0,093	0,075	t = -1,24	p = 0,225	8,43

Biblioteca (Sim=1)	-0,206	1,068	t = -0,19	p = 0,849	4,85
-----------------------	--------	-------	-----------	-----------	------

⚠ Esta tabela possui muitas colunas e foi cortada para impressão. Para visualizá-la completa, acesse o artigo original em: <https://revistatopicos.com.br/artigos/sala-de-aula-invertida-no-ensino-de-bioestatistica-na-graduacao-em-medicina-estudo-quase-experimental-de-desempenho-academico-e-variaveis-de-engajamento?noblockage>

$R^2=0,307$; $R^2_{aj}=0,232$; $F(3,28)=4,125$; $p=0,015^*$ $p < 0,05$. VIF > 5 indica multicolinearidade preocupante. **Fonte:** Elaborado pelos autores com suporte do *software* estatístico Jamovi, 2026.

A análise de multicolinearidade revela VIF = 8,43 para Frequência e VIF = 4,85 para Biblioteca, valores elevados que refletem a alta correlação entre essas variáveis ($r_s = 0,780$). Nesse cenário, os betas individuais perdem estabilidade e não devem ser interpretados isoladamente; as estimativas confiáveis são o modelo global ($F = 4,125$; $p = 0,015$; $R^2 = 0,307$) e o coeficiente do Laboratório (VIF = 3,07; $p = 0,098$), que indica tendência positiva robusta à multicolinearidade. Em amostras maiores ($n \geq 80$), técnicas como ridge regression ou análise de componentes principais (Hair et al., 2019) seriam indicadas para contornar a colinearidade.

Para explorar efeitos de moderação, foram ajustados modelos com termos de interação com variáveis centradas na média (para redução de multicolinearidade não-essencial). Os resultados são apresentados na Tabela 7 e visualizados na Figura 5.

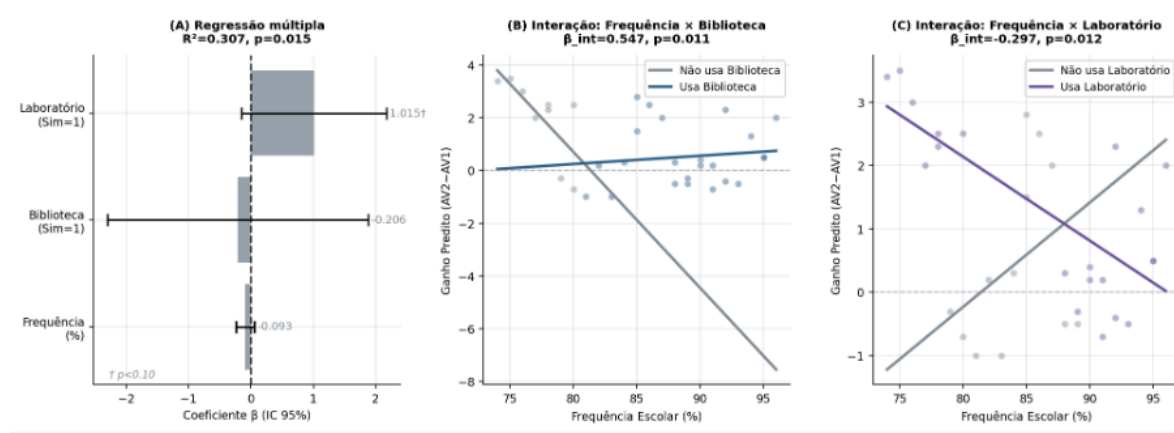
Tabela 7. Modelos com termos de interação: efeitos de moderação pela frequência ($n = 32$)

Modelo de Interação	β (interação)	p-valor	R ² do Modelo	Interpretação pedagógica
Frequência × Biblioteca	$\beta = +0,547$	$p = 0,011^*$	0,394	Bib. mais benéfica com alta frequência
Frequência × Laboratório	$\beta = -0,297$	$p = 0,012^*$	0,449	Lab. mais benéfico com baixa frequência

Variáveis centradas na média. * $p < 0,05$.

Fonte: Elaborado pelos autores com suporte do *software* estatístico Jamovi, 2026.

Figura 5. Modelos de Regressão: Preditores do ganho de desempenho. (A) Coeficientes β com IC 95% do modelo de regressão múltipla. (B) Interação Frequência × Biblioteca: trajetórias previstas por nível de uso. (C) Interação Frequência × Laboratório. Pontos = dados observados; linhas = trajetórias previstas.



Fonte: Elaborado pelos autores com suporte da linguagem Python, 2026.

A interação Frequência × Biblioteca ($\beta = +0,547$; $p = 0,011$; $R^2 = 0,394$) indica que o benefício marginal da biblioteca é crescente com a frequência: estudantes de alta frequência que não usaram a biblioteca tiveram perdas relativas, enquanto os de baixa frequência

que a utilizaram apresentaram ganhos ainda mais modestos do que os que não a utilizaram reforçando o papel compensatório de outros recursos para esse perfil.

A interação Frequência × Laboratório ($\beta = -0,297$; $p = 0,012$; $R^2 = 0,449$) revela que o laboratório é especialmente benéfico para estudantes de menor frequência, quanto menor a presença nas aulas, maior o benefício relativo do laboratório. Em contrapartida, estudantes de alta frequência que usaram o laboratório tenderam a apresentar ganho menor, reflexo do efeito de teto sobre o desempenho inicial. Esses resultados, com incremento do R^2 para 0,39–0,45 nos modelos com interação, evidenciam que relações não-lineares e de moderação são centrais para compreender como o engajamento extraclasse opera na SAI, dado que análises univariadas e lineares simples não capturam adequadamente.

4.7. Análise de Cluster: Perfis de Engajamento e Desfechos

A análise K-means foi realizada com padronização Z-score das variáveis (Frequência, Biblioteca, Laboratório). O coeficiente de silhueta indicou $k = 3$ como solução ótima (silhueta = 0,753), com equilíbrio entre separação estatística dos grupos e interpretabilidade pedagógica. A Tabela 8 descreve os perfis resultantes.

Tabela 8. Perfis dos clusters K-means de engajamento extraclasse ($n = 32$)

Cluster	n	Freq. Média	% Bib.	% Lab.	Ganh. Médi
Cluster A	10	85,0%	100%	0%	0,63

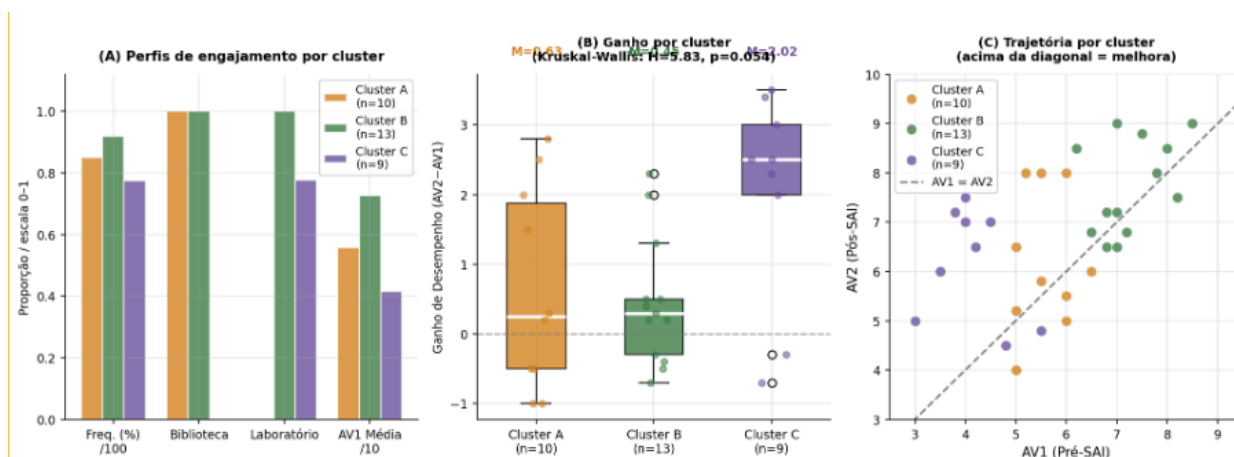
Cluster B	13	92,0%	100%	100%	0,44
Cluster C	9	77,4%	0%	77,8%	2,02

⚠ Esta tabela possui muitas colunas e foi cortada para impressão. Para visualizá-la completa, acesse o artigo original em: <https://revistatopicos.com.br/artigos/sala-de-aula-invertida-no-ensino-de-bioestatistica-na-graduacao-em-medicina-estudo-quase-experimental-de-desempenho-academico-e-variaveis-de-engajamento?noblockage>

Kruskal-Wallis entre clusters: $H=5,83$; $p=0,054$. Comparação par a par Cluster B vs. C: Mann-Whitney $p=0,025^*$.

Fonte: Elaborado pelos autores com suporte do software estatístico Jamovi, 2026.

Figura 6. Análise de Clusters. (A) Perfis de engajamento por cluster (barras agrupadas). (B) Boxplots do ganho por cluster; Kruskal-Wallis $H=5,83$; $p=0,054$; comparação Cluster B vs. C: $p=0,025^*$. (C) Dispersão $AV1 \times AV2$ colorida por cluster; pontos acima da diagonal indicam melhora.



Fonte: Elaborado pelos autores com suporte da linguagem Python, 2026.

O teste de *Kruskal-Wallis* indicou tendência de diferença entre *clusters* ($H = 5,83$; $p = 0,054$), e a comparação par a par revelou

diferença significativa entre Cluster B e Cluster C (*Mann-Whitney* U = 24,5; $p = 0,025$). O Cluster C ("Recuperador via Laboratório") caracterizado por menor frequência (M = 77,4%), ausência de uso da biblioteca e uso predominante do laboratório, registrou o maior ganho médio (M = 2,022), em contraposição ao Cluster B ("Multiengajado"), que apresentou o menor ganho (M = 0,446) apesar de reunir estudantes com maior frequência, maior uso de recursos e, conseqüentemente, o maior desempenho inicial (AVI médio = 7,27). O Cluster A ("Engajado Tradicional"), com frequência intermediária e uso exclusivo da biblioteca, situou-se em posição intermediária (M = 0,630).

O resultado do Cluster C constitui o achado de maior profundidade pedagógica do estudo. Estudantes com o pior perfil de engajamento convencional foram os que mais se beneficiaram da SAI, fenômeno que se explica por três mecanismos articulados: (1) margem de crescimento estrutural: com AVI médio de 4,14, havia maior distância para o máximo e, portanto, maior espaço de ganho absoluto; (2) aprendizagem autodirigida de recuperação: a SAI, ao externalizar o contato inicial com o conteúdo para o espaço extraclasse, pode ter funcionado como mecanismo especialmente eficaz para quem frequentava menos as aulas expositivas, pois esses estudantes compensaram a menor presença com estudo autônomo prévio; (3) papel compensatório do laboratório: o espaço prático de aplicação com retroalimentação imediata substituiu funcionalmente a biblioteca como recurso de consolidação do conhecimento.

Esse padrão é coerente com os resultados de Maryami et al. (2025), que demonstraram, em ensaio randomizado por conglomerados, que estudantes com menor desempenho inicial se beneficiam

desproporcionalmente da SAI, com maior retenção do conhecimento a longo prazo; e com Kiljunen et al. (2024), que identificaram, em universitários com dificuldades de aprendizagem, estratégias alternativas de engajamento igualmente eficazes para a aprendizagem.

A implicação prática imediata desse achado é relevante: em vez de uniformizar as expectativas de engajamento, a gestão pedagógica da SAI deve reconhecer e acolher perfis alternativos de estudo particularmente o uso do laboratório como recurso central, criando condições para que estudantes em situação de risco pedagógico possam se beneficiar do modelo sem que a menor frequência se torne um impeditivo ao desempenho. Nas novas DCNs de Medicina (CNE/CES nº 3/2025), que reafirmam a literacia estatística como pilar da formação e a adoção de metodologias ativas como diretriz estrutural, a SAI com suporte multimodal de recursos deixa de ser opção e passa a ser exigência formativa (Oliveira et al., 2026; Barros et al., 2026).

4.8. Síntese dos Achados e Limitações

A Tabela 9 sintetiza os principais resultados estatísticos de todos os blocos de análise.

Tabela 9. Síntese dos achados estatísticos por bloco de análise (n = 32)

Bloco de Análise	Resultado Central	Estatística Principal	Interpretação
AV1 vs. AV2	Ganho médio = 0,95; d = 0,668	t(31)=3,78; p<0,001	Efeito médio a grande; sig. confirmada por Wilcoxon

Frequência × Ganho	$r=-0,411$; $\beta=-0,090$	Pearson $p=0,019$	Efeito de composição, não causal; reflexo da proficiência inicial
Biblioteca × Ganho	Não-usuários: $M=2,02$ vs. $M=0,53$	Mann-Whitney $p=0,017^*$	Efeito de teto diferenciado; OR logístico n.s. (magnitude $\neq$ direção)
Laboratório × Ganho	Usuários: $M=1,25$ vs. $M=0,44$	Mann-Whitney $p=0,076$	OR=4,00 ($p=0,085$); recurso compensatório, tendência relevante
Plataforma Digital	Adesão 100%	Sem variabilidade	Sucesso operacional da SAI; análise comparativa inviável
Regressão Múltipla	$R^2=0,307$; VIF(Freq)=8,43	$F(3,28)=4,125$; $p=0,015^*$	Modelo global sig.; multicolinearidade Freq-Bib. compromete betas
Interações	Freq×Bib $p=0,011$; Freq×Lab $p=0,012$	Ambas $p<0,05^*$	$R^2=0,39-0,45$; relações de moderação não-lineares confirmadas
Clusters K-means	Cluster C: Ganho $M=2,02$	B vs. C: $p=0,025^*$	Recuperação via laboratório, perfil de maior benefício da SAI

* $p < 0,05$.

Fonte: Elaborado pelos autores, 2026.

As conclusões deste estudo são limitadas por: (1) ausência de grupo controle, variáveis de confusão (maturação, efeito de prática) não são controladas; (2) amostra por conveniência ($n = 32$), poder estatístico reduzido, especialmente para subgrupos pequenos (Biblioteca Não: $n=9$; Frequência Baixa: $n=1$); (3) multicolinearidade elevada ($VIF > 5$),

impede interpretação isolada de coeficientes da regressão múltipla; (4) plataforma digital sem variabilidade, logs de acesso não disponibilizados; (5) ausência de dados de autorregulação e motivação, impede triangulação dos mecanismos identificados; (6) violação da normalidade das diferenças, mitigada pela análise dupla, porém requer cautela na interpretação paramétrica. Estudos futuros com delineamentos controlados, amostras multicêntricas ($n \geq 80$), ANCOVA com AVI como covariável, variáveis contínuas de uso do AVA e instrumentos validados de autorregulação são necessários para consolidar e ampliar as evidências aqui apresentadas.

5. CONCLUSÃO/CONSIDERAÇÕES FINAIS

O estudo avaliou os efeitos da Sala de Aula Invertida sobre o desempenho em Bioestatística de 32 estudantes de Medicina, em delineamento quase-experimental de grupo único pré-teste/pós-teste, e de variáveis de engajamento extraclasse. O achado confirmatório é robusto, a SAI associou-se à ganho médio de 0,95 ponto (IC 95%: [0,44; 1,46]), significativo [$t(31) = 3,780$; $p < 0,001$; Wilcoxon $W = 121,5$; $p = 0,008$] e de efeito médio a grande ($d = 0,668$), com queda dos estudantes com nota $< 5,0$ de 8 (25,0%) para 3 (9,4%).

As análises extraclasse, de caráter exploratório, apontaram associações explicáveis, em grande parte, por um mecanismo comum, o acoplamento do escore de mudança ao desempenho basal e a regressão à média. Assim, a correlação negativa da frequência e o maior ganho dos não-usuários da biblioteca refletem a menor proficiência inicial desses subgrupos, não efeito causal. O laboratório mostrou tendência positiva (OR = 4,00), moderada pela frequência, a plataforma digital teve adesão universal (100%), sinal de sucesso operacional.

A análise de cluster identificou três perfis, o "Recuperador via Laboratório" (Cluster C) menor frequência e uso de laboratório teve o maior ganho ($M = 2,02$), sugerindo benefício a estudantes em risco que adotam estratégias alternativas. Esse perfil, porém, coincide com os não-usuários da biblioteca e com o menor desempenho inicial achado exploratório, a interpretar à luz da regressão à média. Os modelos de interação ($R^2 = 0,39-0,45$) sugeriram efeito não-linear e moderado do engajamento, com implicações para intervenções personalizadas.

As limitações, ausência de grupo controle, amostra por conveniência, multicolinearidade elevada e escores de mudança acoplados ao basal impõem cautela na generalização e recomendam, como reanálise prioritária, o ajuste por AVI (ANCOVA). Investigações controladas, com amostras maiores, variáveis contínuas de uso do AVA e instrumentos de autorregulação, poderão consolidá-las. Nas Diretrizes Curriculares Nacionais de 2025, que reafirmam a literacia estatística como pilar da formação médica, a adoção sistemática de metodologias ativas no ensino de Bioestatística firma-se como direção formativa estratégica.

REFERÊNCIAS BIBLIOGRÁFICAS

ABEYSEKERA, L.; DAWSON, P. Motivation and cognitive load in the flipped classroom: definition, rationale and a call for research. **Higher Education Research & Development**, v. 34, n. 1, p. 1-14, 2015. DOI: 10.1080/07294360.2014.934336.

ALTHUBAITI, A.; ALTHUBAITI, S. M. Flipping the Online Classroom to Teach Statistical Data Analysis Software: A Quasi-Experimental Study. **SAGE Open**, v. 14, n. 1, 2024. DOI: 10.1177/21582440241235022.

ANDERSON, Lorin W.; KRATHWOHL, David R. **A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition.** Addison Wesley Longman, Inc., 2001.

BARROS, L. et al. Bioestatística e medicina baseada em evidências: desafios pedagógicos e a necessidade de reforma curricular frente às diretrizes curriculares nacionais 2025. **Caderno Pedagógico**, v. 23, n. 2, p. 01-24, 2026. DOI: 10.54033/cadpedv23n2-036.

BERGMANN, Jon; BERGMANN, Jonathan; SAMS, Aaron. **Flip your classroom, revised edition: Reach every student in every class every day.** ASCD, 2023.

BRASIL. Lei nº 14.874, de 28 de maio de 2024. Dispõe sobre a pesquisa com seres humanos. **Diário Oficial da União**, Brasília, DF, 29 maio 2024.

BRASIL. Ministério da Educação. Conselho Nacional de Educação. **Diretrizes Curriculares Nacionais dos Cursos de Graduação em Medicina.** Brasília, DF: MEC/CNE, 2001; 2014; 2025.

CAMPBELL, D. T.; STANLEY, J. C. **Delineamentos experimentais e quase-experimentais de pesquisa.** São Paulo: EPU, 2015.

COHEN, Jacob. **Statistical power analysis for the behavioral sciences, Rev.** Lawrence Erlbaum Associates, Inc, 1977.

COHEN, Jacob. **Statistical power analysis for the behavioral sciences.** routledge, 2013.

DENG, Ruiqi; FENG, Siqu; SHEN, Suqin. Improving the effectiveness of video-based flipped classrooms with question-embedding. **Education and Information Technologies**, v. 29, n. 10, p. 12677-12702, 2024. <https://doi.org/10.1007/s10639-023-12303-5>

EVARISTO-CHIYONG, T.; MATTOS-VELA, M. Flipped learning for teaching biostatistics to Peruvian dental students. **Journal of Oral Research**, v. 8, n. 2, p. 159-165, 2019. DOI: 10.17126/joralres.2019.025.

FIELD, A. **Discovering statistics using IBM SPSS statistics**. [S.l.]: Sage Publications Limited, 2024.

FREEMAN, S. et al. Active learning increases student performance in science, engineering, and mathematics. **Proceedings of the National Academy of Sciences**, v. 111, n. 23, p. 8410–8415, 2014. DOI: 10.1073/pnas.1319030111.

GOPALAN, C. et al. Exploring educational transformations through the Innovative Flipped Learning Instruction Project Symposium. **Advances in Physiology Education**, v. 48, n. 3, p. 505-511, 2024. DOI: 10.1152/advan.00027.2024.

GRAWE, Paul H. Nonparametric Statistics for the Behavioral Sciences. **Numeracy: Advancing Education in Quantitative Literacy**, v. 9, n. 1, 2016. DOI. 10.5038/1936-4660.9.1.7

HAIR, J. F. et al. **Multivariate data analysis**. 8. ed. Hampshire: Cengage Learning, 2019.

HAZIKI, N. H.; ABDULLAH, A. H.; HAMZAH, M. H. Flipped Classrooms in Higher Education: A Meta-Analysis of Their Impact on

Mathematics Performance. **Malaysian Journal of Mathematical Sciences**, v. 19, n. 1, p. 125–146, 2025.

ILGAZ, A. The Impact of Flipped Classroom and Kahoot! on Students' Motivation in Biostatistics Education: An Action Research Study. **Journal of Education and Research in Nursing**, v. 22, n. 1, p. 8–18, 2025.

JANNUZZI, P. M. Delineamentos experimentais na avaliação de políticas públicas: usos e abusos. **Estudos em Avaliação Educacional**, v. 34, e09956, 2023. DOI: 10.18222/eae.v34.9956.

KHADAWARDI, K. et al. Evaluating Medical Student Engagement in Flipped Classrooms: Insights on Motivation and Peer Learning. **Journal of Medical Education and Curricular Development**, v. 12, 2025. DOI: 10.1177/23821205251320756.

KILJUNEN, J. et al. Higher education and the flipped classroom approach: efficacy for students with a history of learning disabilities. **Higher Education**, v. 88, p. 1127–1143, 2024. DOI: 10.1007/s10734-023-01162-1.

LI, S. et al. Effectiveness of Flipped Classrooms for K-12 Students: Evidence From a Three-Level Meta-Analysis. **Review of Educational Research**, 2025. DOI: 10.3102/00346543241261732.

LO, C. K. et al. The effectiveness of the flipped classroom: A second-order meta-analysis. **Teaching and Teacher Education**, 2025. DOI: 10.1016/j.tate.2025.104889.

MARYAMI, M.; HOSSEINIMOGHADAM, M.; ROSTAMI, K. The Impact of Flipped Classroom Approach on Critical Thinking, Self-Esteem, and

Learning Retention of Nursing Undergraduate Students: A Cluster Randomized Study. **Health Science Reports**, v. 8, 2025. DOI: 10.1002/hsr2.71271.

MCLAUGHLIN, Jacqueline E. et al. The flipped classroom: a course redesign to foster learning and engagement in a health professions school. **Academic medicine**, v. 89, n. 2, p. 236-243, 2014.

MENEZES, A. A. da S.; PETRACONI, N.; MARTINS, C. B. Planejamento e análise da qualidade de uma avaliação de estatística descritiva em cursos da área da saúde. **Revista de Educação Matemática e Tecnológica Iberoamericana – EM TEIA**, v. 17, n. 1, p. 75-98, 2026. DOI: <https://doi.org/10.51359/2177-9309.2026.268500>

NAING, C. et al. The effects of flipped classrooms to improve learning outcomes in undergraduate health professional education: A systematic review. **Campbell Systematic Reviews**, v. 19, n. 3, p. e1339, 2023.

OLANA, T.; BACHA, K.; LEMMA, A. The impact of flipped classroom strategy on secondary school students' achievement in biology. **Discover Education**, v. 4, article 494, 2025. DOI: 10.1007/s44217-025-00942-4.

OLIVEIRA, S. S. de et al. Novas Diretrizes Curriculares Nacionais do curso de graduação em Medicina de 2025: um convite para a transformação da educação médica no Brasil. **Revista Brasileira de Educação Médica**, v. 50, n. 1, p. e050, 2026. <https://doi.org/10.1590/1981-5271v50.1-editorial>

SHI, X. Y.; YIN, Q.; WANG, Q. W.; LU, B. R.; LI, G. X.; HUANG, S. H.; SUN, Z. G. Is the flipped classroom more effective than the traditional

classroom in clinical medical education: a systematic review and meta-analysis. **Frontiers in Education**, v. 9, 2025. DOI: 10.3389/feduc.2024.1485540.

SILVA, R. G. de P. et al. Competência em medicina baseada em evidências de uma universidade pública: influência do período de graduação. **Revista Brasileira de Educação Médica**, v. 49, n. 4, p. e173, 2025.

SPAIC, Dragan et al. The flipped classroom in medical education: systematic review and meta-analysis. **Journal of Medical Internet Research**, v. 27, p. e60757, 2025.

WANG, Z.; LU, J.; YU, S. The impact of the flipped classroom on the motivation and academic performance of Chinese college English learners. **PLoS ONE**, v. 20, n. 5, 2025. DOI: 10.1371/journal.pone.0322094.

¹ Docente do Programa de Mestrado Profissional em Ensino em Saúde do Centro Universitário Doutor Leão Sampaio Campus Lagoa Seca. E-mail: [acesse o artigo original para visualizar o e-mail](#).

² Discente do Curso Superior de Direito do Centro Universitário Doutor Leão Sampaio. Campus Lagoa Seca. E-mail: [acesse o artigo original para visualizar o e-mail](#).

³ Discente do Curso Superior de Direito do Centro Universitário Doutor Leão Sampaio. Campus Lagoa Seca. E-mail: [acesse o artigo original para visualizar o e-mail](#).

⁴ Discente do Programa de Mestrado Profissional em Ensino em Saúde do Centro Universitário Doutor Leão Sampaio Campus Lagoa Seca. E-mail: [acesse o artigo original para visualizar o e-mail](#).

⁵ Discente do Curso Superior de Medicina do Centro Universitário Doutor Leão Sampaio. Campus Lagoa Seca. E-mail: [acesse o artigo original para visualizar o e-mail](#).