

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**Instituto de Ciências Exatas**  
**Programa de Pós-Graduação em Física**

Tais Bastani Ribeiro

**IMPACTO DE METODOLOGIAS DE ENSINO E GÊNERO NA APRENDIZAGEM:**  
**uso de metodologias ativas em Fundamentos de Eletromagnetismo**

Belo Horizonte  
2023

Tais Bastani Ribeiro

**IMPACTO DE METODOLOGIAS DE ENSINO E GÊNERO NA APRENDIZAGEM:  
uso de metodologias ativas em Fundamentos de Eletromagnetismo**

Dissertação apresentada ao Programa de Pós-Graduação em Física do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Mestre em Física.

Orientador: Ubirajara Agero Batista

Belo Horizonte

2023

Dados Internacionais de Catalogação na Publicação (CIP)

R484i Ribeiro, Tais Bastani.  
Impacto de metodologias de ensino e gênero na aprendizagem: uso de metodologias ativas em Fundamentos de Eletromagnetismo / Tais Bastani Ribeiro. – 2023.  
52f. : il.

Orientador: Ubirajara Agero Batista.  
Dissertação (mestrado) – Universidade Federal de Minas Gerais,  
Departamento de Física.  
Bibliografia: f. 46-50.

1. Física, Ensino. 2. Métodos de ensino. 3. Aprendizagem I. Título. II. Batista, Ubirajara Agero. III. Universidade Federal de Minas Gerais, Departamento de Física.

CDU – 53(07) (043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

### **FOLHA DE APROVAÇÃO**

A presente dissertação, intitulada “**Impacto de metodologias de ensino e gênero na aprendizagem: uso de metodologias ativas em Fundamentos de Eletromagnetismo**”, de autoria de **TAIS BASTANI RIBEIRO**, submetida à Comissão Examinadora, abaixo-assinada, foi aprovada para obtenção do grau de **MESTRE EM FÍSICA** em vinte e nove de junho de 2023.

Belo Horizonte, 29 de junho de 2023.

Prof. Ubirajara Agero Batista

Orientador da aluna

Departamento de Física/UFMG

Prof. Elmo Salomão Alves

Departamento de Física/UFMG

Prof. Luis Gustavo D'Carlos Barbosa

Faculdade de Educação/UFMG

Dra. Juliana Caldeira Brant

Departamento de Física/UFMG



Documento assinado eletronicamente por **Ubirajara Agero Batista, Servidor(a)**, em 30/06/2023, às 12:52, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Juliana Caldeira Brant, Usuário Externo**, em 30/06/2023, às 15:06, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Elmo Salomão Alves, Coordenador(a)**, em 01/07/2023, às 10:05, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Luis Gustavo D Carlos Barbosa, Professor do Magistério Superior**, em 04/07/2023, às 12:50, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **2429870** e o código CRC **1315047B**.

## AGRADECIMENTOS

Agradeço ao meu orientador Bira por apoiar e acreditar neste trabalho.

Aos professores e Júlia e Lucas pela colaboração imprescindível para este trabalho.

Aos professores da FACE e FaE, pela ajuda na construção de um trabalho interdisciplinar.

Ao André, pelo companheirismo, suporte e amor.

Aos amigos da Física, que me deram muito palpite e café. Em especial João Rafael por cada leitura e Olímpio e Sabrina por cada almoço longo.

Aos meus pais e irmãs, por terem me dado todo o apoio e suporte em toda jornada da Física.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

## RESUMO

Metodologias ativas de ensino são alternativas a aulas tradicionais, aplicadas com o propósito de melhorar o aprendizado. Pesquisadores estudam o impacto de metodologias ativas no aprendizado conceitual, na retenção de estudantes e no *gap* de gênero, fenômeno em que mulheres consistentemente têm notas menores que homens em avaliações de física. Dentre as metodologias ativas mais utilizadas no ensino de Física está *Peer Instruction* (PI), que consiste em atividades preparatórias pré aula e discussões em grupo durante a aula.

O PI é aplicado em uma disciplina de Física básica da Universidade Federal de Minas Gerais, e este uso foi adaptado no período de Ensino Remoto Emergencial (ERE), durante a pandemia de COVID-19. Neste trabalho, comparamos o aprendizado conceitual e nota final de estudantes dessa disciplina em metodologia tradicional, PI e ERE. Também comparamos aprendizado conceitual e nota final por gênero em cada metodologia. Para isso, aplicamos o teste *Brief Electricity and Magnetism Assessment* (BEMA) no primeiro e último dia de aula da disciplina, no modelo pré e pós teste, para um total de 1515 estudantes, sendo 353 de aulas tradicionais, 245 de PI e 917 do ERE. Calculamos então, para cada metodologia e gênero, o ganho normalizado médio  $g$  e o  $d$  de Cohen do BEMA, a distribuição do ganho e da nota final e as comparamos com o método de distribuição relativa.

Na análise de dados, vemos que o aumento relativo no ganho médio dos estudantes com o uso de PI é de 32%. No entanto, não vemos este aumento na nota final. Para o ERE, há um aumento expressivo na nota final comparado com as outras duas metodologias, mas o ganho é similar ao tradicional. Ao separarmos o ganho do ERE por semestre, vemos uma queda a cada período, sendo muito menor no último semestre. Separando por gênero, vemos que há um *gap* de gênero em todas as metodologias de ensino, e encontramos indícios de que a metodologia ativa pode ter um *gap* de gênero menor, mas ainda considerável. Em nenhuma das metodologias há diferença da nota final entre gêneros.

No todo, concluímos que o uso de PI melhora o aprendizado conceitual de estudantes, e que o ensino durante o ERE foi menos efetivo ao longo dos semestres. A nota final em uma disciplina não se mostrou um bom parâmetro de comparação de aprendizagem entre grupos diferentes. Por fim, vemos que o uso de PI pode ajudar a reduzir o *gap* de gênero, mas não de forma expressiva.

**Palavras-chave:** *Peer Instruction*, *gap* de gênero, aprendizagem conceitual, ensino remoto emergencial

## ABSTRACT

Active methodologies are alternatives to traditional classes used with the purpose of enhance students' learning. Researches study the impact of active methodologies on conceptual learning, retention of students and gender gap, phenomenon in which women consistently score lower than men in Physics assessments. Among the most used active methodologies in Physics teaching is Peer Instruction (PI), which consists of preparatory activities before the class and group discussions during class.

PI is used in a introductory Physics course of Universidade Federal de Minas Gerais, and this use was adapted during the period of Emergency Remote Teaching (ERT), during the COVID-19 pandemic. In this work, we compare the conceptual learning and the final grade of students in traditional classes, PI classes and ERT. We also compare the conceptual learning and course scores by gender in each methodology. For this, we applied the Brief Electricity and Magnetism Assessment (BEMA) in the first and last class of the course, with a pretest and posttest design, for 353 students of traditional classes, 245 students of PI clases and 917 of ERT. From the pretest and posttest, we calculated, for each methodology and gender, the normalized gain  $g$ , Cohen's  $d$  from BEMA, the distribution for the final grade and  $g$  and compared these distributions with the relative distribution method.

On the data analysis, we see a relative increase on the mean gain of 32% with the use of PI, however we don't see this increase on the final grade. On the ERT, there is a significant increase on the final grade compared to the other two methodologies, but  $g$  is similar to  $g$  in traditional classes. When separating  $g$  from ERT by semester, we see a decrease in each semester, being much lower in the last semester. When separating these data by gender, we see a gender gap in all methodologies, and we found evidence that PI might have a smaller gender gap, but it's still significant. There isn't a gender gap in final grades in neither of the methodologies.

Overall, we conclude that the use of PI improves students' conceptual learning, and that teaching during the ERT is less effective over the semesters. The final score in a course is not a good parameter for comparing learning between different groups. Finally, we see that the use of active methodologies may help decreasing gender gap, but not by a large amount.

**Keywords:** Peer Instruction, gender gap, conceptual learning, emergency remote teaching.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>9</b>
<b>1.1</b>	<b><i>Peer Instruction</i></b>	<b>10</b>
1.1.1	<i>Peer Instruction</i> e outras metodologias	12
<b>1.2</b>	<b>Ensino Remoto Emergencial</b>	<b>13</b>
<b>1.3</b>	<b><i>Gap</i> de gênero</b>	<b>15</b>
<b>1.4</b>	<b>Nota final</b>	<b>15</b>
<b>2</b>	<b>MÉTODOS</b>	<b>17</b>
<b>2.1</b>	<b>Fundamentos de Eletromagnetismo</b>	<b>17</b>
<b>2.2</b>	<b>Nota final</b>	<b>17</b>
<b>2.3</b>	<b><i>Brief electricity and magnetism assessment</i></b>	<b>18</b>
<b>2.4</b>	<b>Ganho normalizado</b>	<b>18</b>
<b>2.5</b>	<b>D de cohen</b>	<b>19</b>
<b>2.6</b>	<b>Estimativa de densidade Kernel</b>	<b>20</b>
<b>2.7</b>	<b>Distribuição relativa</b>	<b>21</b>
<b>3</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>22</b>
<b>3.1</b>	<b><i>Peer Instruction</i> vs tradicional</b>	<b>22</b>
<b>3.2</b>	<b>Ensino Remoto Emergencial</b>	<b>26</b>
3.2.1	ERE vs métodos presenciais	26
3.2.2	Por semestre	31
<b>3.3</b>	<b><i>Gap</i> de gênero</b>	<b>34</b>
<b>4</b>	<b>CONCLUSÕES</b>	<b>44</b>
	<b>REFERENCIAS</b>	<b>46</b>
	<b>APÊNDICE</b>	<b>51</b>

## 1 INTRODUÇÃO

O ensino de Física é usualmente caracterizado pelo uso de metodologias tradicionais, que consiste tipicamente pela exposição do conteúdo pelo professor, com pouca interação dos estudantes. Contudo, esse tipo de aula não contribui muito para o desenvolvimento de conceitos de Física próximos do consenso científico, desenvolvendo sobretudo habilidades de resolução de problemas quantitativos [1]. Então, professores e pesquisadores de ensino de Física voltaram-se para aulas interativas a fim de aumentar o aprendizado conceitual. As metodologias ativas são caracterizadas pela participação dos estudantes no processo de aprendizagem a partir de atividades que promovem a reflexão sobre o conteúdo estudado.

A pesquisa em ensino mostra um resultado positivo de metodologias ativas quando se mede o aprendizado dos estudantes [2]. Há também outros diversos benefícios, como o engajamento dos estudantes, criação de um ambiente de cooperação e melhor memorização do conteúdo [3]. Em universidades brasileiras, o uso de metodologias ativas também promove um aumento no ganho conceitual [4, 5].

Outra questão investigada é se metodologias ativas afetam o chamado *gap* de gênero—o fenômeno em que, nas avaliações de Física, as mulheres consistentemente têm notas médias menores que os homens. Não há um consenso sobre a influência de métodos ativos sobre a diferença entre homens e mulheres, e há estudos que mostram tanto aumento quanto diminuição do *gap* [6].

No Departamento de Física (DF) da Universidade Federal de Minas Gerais (UFMG), há a predominância de aulas tradicionais, mas metodologias ativas têm sido usadas por alguns professores em disciplinas introdutórias de Física, em particular Fundamentos de Eletromagnetismo. A metodologia ativa mais usada é o *Peer Instruction* (PI), em conjunto com *Just in Time Teaching* (JiTT). Entre 2017 e 2019, foram aplicados testes conceituais para estudantes dessa disciplina, afim de avaliar o aprendizado e comparar as duas metodologias.

Os dados de Fundamentos de Eletromagnetismo foram analisados por Bastani Ribeiro [7], encontrando que metodologias ativas promovem um aumento no aprendizado conceitual, mas não reduz o *gap* de gênero. No entanto, uma das limitações desse estudo é que há apenas dois professores que aplicam metodologia ativa, e um dos *gaps* da literatura sobre o uso do *Peer Instruction* são pesquisas sobre a influência dos professores e suas

crenças sobre o método. Afim de contribuir com essa discussão, tivemos a intenção de, neste trabalho, observar o aprendizado quando professores de aulas tradicionais usam o *Peer Instruction*. Todavia, antes que isso pudesse ser feito, deu-se início a pandemia de COVID-19 e o regime de Ensino Remoto Emergencial, que perdurou por quatro semestres letivos.

Diante do contexto apresentado, procuramos usar ferramentas estatísticas mais avançadas para comparar o aprendizado conceitual e a nota final em aulas tradicionais e aulas que combinam o uso de *Peer Instruction* com *Just in Time Teaching*. Além disso, também pretendemos investigar se essa combinação afeta o *gap* de gênero. Sob os mesmos critérios, pretendemos avaliar o impacto do ensino remoto emergencial - em particular, adaptado a partir do PI - sobre o aprendizado de estudantes.

A seguir, introduziremos as metodologias de ensino usadas e os métodos empregados para a análise.

## 1.1 *Peer Instruction*

Uma das metodologias ativas mais usadas na Física, especialmente em contexto universitário, é o ***Peer Instruction*** (PI), desenvolvido na Universidade de Harvard pelo físico Eric Mazur, motivado pelo resultado inferior ao esperado da aplicação de testes conceituais em suas turmas. A metodologia tem como objetivo melhorar a aprendizagem conceitual dos estudantes, além de “explorar a interação entre os estudantes durante as aulas expositivas e focar a atenção dos estudantes nos conceitos que servem de fundamento” [8]. A dinâmica do *Peer Instruction* segue como mostrado no fluxograma da figura 1 e descrito abaixo:

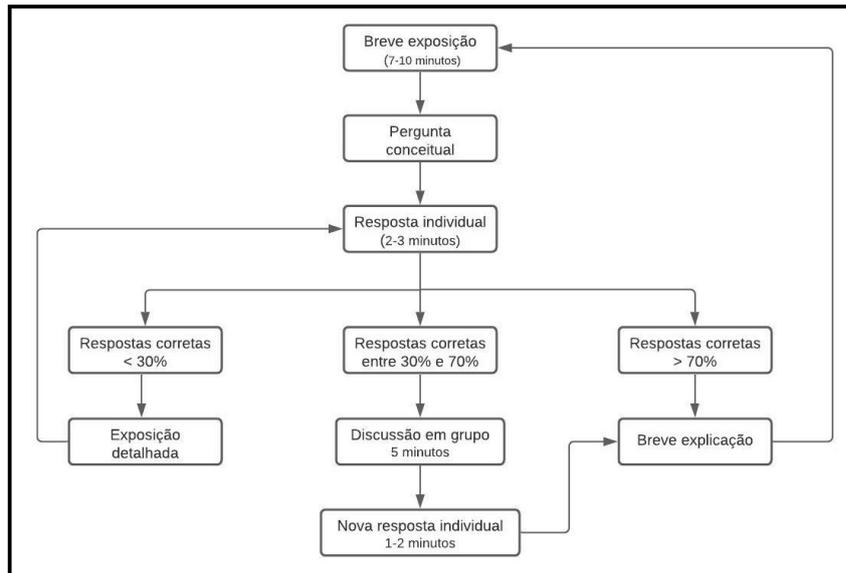
- 1º passo: exposição breve do conteúdo estudado (7-10 minutos);
- 2º passo: estudantes respondem individualmente uma pergunta conceitual curta - chamada ConcepTest - em um sistema de resposta - *clickers*, *flashcards*, dispositivos com acesso à internet<sup>1</sup>(2-3 minutos). Então, há 3 caminhos a serem seguidos:
  - Se 70% ou mais dos estudantes responderem a pergunta corretamente, há uma breve explicação sobre a questão e a aula segue desde o 1º passo;
  - Se 30% ou menos dos estudantes responderem a pergunta corretamente, há uma nova exposição sobre o assunto, mais detalhada, e a aula segue desde o 1º passo;

---

<sup>1</sup> Destacamos o aplicativo Meliora - meliora.ufmg.br, desenvolvido na UFMG.

- Se entre 30% e 70% dos estudantes responderem a pergunta corretamente, estudantes discutem em pares ou pequenos grupos sobre a questão (5 minutos), e respondem a pergunta conceitual novamente.

Figura 1 – Fluxograma ilustrando aula típica com a metodologia *Peer Instruction*.



Indicamos os tempos médios estimados e ressaltamos que os tempos variam de acordo com o assunto e a dinâmica do professor na aula. Destacamos também algumas orientações para fazer boas perguntas conceituais [9]:

- focar em um único conceito central;
- não ser ambígua;
- não ser fácil demais ou difícil demais;
- fornecer respostas incorretas plausíveis, baseadas em equívocos comuns de estudantes;
- não se resumir à aplicação de números em equações.

Desde 1990, na Universidade de Harvard, foram aplicados em disciplinas de Física básica testes conceituais no modelo pré e pós teste. Em 1991, o PI passou a ser usado nessas disciplinas, substituindo as aulas tradicionais. Desde então, essa metodologia começou a ser utilizada em diversas universidades dos Estados Unidos, principalmente em Física mas também em matemática, biologia, química e outros cursos da área STEM (Science, Technology, Engineering and Mathematics) [10], com variações aplicadas pelos professores dependendo do contexto em sala de aula [11]. No Brasil, o *Peer Instruction* também

passou a ser aplicado e estudado, tanto em contexto escolar [12,13] como em contexto universitário [5].

O aspecto mais analisado com o uso do *Peer Instruction* é a aprendizagem conceitual de estudantes. Segundo Crouch e Mazur (2001), o ganho aumentou expressivamente com o uso de PI, mas sem comprometer o desempenho na resolução de problemas quantitativos clássicos, no qual também se viu aumento [9]. Com o uso do PI em outras universidades, corroborou-se o aumento da aprendizagem conceitual, na habilidade de resolução de problemas e no desempenho acadêmico [14].

### 1.1.1 *Peer Instruction* e outras metodologias

Com a adoção crescente do *Peer Instruction*, também popularizou-se o uso dessa metodologia com outros métodos de ensino ativo. Uma das metodologias adotada em conjunto com o PI é o *Tutorials in Introductory Physics* [15] - Tutoriais, em tradução livre. Os Tutoriais são listas de exercícios desenvolvidas para resolução em pequenos grupos, em horários extra após a aula regular. Os estudantes são incentivados a discutir e chegar em consensos quanto à solução dos problemas apresentados e, os professores ou monitores, a não dar respostas diretas às dúvidas de estudantes. Um uso dessa metodologia é descrito com detalhes por Finkelstein e Pollock (2005) [16].

Dentre as combinações mais comuns, está o uso de PI com *Just in Time Teaching* (JiT) [17], ou Ensino sob Medida, em tradução utilizada em estudos nacionais [18]. que tem como objetivo incentivar a preparação prévia dos estudantes e, assim, aumentar seu aprendizado. Na proposta de Novak, estudantes estudam um material indicado antes da aula (capítulo de livro-texto, notas de aula, etc) e, após o estudo, respondem um teste conceitual curto, com o propósito de identificar os principais erros e dúvidas. A aula é preparada com base nestes resultados. Lopes (2014) estudou em sua dissertação sete períodos acadêmicos de uso de PI com JiTT, encontrando que o uso dessa combinação leva à um melhor desempenho de estudantes, comparado com o uso de apenas uma das metodologias.

O *Peer Instruction* prevê, tradicionalmente, que estudantes leiam o conteúdo previamente, mas o *Just in Time Teaching* detalha esse estudo prévio e acrescenta o *feedback* de estudantes. Então, convencionou-se adotar as duas metodologias juntas. No departamento de Física, os professores que adotam metodologias ativas usam o PI com JiTT, pedindo que estudantes respondam duas perguntas sobre o conteúdo lido antes da aula. Em uma terceira pergunta, os professores perguntam se há dúvidas ou qual a parte mais interessou o estudante. No entanto, por simplicidade, apresentamos os resultados apenas como *Peer Instruction*.

## 1.2 Ensino Remoto Emergencial

Em março de 2020, a Organização Mundial de Saúde (OMS) declarou a pandemia de COVID-19. No mundo todo, atividades educacionais, em todos os níveis, foram forçadas a migrar para o formato remoto, instaurando o Ensino Remoto Emergencial (ERE), nomenclatura adotada por diversas instituições. O nome faz-se necessário para diferenciar o ERE da Educação à Distância (EaD). Segundo Coqueiro e Sousa (2021), a EaD envolve atividades em um ambiente virtual de aprendizagem (AVA), realizadas com planejamento prévio e metodologias próprias. Além disso, no Brasil, a EaD tem uma legislação própria que a regula. Já no ERE, devido ao seu caráter emergencial, não há planejamento prévio, e as atividades podem ser síncronas ou assíncronas, a depender da instituição ou professor [19].

Nos últimos dois anos, os efeitos da pandemia na educação têm sido estudados de diversos pontos de vista. Para White, apesar de as aulas disponibilizadas de forma online e assíncrona terem sido assistidas múltiplas vezes, o desempenho de estudantes na prova referente a essas aulas foi inferior ao desempenho após aulas presenciais [20]. Acrescentando ao debate sobre atividades síncronas ou assíncronas, Guo afirma que estudantes que participaram de atividades síncronas optativas tiveram uma transição para o remoto mais tranquila, com uma queda menor nas notas e um ganho conceitual maior do que estudantes que participaram apenas de atividades assíncronas obrigatórias. [21].

Outro aspecto estudado é a influência de metodologias ativas sobre o ERE. Em um formato adaptado de *Peer Instruction* para ensino remoto, de Paula et al. (2021), observaram que as aulas síncronas, nas quais estudantes respondiam perguntas em grupos referentes a conteúdos assistidos previamente, promoveram participação e compreensão dos tópicos. Neste caso, esse era o primeiro semestre em que foi usada alguma metodologia ativa - a transição para o PI estava prevista antes da pandemia [22]. Já no estudo de Aguiar et al. (2022), as aulas presenciais antes da pandemia já eram ministradas com metodologias ativas, com aulas intercalando exposições curtas e atividades em grupos, como exercícios e experimentos. Esse formato foi adaptado para o ERE. Nos dois regimes, estudantes de graduação e mestrado profissional responderam um teste conceitual de termodinâmica. Então, os autores compararam as médias do teste entre os dois regimes, não encontrando diferença estatística no desempenho de estudantes no formato presencial e no formato remoto. No entanto, é ressaltado que a pesquisa mede apenas um aspecto das aulas e que o estudo não implica que o ensino remoto não traz prejuízos para a formação acadêmica [23]. É importante perceber que os estudos que, de alguma forma, medem o desempenho de estudantes durante o ensino remoto emergencial trazem dados de apenas um semestre, em geral o primeiro do formato remoto. Portanto, é incerto o efeito do ERE no desempenho de estudantes a longo prazo. No Brasil, a UFMG, assim como outras instituições de ensino superior, manteve as atividades remotas durante 2020 e 2021, com o retorno presencial sendo realizado apenas no primeiro semestre de 2022 [24].

Para a adaptação da disciplina de Fundamentos de Eletromagnetismo, os estudantes foram separados em turmas de 40-50 pessoas, como no período presencial, e as aulas foram padronizadas em modelo inspirado no uso de metodologias ativas, como descrito: Para cada tópico de um capítulo, havia disponível para os estudantes, na plataforma de aprendizagem virtual *Moodle*<sup>2</sup>:

- gravações de aulas de diferentes professores e instituições no *Youtube*;
- uma apostila criada pelos professores da disciplina, desenvolvida antes da pandemia para estudos independentes - estudantes também poderiam seguir qualquer livro-texto do nível de Física 3 a escolha;
- um teste de conhecimento com questões conceituais e quantitativas que deveria ser respondido após os estudos do tópico. O teste provia um *feedback*: ao responder errado, o estudante tinha uma dica indicando o raciocínio para a resposta correta. Os testes podiam ser feitos inúmeras vezes.

Além dos materiais de cada tópico, também eram disponibilizadas listas de exercícios, que cobriam múltiplos capítulos.

O horário tradicional de aula - duas sessões síncronas por semana, de até 1:40h cada - era reservado para os estudantes tirarem dúvidas sobre os estudos dos tópicos, que deviam ser estudados previamente.

Os estudantes foram avaliados principalmente através de provas tradicionais ou orais, ambas online, 2 ou 3 por semestre. As provas tradicionais eram disponibilizadas no *Moodle*, com 3-4 questões quantitativas, para serem resolvidas em 1:40h. Nas provas orais, os estudantes deveriam explicar, em sessão aberta, a solução de uma questão sorteada das listas de exercícios. Se os estudantes não soubessem a solução, uma nova questão poderia ser sorteada. Quem preferisse, podia solicitar a realização da prova oral em sessão privada, somente com o professor, mas a maioria o fez em sessão aberta.

Os testes de conhecimento valiam, no total, 10-15% da nota final. Os estudantes que responderam o teste conceitual BEMA (descrito com detalhes na seção 2.3) recebiam 3% da nota final, independente do resultado do BEMA. Também era pedido que os estudantes respondessem uma autoavaliação, em que tinham que classificar em níveis os seus conhecimentos sobre conceitos centrais de eletromagnetismo. O restante da nota era dividido entre as provas e trabalhos em grupos, nos quais os estudantes deveriam produzir um conteúdo (vídeo, apresentação, podcast, página web, etc) sobre determinado tópico.

A distribuição da nota entre as provas e trabalhos mudou ao longo dos semestres. Na tabela 1, vemos essa distribuição detalhada, com o valor total de cada atividade. Cada semestre tem o total de 100 pontos.

---

<sup>2</sup> moodle.org

Tabela 1 – Distribuição de pontos em cada semestre do Ensino Remoto Emergencial

Semestre	2020/1	2020/2	2021/1	2021/2
Autoavaliação e BEMA	5	5	5	5
Testes de conhecimento	15	10	10	15
Produção de conteúdo	40	40	25	20
Prova tradicional	20	15	30	-
Prova oral	20	30	30	60

Na análise de dados, consideramos o ERE como uma metodologia de ensino diferente da tradicional e PI, por ter sido padronizado e aplicado em um momento único, distinto dos semestres anteriores.

### 1.3 Gap de gênero

Um resultado comum na pesquisa em ensino de Física é que mulheres apresentam menores notas em testes conceituais e outras formas de avaliação. Esse fenômeno é chamado de **gap de gênero**. Sob o pressuposto de que ambientes cooperativos e discussão entre estudantes é um passo a ser tomado em direção à fechar o *gap* de gênero [25], Lorenzo, Crouch e Mazur investigaram se o PI tinha alguma influência sobre o *gap*. Os resultados mostraram que tanto homens quanto mulheres obtiveram notas menores em aulas tradicionais do que em aulas com PI e com o uso conjunto de PI e Tutoriais. O *gap* de gênero é visto tanto no pré teste quanto no pós teste de aulas tradicionais, mas em aulas interativas a diferença de nota no pós-teste entre homens e mulheres diminuiu, chegando a tornar-se estatisticamente insignificante entre estudantes de PI com Tutoriais. [26] Usando a mesma combinação de metodologias ativas, Pollock et al. também avaliou seu efeito no *gap* de gênero. No entanto, o resultado encontrado foi diferente: em turmas com PI, o *gap* de gênero manteve-se, indicando que metodologias ativas não são suficientes para reduzir o *gap* de gênero e outros fatores, como contexto educacional e os próprios professores, podem influenciar na diferença [27]. Essa conclusão vai ao encontro de outros estudos que mostram que o *gap* de gênero é multifatorial [6].

### 1.4 Nota final

A nota final em uma disciplina tem um papel central na educação formal, sendo uma forma de quantificar o conhecimento de estudantes em certo assunto e determinar se são considerados aptos a prosseguir para outras disciplinas e atividades que requerem aquele conhecimento. Há a preocupação se metodologias ativas contribuem para, além do aprendizado conceitual, uma melhora no desempenho medido pela nota final na disciplina. Estudos relatam que, ao comparar testes conceituais com notas de exames das disciplinas,

viu-se que apesar de tanto os resultados dos testes conceituais como as notas finais dos exames serem instrumentos capazes de medir ganho de aprendizagem gerados pelo uso de metodologias ativas, as notas dos exames são instrumentos menos sensíveis [28].

Como a nota final de um estudante é um dos aspectos que define sua retenção, os aspectos que influenciam a nota final também podem influenciar na retenção. A retenção é um grande problema em disciplinas introdutórias ofertadas pelo DF. Segundo relatório da Pró-Reitoria de Graduação (Prograd) da UFMG [29], entre 2011 e 2018, apenas 52% dos estudantes de Fundamentos de Eletromagnetismo foi aprovado. 7,8% abandonou a disciplina, sendo reprovados por frequência, 10,3% trancou a disciplina e 29,9% foi reprovado por nota. Em 2020, primeiro ano de ERE, essa aprovação saltou para 81,8% e, em 2021, foi para 65,4%. Considerando esta situação, consideramos que estudar o impacto da metodologia de ensino na nota final é um passo para entender a retenção em física básica.

## 2 MÉTODOS

Para comparar as metodologias de ensino apresentadas, fazemos uma análise estatística a partir das ferramentas descritas neste capítulo.

### 2.1 Fundamentos de Eletromagnetismo

A disciplina Fundamentos de Eletromagnetismo (Fund. Eletro) ofertada pelo Departamento de Física da UFMG é a disciplina introdutória de Eletromagnetismo, equivalente à Física III, onde se cobre os conteúdos: carga elétrica, campo elétrico e a lei de Gauss; potencial elétrico, capacitores e dielétricos; corrente e resistência elétricas; campo magnético e lei de Ampère; lei de Faraday e indutância; equações de Maxwell. Essa disciplina está presente nos currículos dos cursos de graduação em Física, Matemática, Matemática Computacional, Química, Química Tecnológica, Ciência da Computação e Engenharias.

Em cada semestre, há 9-11 turmas de Fund. Eletro, cada uma com 40-60 estudantes, distribuídas entre 6 a 7 professores, a depender da demanda. A carga horária é de 60h, com duas aulas semanais de 1:40h. Apenas cursa a disciplina estudantes aprovados em Fundamentos de Mecânica - disciplina introdutória de mecânica, equivalente à Física I - e Calculo II - disciplina que cobre sequência, séries, coordenadas polares, derivadas parciais [30].

### 2.2 Nota final

Em Fundamentos de Eletromagnetismo, no período de 2017 a 2019, os professores, tanto tradicionais quanto que usam *Peer Instruction*, têm liberdade para escolher suas avaliações e a distribuição da nota final. Cada estudante deve ser avaliado ao final da disciplina com uma nota de 0 a 100, e cada atividade, como prova ou trabalho em grupo, deve ser avaliada em no máximo 40 pontos. As formas mais comuns de distribuição da nota final, em ambos os métodos, são: três provas valendo 1/3 da nota final cada; três provas valendo 30 pontos cada, e 10 pontos de atividades como listas de exercícios ou trabalhos. No entanto, nem todos os professores seguem este modelo. Cada professor é responsável por desenvolver e avaliar as atividades de seus estudantes. Para um estudante

ser aprovado na disciplina, precisa ter uma nota final  $> 60$ . O estudante que tiver nota final entre 40 e 60, têm direito a realizar um exame especial, uma prova que cobre todo o conteúdo do curso. Caso o estudante tenha nota igual ou maior a 60 no exame especial, sua nota final na disciplina é 60. Estudantes que tiverem nota final  $< 40$  ou nota  $< 60$  no exame especial são considerados reprovados [31]. A distribuição de nota final durante o ERE foi padronizada, conforme descrito em 1.2. Então temos, para cada estudante, uma nota final entre 0 e 100.

### 2.3 *Brief electricity and magnetism assessment*

Na pesquisa em ensino de Física, os pesquisadores tipicamente aplicam os chamados testes conceituais - testes padronizados, desenvolvidos com base em pesquisa, com questões conceituais ou semiquantitativas. Como exemplos de testes conceituais, temos o *Force Concept Inventory* (FCI) [32] e o *Force and Motion Conceptual Evaluation* (FMCE) [33] que cobrem conteúdos de mecânica newtoniana; o *Brief Electricity and Magnetism Assessment* (BEMA) [34] e o *Conceptual Survey of Electricity and Magnetism* (CSEM) [35], que cobrem conteúdos de eletromagnetismo. Há também testes sobre outros conteúdos de Física, como termodinâmica e quântica, e conceitos mais específicos, como densidade, energia, processamento de dados. Uma lista mais detalhada dos testes pode ser encontrada, em inglês, na plataforma PhysPort - <https://www.physport.org/Assessment.cfm>.

Em Fundamentos de Eletromagnetismo, usamos o BEMA, um teste com 31 questões de múltipla escolha, conceituais e semi-quantitativas, planejadas para medir o conhecimento de conceitos básicos de eletromagnetismo. BEMA é um teste confiável e com um poder discriminatório adequado, portanto a nota do estudante no teste é uma medida confiável do seu conhecimento no assunto [34]. A tradução do BEMA para português está disponível em [physport.org](https://www.physport.org). Duas questões do BEMA, em português, são mostradas na Fig.2.3: Aplicamos o BEMA no primeiro dia de aula e no final da disciplina, no modelo pré e pós teste, que nos permite o contraste do desempenho estudantes com o curso, e não apenas o seu desempenho final. Assim temos, com esse modelo, uma forma de quantificar o aprendizado conceitual dos estudantes.

### 2.4 Ganho normalizado

Uma maneira de medir o aprendizado dos estudantes a partir do pré e pós teste é calcular o ganho de hake ou **ganho normalizado** ( $g$ ), usado amplamente na pesquisa em ensino de Física. Essa medida foi popularizada por Richard Hake em um estudo comparando aulas tradicionais e de metodologias ativas [36]. Considerando a quantidade de questões que o estudante não acertou no pré-teste como medida do quanto ele ainda

Figura 2 – Duas primeiras questões do BEMA.

Dois objetos pequenos, cada um com uma carga líquida  $+Q$ , fazem uma força de magnitude  $F$  entre si:



Substitui-se um dos objetos por outro cuja carga líquida é  $+4Q$ :



1. A magnitude original da força sobre a carga  $+Q$  era  $F$ . Qual é a magnitude da força sobre a carga  $+Q$  agora?

- |            |                              |
|------------|------------------------------|
| (a) $4F$   | (e) $F$                      |
| (b) $5F/2$ | (f) $F/4$                    |
| (c) $3F$   | (g) Diferente das anteriores |
| (d) $2F$   |                              |

2. Qual é a magnitude da força sobre a carga  $+4Q$ ?

- |            |                              |
|------------|------------------------------|
| (a) $4F$   | (e) $F$                      |
| (b) $5F/2$ | (f) $F/4$                    |
| (c) $3F$   | (g) Diferente das anteriores |
| (d) $2F$   |                              |

Fonte: Tradução em português do BEMA obtido na plataforma PhysPort, em [www.physport.org](http://www.physport.org). Acesso em fev/2019.

poderia aprender com o curso, o ganho normalizado é o percentual desta quantidade que ele acerta no pós-teste. Assim, ele é definido como:

$$g = \frac{N_{pos} - N_{pre}}{N_{max} - N_{pre}} \quad (2.1)$$

onde onde  $N_{pre}$  é a nota do pré teste,  $N_{post}$  é a nota do pós teste e  $N_{max}$  é a nota máxima no BEMA.

Há duas formas de calcular o  $g$ . No seu trabalho, Hake apresentou o ganho das médias, calculando  $g$  usando a média do  $N_{pre}$  e  $N_{pos}$ . Também é possível calcular  $g$  pela média de ganhos, computando  $g$  para cada estudante e então tirando a média. Os resultados dessas duas formas de calcular  $g$  são similares para grupos maiores que 50 estudantes [37]. Neste estudo, o  $g$  é calculado pela média de ganhos. Quanto mais próximo de 1 for  $g$ , maior a aprendizagem conceitual do estudante durante o curso. Se o pós-teste for menor do que o pré-teste, o  $g$  é negativo, então consideramos que o estudante teve uma perda conceitual durante o curso.

## 2.5 D de cohen

**D de cohen** ( $d$ ) é uma medida de tamanho de efeito amplamente usada em ciências sociais, e mede o tamanho da diferença entre a média de duas variáveis, normalizado pelo desvio padrão. É definido como:

$$d = \frac{m_1 - m_2}{s_{pool}} \quad (2.2)$$

onde  $m_1$  e  $m_2$  são as médias das duas variáveis a serem comparadas e  $S_{pool}$  é o desvio padrão agrupado.

$$s_{pool} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \quad (2.3)$$

onde  $n_1$  e  $s_1$  são, respectivamente, o tamanho da amostra e desvio padrão de  $m_1$  e  $n_2$  e  $s_2$  são o tamanho da amostra e desvio padrão para  $m_2$ .

Há quatro formas diferentes de calcular  $d$ , segundo Goulet-Pelletier et al. [38]. Na pesquisa em ensino de Física, a forma mais comum é comparar a nota do pré e pós teste para um grupo, e comparar o resultado entre grupos. Assim,  $m_1$ ,  $n_1$  e  $s_1$  são, respectivamente, a média, número de respondentes e desvio padrão do pós teste e  $m_2$ ,  $n_2$  e  $s_2$  são a média, número de respondentes e desvio padrão do pré teste. Então, quando usado nesse cenário,  $d$  reporta se o curso foi efetivo e quantifica esse efeito. Cohen originalmente definiu um  $d$  de 0,2 como pequeno, 0,5 como médio e 0,8 como grande [39]. No entanto, essas definições não devem ser seguidas rigidamente e é importante considerar o contexto do estudo [38].

O uso do ganho normalizado  $g$  e o  $D$  de cohen  $d$  para medir e comparar aprendizado têm sido discutido na pesquisa em ensino de física. Enquanto  $g$  é uma medida mais popular nesse campo, seu uso têm sido questionado. Nissen et al. argumenta que  $g$  é enviesado pelo pré teste, gerando resultados enganosos sobre intervenções ou comparações entre grupos [40]. No entanto, segundo Coletta e Steinert,  $g$  não é enviesado pelo pré teste, e  $d$  é menos informativo [41]. Neste trabalho, optamos por focar a análise no  $g$ , usando essa medida como variável dependente na análise estatística, mas também reportamos e avaliamos  $d$ .

## 2.6 Estimativa de densidade Kernel

A Estimativa de densidade Kernel (EDK) é uma forma não-paramétrica de estimar a densidade de probabilidade de uma variável aleatória [42]. Ela é definida por

$$\hat{f}(x_0) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x_i - x_0}{h}\right) \quad (2.4)$$

onde  $x_0$  é o ponto no qual avaliamos a estimativa de densidade,  $x_i$  são as observações de  $x$  na amostra,  $K$  é uma função kernel a escolha (triangular, quadrática, gaussiana,

Epanechnikov, etc) e  $h$  é um parâmetro de suavização chamado *bandwidth*. A função é estimada da seguinte forma:

- constrói-se uma curva kernel centrada em cada ponto, com escala determinada por  $h$ ;
- soma-se essas curvas, obtendo a curva de densidade de probabilidade.

Utilizamos a EDK para estimar a densidade de probabilidade do ganho normalizado, adotando  $h = 1$  e, para  $K$ , adotamos a gaussiana. Junto à curva de densidade de probabilidade feita por EDK, mostramos um histograma, onde dividimos o ganho do grupo em 10 faixas iguais e calculamos a porcentagem de estudantes em cada faixa de ganho. Então, apresentamos o gráfico de distribuição para o ganho normalizado e nota final, onde temos o histograma e, sobreposto ao histograma, a curva de densidade de probabilidade.

## 2.7 Distribuição relativa

A densidade relativa é uma ferramenta estatística descritiva e não paramétrica, desenvolvida por Handcock e Morris usada para comparar duas distribuições de uma mesma variável para grupos diferentes. [43]. Para fazer o gráfico de densidade relativa discreto, seguimos como descrito:

- separamos o grupo de referência em decis;
- vemos a faixa de ganho de cada decil;
- calculamos a porcentagem de estudantes do grupo de comparação em cada faixa de ganho.

No eixo x do gráfico, optamos por mostrar a faixa de ganho de cada decil. A primeira e a última faixa do gráfico representam, respectivamente, a porcentagem de estudantes do grupo de comparação com ganhos menores que o menor ganho do grupo de referência e a porcentagem de estudantes do grupo de comparação com ganhos maiores que o maior ganho do grupo de referência. Assim, temos a proporção de estudantes do grupo de comparação em cada decil do grupo de referência. Quando o resultado para um decil é  $> 10\%$ , temos uma sobre-representação de estudantes do grupo de comparação. Quando o resultado é  $< 10\%$ , temos uma sub-representação de estudantes do grupo de comparação. Quando o resultado é  $= 10\%$ , temos uma distribuição proporcional entre os dois grupos. Então, usamos o método da distribuição relativa para comparar a distribuição do ganho normalizado entre diferentes grupos.

### 3 RESULTADOS E DISCUSSÃO

Para comparar os métodos de ensino usados em aulas presenciais e o Ensino Remoto Emergencial, coletamos a nota final, o pré e pós teste para 353 estudantes de 15 turmas de 7 professores que utilizam o método tradicional, 245 estudantes de 10 turmas de 2 professores que utilizam o *Peer Instruction* e 917 estudantes de 36 turmas de 9 professores no período do Ensino Remoto Emergencial. Os dados dos métodos tradicional e *Peer Instruction* foram colhidos entre 2017/1 e 2019/2 e os dados do Ensino Remoto Emergencial foram colhidos entre 2020/1 e 2020/2, segundo o calendário adaptado. Na tabela 2, apresentamos a média da nota final, pré teste e pós teste. Os dados apresentados aqui são de estudantes que responderam ambos pré e pós teste.

Primeiramente, utilizamos os resultados do pré e pós teste para calcular o  $g$  e o  $d$ , e, a partir dos resultados dessas medidas também apresentados na tabela 2, utilizamos o  $g$  e a nota final como variáveis resposta e o método de ensino e gênero como variáveis explicativas nas ferramentas estatísticas descritas no capítulo 2. Avaliamos a significância estatística da diferença entre as médias de  $g$  e nota final com o teste t. Por completude, apresentamos o p-valor do teste, e escolhemos adotar o critério padrão de significância de 5%. Inicialmente, comparamos apenas os dois métodos presenciais entre si. Em seguida, adicionamos o Ensino Remoto Emergencial, comparando o resultado do período com os dois métodos presenciais e posteriormente analisamos apenas o ERE, usando o semestre como variável resposta. Por fim, adicionamos a variável gênero, investigando o  $gap$  de gênero em cada método.

Uma versão completa da tabela, com os resultados do pré e pós teste, é mostrada no apêndice 4

#### 3.1 *Peer Instruction* vs tradicional

A tabela 2 mostra que o ganho para estudantes do método tradicional é 0,216, enquanto para estudantes do *Peer Instruction* é 0,283. A diferença de 0,068 representa um aumento de aproximadamente 32% no ganho dos estudantes de PI em relação ao método tradicional. O teste-t dessa diferença apresenta o p-valor  $< 0,001$ , então descartamos a hipótese de as médias de  $g$  serem iguais. O aumento no ganho com o uso de uma metodologia ativa é consideravelmente menor que o aumento encontrado por Hake, com o

Tabela 2 – Média da nota final e ganho normalizado para o métodos tradicional, *Peer Instruction* e Ensino Remoto Emergencial separados por gênero masculino (Masc.) e feminino (Fem.). Número de respondentes (N), média (m) e desvio padrão ( $\sigma$ ) da nota final e ganho normalizado ( $g$ ) e D de Cohen ( $d$ ) de estudantes que completaram o pré e o pós teste.

Método	Gênero	N	Nota final		$g$		$d$
			m	$\sigma$	m	$\sigma$	
Tradicional	Masc.	237	70,0	19,6	0,239	0,242	0,932
	Fem.	116	68,1	20,3	0,166	0,177	0,864
	Ambos	353	69,4	19,8	0,215	0,225	0,881
<i>Peer Instruction</i>	Masc.	141	70,2	20,2	0,318	0,229	1,248
	Fem.	104	69,0	20,7	0,237	0,226	1,123
	Ambos	245	69,7	20,4	0,283	0,231	1,172
Ensino Remoto Emergencial	Masc.	612	82,2	13,9	0,233	0,279	0,866
	Fem.	301	82,8	13,8	0,150	0,229	0,782
	Ambos	917	82,4	13,9	0,205	0,224	0,812

$g$  sendo considerado baixo para ambos os métodos [36]. No entanto, ao compararmos com estudos nacionais, vemos um aumento relativo similar [44].

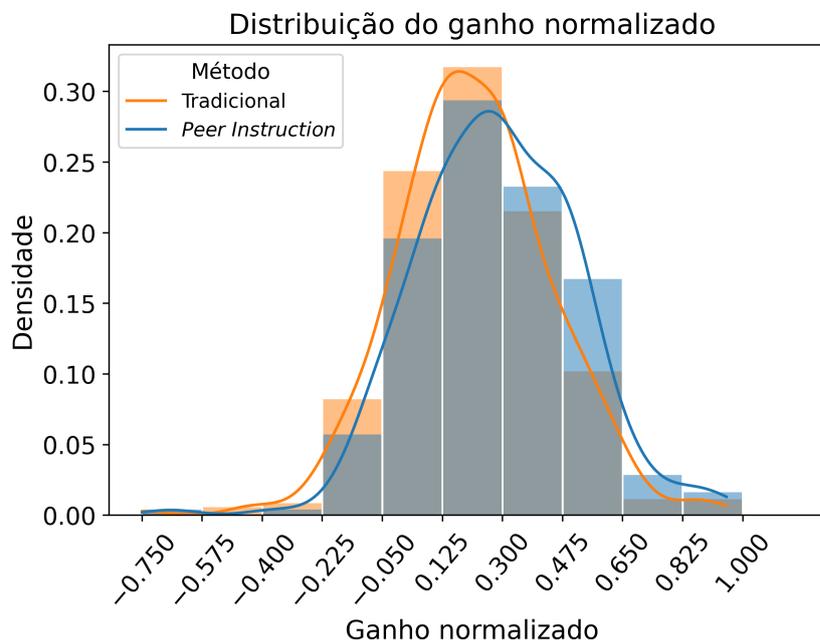
A diferença no efeito do método sobre os testes também pode ser vista pelo valor de  $d$ , na tabela 2, onde PI apresenta  $d = 1,172$  e tradicional apresenta  $d = 0,881$ . Vemos um valor de  $d$  que é considerado grande em ambos os casos, um resultado esperado, pois indica que o curso, independente da metodologia, influencia no resultado do pós teste em relação ao pré teste. Contudo, o  $d$  maior para PI indica que essa metodologia apresenta uma influencia maior no resultado do pós teste.

Para explorar a distribuição do ganho normalizado, fizemos um histograma que divide o ganho em 10 faixas de 0,175 e mostra a porcentagem de estudantes em cada faixa de ganho. Sobreposto ao histograma, plotamos uma estimativa de densidade kernel do ganho normalizado. O resultado é visto na Fig.3.

Na Fig.3, a curva e barras laranjas representam a metodologia tradicional e a curva e barras azuis representam o *Peer Instruction*. Em cinza, vemos a sobreposição das barras. No gráfico, vemos que há um deslocamento nas distribuições: a curva que mostra o  $g$  dos estudantes submetidos ao PI está deslocada para a direita comparado com a curva que mostra o  $g$  dos estudantes do método tradicional de ensino. O resultado mostrado indica que, a qualquer valor de  $g$ , estudantes do PI têm probabilidades maiores de ter ganhos maiores que estudantes de aulas tradicionais. Essa característica na distribuição é similar à encontrada em outros estudos [2].

Para explorar a diferença entre as distribuições, fazemos um gráfico de distribuição relativa discreta, usando estudantes do método tradicional como grupo de referência e estudantes do *Peer Instruction* como grupo de comparação. Então, separamos estudantes

Figura 3 – Histograma e EDK do ganho normalizado para os métodos tradicional e *Peer Instruction*. O ganho normalizado é separado em dez faixas de 0,175.

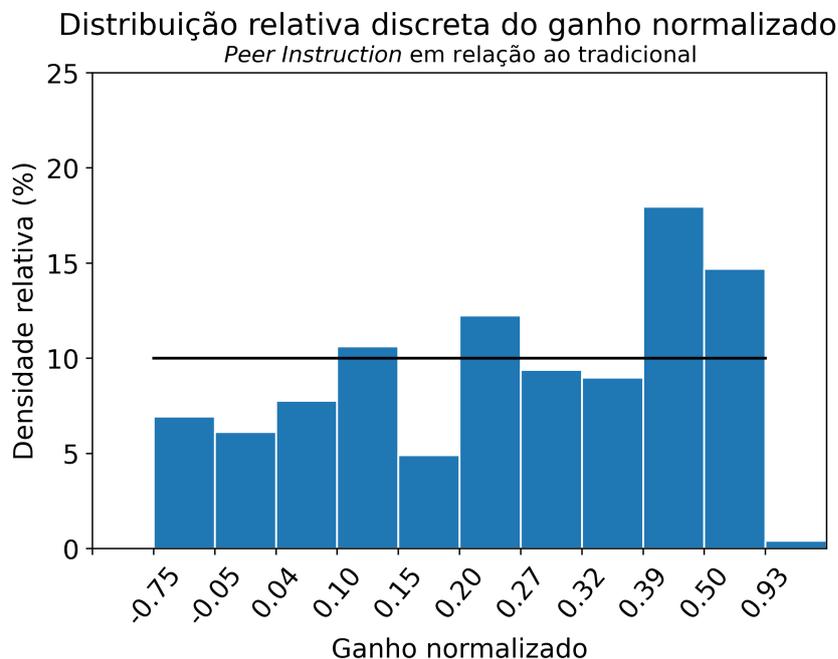


do método tradicional em grupos definidos pelos quantis de ganho a cada 10%, computamos a faixa de ganho do quantil e calculamos a porcentagem de estudantes do PI nessa faixa de ganho.

No gráfico mostrado na Fig.4, o primeiro decil do grupo de referência representa 10% dos estudantes com menor ganho normalizado no método tradicional, com  $g$  entre  $-0,75$  e  $-0,05$ , conforme mostrado no eixo x do gráfico. Para essa faixa de ganho, há aproximadamente 7% dos estudantes do PI. O último decil do grupo de referência representa 10% dos estudantes com maior ganho normalizado no método tradicional, com  $g$  entre  $0,50$  e  $0,93$ . Para essa faixa de ganho, nós temos aproximadamente 15% de estudantes do PI. A última barra representa os estudantes do PI com ganhos maiores que  $0,93$ . Não há estudantes do PI com ganhos menores que  $-0,75$ , que seriam representados pela primeira barra, nesse caso vazia. Observando também os outros decis dos estudantes do método tradicional, vemos que há uma tendência de sub-representação de estudantes do PI em faixas de ganho mais baixos, e uma sobrerrepresentação de estudantes do PI em faixas de ganho mais altas. Então, estudantes de PI estão concentrados em ganhos mais altos em relação aos estudantes do método tradicional.

Agora, usamos os mesmos métodos para analisar a diferença da nota final entre as duas metodologias de ensino. A média da nota é apresentada na tabela2, mostrando uma diferença de 0,3 na nota final média entre os métodos. O teste-t dessa diferença apresenta um p-valor= 0.847, então falhamos em descartar a hipótese de que as notas finais sejam

Figura 4 – Distribuição relativa discreta do ganho normalizado, com estudantes do método tradicional como grupo de referência e estudantes do *Peer Instruction* como grupo de comparação.



iguais. A distribuição da nota final nos dois métodos é mostrada na Fig.5.

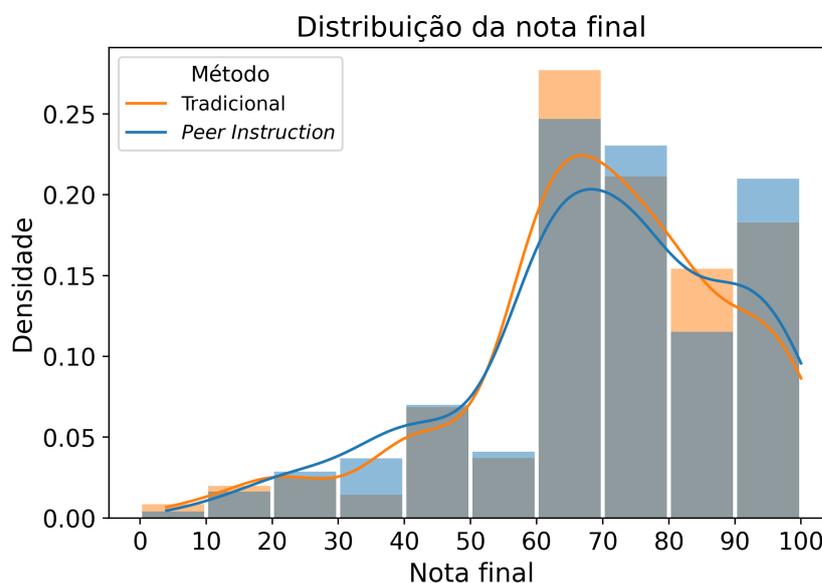
Nas distribuições da Fig.5, vemos que também não há diferença significativa na distribuição da nota final entre os métodos.

Na Fig.6, vemos que, ao contrário do ganho normalizado, na nota final não há uma tendência crescente da distribuição relativa, e sim um comportamento aleatório. É interessante notar que 10% dos estudantes do método tradicional têm nota entre 60 e 61 pontos, indicando um acúmulo de estudantes nessa faixa, indicando o senso comum de que as notas diversas vezes são arredondadas para 60. Já no PI, a quantidade de estudantes é aproximadamente 4% nessa faixa, mas aproximadamente 12% dos estudantes têm entre 61 e 66 pontos.

É importante lembrar que os dados apresentados tanto para a nota final quanto para  $g$  são apenas de estudantes que responderam o pré e pós teste, sendo apenas uma parcela do total de estudantes. Portanto, temos uma amostra sujeita ao viés de seleção, que não inclui, por exemplo, estudantes menos participativos, estudantes que não frequentam as aulas ou que abandonaram a disciplina. Porém, consideramos estes efeitos similares em ambos os métodos, e temos como foco a comparação entre os valores e não o valor absoluto.

Comparando a nota final com o ganho normalizado, vemos que o efeito positivo do

Figura 5 – Histograma e EDK da nota final para os métodos tradicional e *Peer Instruction*. A nota final é separada em dez faixas de 10.



uso de PI no ganho não é refletido na nota final. Como o  $g$  é uma medida feita através de um teste padronizado e avaliado [34], é plausível assumir que  $g$  reflita melhor o aprendizado de estudantes do que as notas finais que, por sua vez, dependem do sistema de avaliação individual e padrões de conhecimento de cada professor, que é ajustado tendo em vista o desempenho por ele observado em estudantes de uma turma. Dessa forma, vemos indícios de que a nota final não é uma boa medida para comparação para estudantes de grupos diferentes.

Embora outros fatores não refletidos nestes dados, como taxa de trancamento, influenciem também na retenção, o resultado da nota final sugere uma aprovação similar entre os dois métodos, considerando que a aprovação é feita através da nota final.

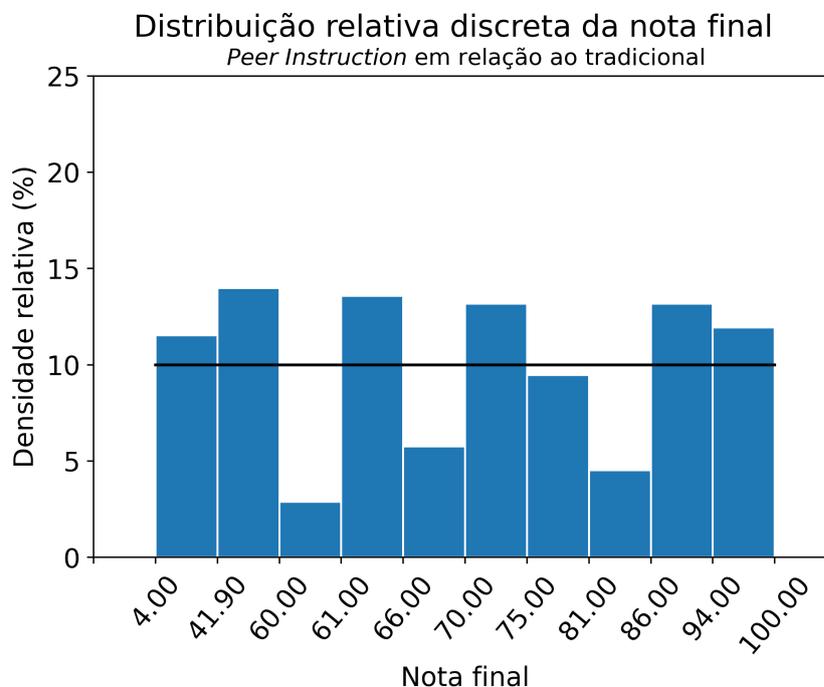
## 3.2 Ensino Remoto Emergencial

### 3.2.1 ERE vs métodos presenciais

Então, utilizamos as mesmas ferramentas estatísticas para comparar os dois métodos anteriores com o Ensino Remoto Emergencial. Vemos na tabela 2 que o número de respondentes dos dois testes no ERE é consideravelmente maior que no período presencial. Consideramos que o número alto de respondentes se deu pelo incentivo de nota ao responder o teste e pelo teste ter sido disponibilizado online por um período maior - no entanto, o tempo para fazer o teste ao começar era o mesmo do presencial, 45 minutos.

Primeiramente, comparamos o Ensino Remoto Emergencial com o *Peer Instruction*.

Figura 6 – Distribuição relativa discreta da nota final, com estudantes do método tradicional como grupo de referência e estudantes do *Peer Instruction* como grupo de comparação.

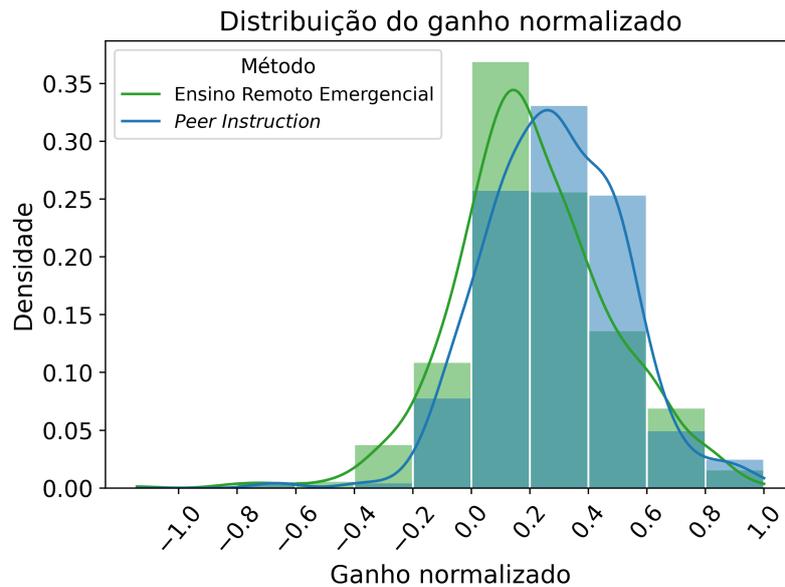


A tabela 2 indica que a diferença entre estes dois métodos é de 0,078. O teste-t dessa diferença apresenta o  $p\text{-valor} < 0,001$ , então descartamos a hipótese de as médias de  $g$  serem iguais. O  $d$  do ERE também é inferior ao  $d$  do PI, com valor de 0,812.

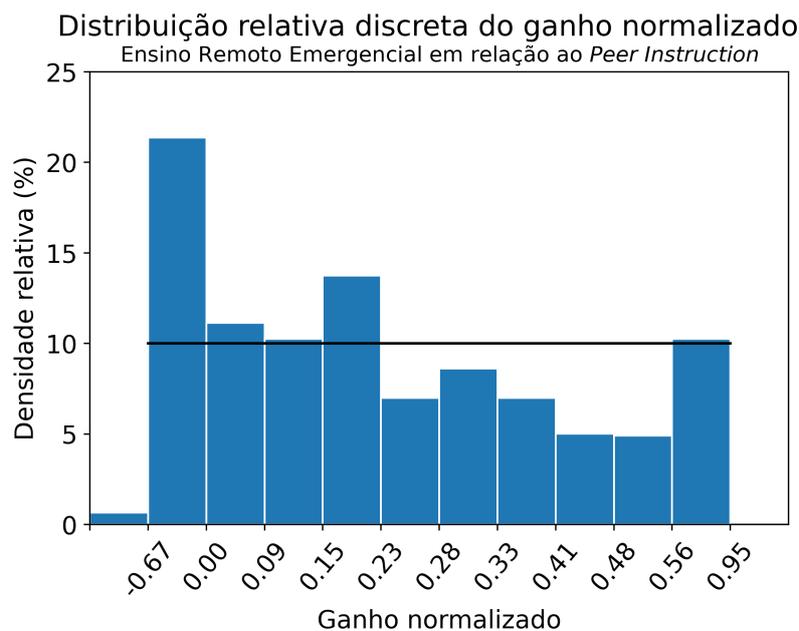
A Fig.7 mostra a distribuição do ganho normalizado do ERE e do *PI* e a distribuição relativa discreta do ganho, usando o ERE como grupo de referência e o *PI* como grupo de comparação.

Na Fig.7a, vemos, novamente, um deslocamento nas distribuições, com estudantes de *PI* tendo maior probabilidade de ter ganhos normalizados maiores em relação à estudantes do ERE, exceto pela faixa de ganho de 0,6 a 0,8, onde há mais estudantes do ERE. Esse aumento de estudantes em ganhos mais altos é também visto na Fig. 7b, onde vemos que o ultimo decil de ganho, com  $g$  entre 0,56 e 0,95, é a única para  $g > 0,23$  em que não há sub-representação dos estudantes do ERE em relação à estudantes de *PI*.

Então, comparamos o Ensino Remoto Emergencial com o método tradicional. A tabela 2 mostra que a média do  $g$  para estudantes do ERE é apenas um pouco inferior ao do método tradicional, com uma diferença de 0,010. O teste-t dessa diferença apresenta um  $p\text{-valor} = 0.528$ , então falhamos em descartar a hipótese de que as médias de  $g$  para os dois métodos sejam iguais. A Fig.8 mostra a distribuição do ganho normalizado do ERE e do *PI* e a distribuição relativa discreta do ganho, usando o ERE como grupo de referência



(a) Histograma e EDK do ganho normalizado para os métodos Ensino Remoto Emergencial e *Peer Instruction*. O ganho normalizado é separado em dez faixas de 0,2.

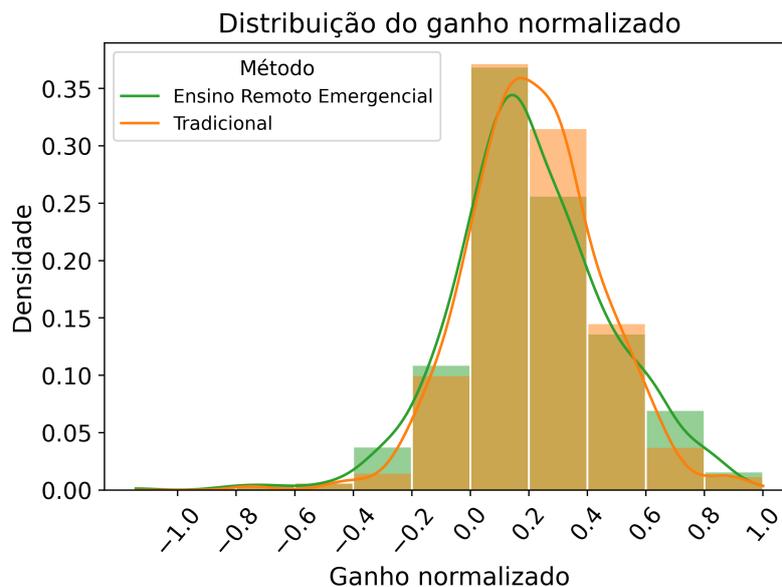


(b) Distribuição relativa discreta do ganho normalizado, com estudantes do *Peer Instruction* como grupo de referência e estudantes do Ensino Remoto Emergencial como grupo de comparação.

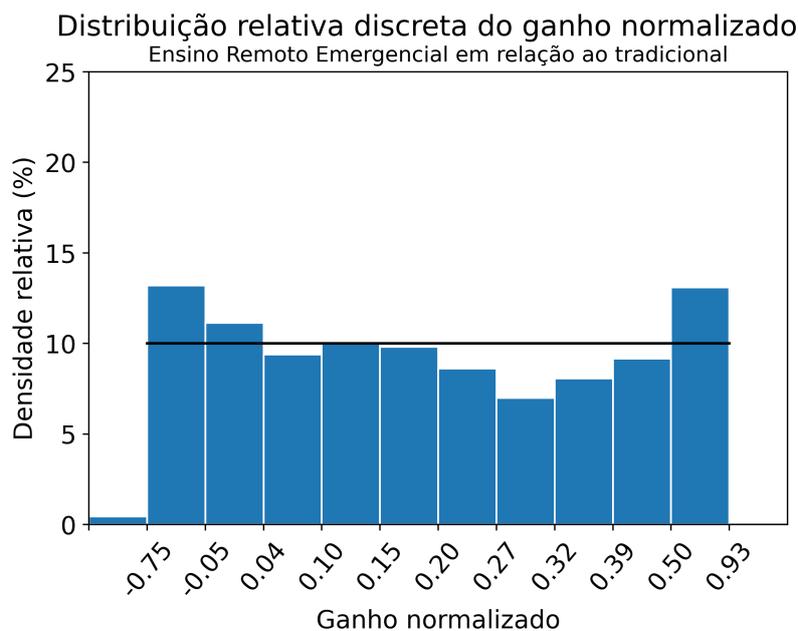
Figura 7 – Comparação do ganho normalizado entre estudantes Ensino Remoto Emergencial e estudantes do *Peer Instruction*.

e o PI como grupo de comparação.

Na Fig.8a vemos que as distribuições do ganho normalizado nos métodos tradicional e do ERE são similares. Para  $g < 0,2$ , a densidade de probabilidade do  $g$  para estudantes do ERE é próxima à do método tradicional. Para  $g > 0,5$ , a curva sugere que melhores estudantes têm melhor desempenho no ERE. A semelhança entre as distribuições também é vista na Fig.8b, onde vemos uma média de igualdade entre as distribuições, sem uma grande sub-representação ou sobre-representação de estudantes do método tradicional em



(a) Histograma e EDK do ganho normalizado para os métodos Ensino Remoto Emergencial e tradicional. O ganho normalizado é separado em dez faixas de 0,2.



(b) Distribuição relativa discreta do ganho normalizado, com estudantes do método tradicional como grupo de referência e estudantes do Ensino Remoto Emergencial como grupo de comparação.

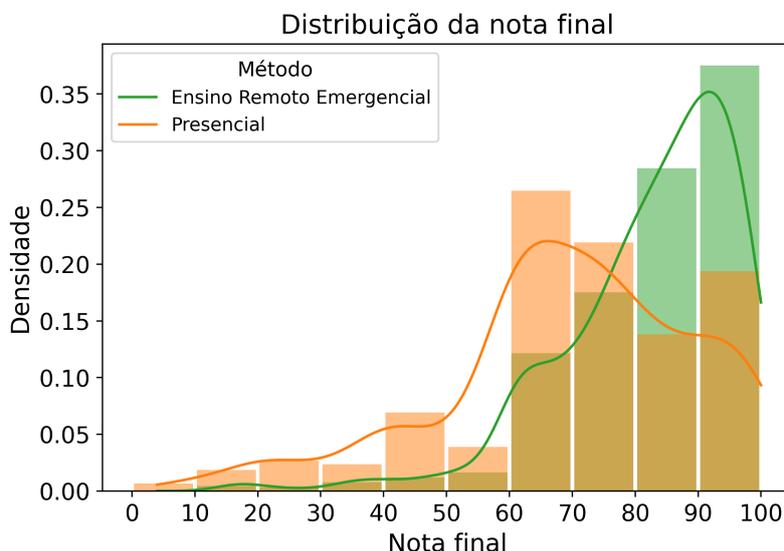
Figura 8 – Comparação do ganho normalizado entre estudantes do Ensino Remoto Emergencial e estudantes do método tradicional.

relação ao ERE.

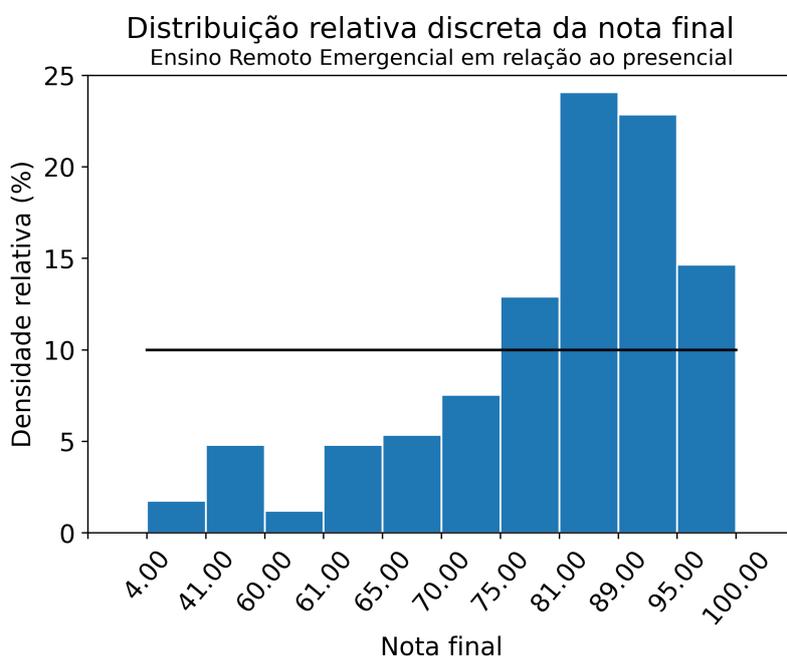
Em seguida, analisamos a nota final no ERE. Na tabela 2, vemos que a nota final têm média muito maior no ERE, com diferença de 12,88 pontos na média. Também na tabela 2, vemos que o desvio padrão da nota final no ERE é muito menor que nos outros dois métodos, indicando uma homogeneização das notas.

Como visto nas figuras 5 e 6, as notas nos dois métodos do período presencial é

similar, então optamos por trata-los em conjunto para comparação com o ERE. Assim, a Fig.9 mostra a distribuição do ganho normalizado do ERE e do período presencial e a distribuição relativa discreta do ganho, usando o ERE como grupo de referência e o período presencial como grupo de comparação.



(a) Histograma e EDK da nota final para o Ensino Remoto Emergencial e o período presencial. A nota final é separada em dez faixas de 10.



(b) Distribuição relativa discreta da nota final, com estudantes do período presencial como grupo de referência e estudantes do Ensino Remoto Emergencial como grupo de comparação.

Figura 9 – Comparação da nota final entre estudantes do Ensino Remoto Emergencial e estudantes do período presencial.

A diferença da nota entre os métodos é vista da distribuição mostrada na Fig9a, com estudantes do ERE tendo probabilidades muito maiores de ter notas > 80. Na Fig.9b, vemos uma grande sub-representação de estudantes do ERE nos primeiros decis, de nota

até 75, e uma grande sobrerrepresentação de estudantes do período presencial em decis de nota acima de 75. Então, há uma grande concentração de estudantes do ERE em notas altas. O resultado da nota durante o ERE pode ser vista também no relatório de retenção da UFMG, que mostra que, em 2020, 81,8% dos estudantes foi aprovado, resultado muito superior à média de 52% nos anos anteriores. Isso sugere que, durante esse período, as avaliações foram mais lenientes, e foram aprovados estudantes com compreensão do conteúdo abaixo do que seria requerido no período presencial anterior.

### 3.2.2 Por semestre

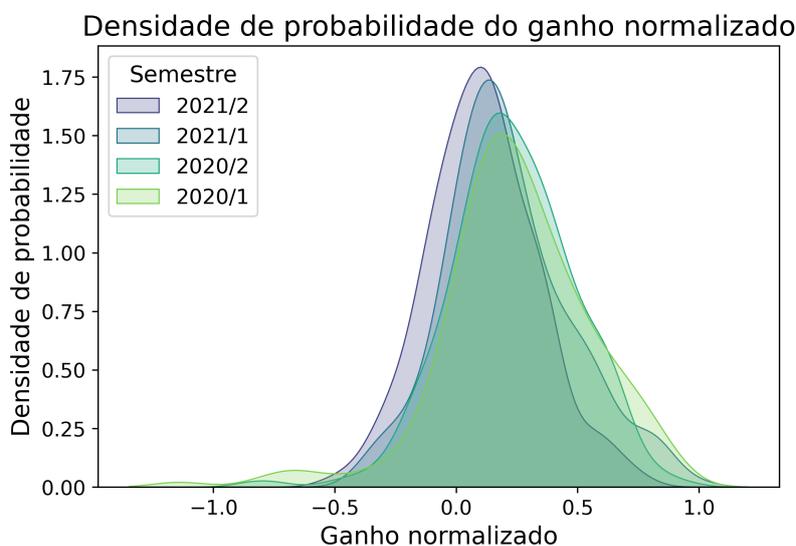
Durante o Ensino Remoto Emergencial, nos interessamos em saber se o semestre se tornou uma variável relevante pois, a cada semestre, os estudantes que cursavam Fundamentos de Eletromagnetismo haviam feito mais disciplinas que são pré-requisito da disciplina no regime de ERE. A partir do relatório de retenção da UFMG [29], vemos a alta aprovação de estudantes também nestas disciplinas, e supomos que, nelas, as avaliações também foram mais lenientes quando comparadas com o período presencial. Assim, separamos o ganho normalizado e a nota final entre os semestres de ERE. O resultado é mostrado na tabela 3. Uma versão completa da tabela, com os resultados do pré e pós teste, é mostrada no apêndice 4.

Tabela 3 – Média da nota final e ganho normalizado para os Ensino Remoto Emergencial, separados por semestre e gênero masculino (Masc.) e feminino (Fem.). Número de respondentes ( $N$ ), média ( $m$ ) e desvio padrão ( $\sigma$ ) da nota final e ganho normalizado ( $g$ ) e  $D$  de Cohen ( $d$ ) de estudantes que completaram o pré e o pós teste.

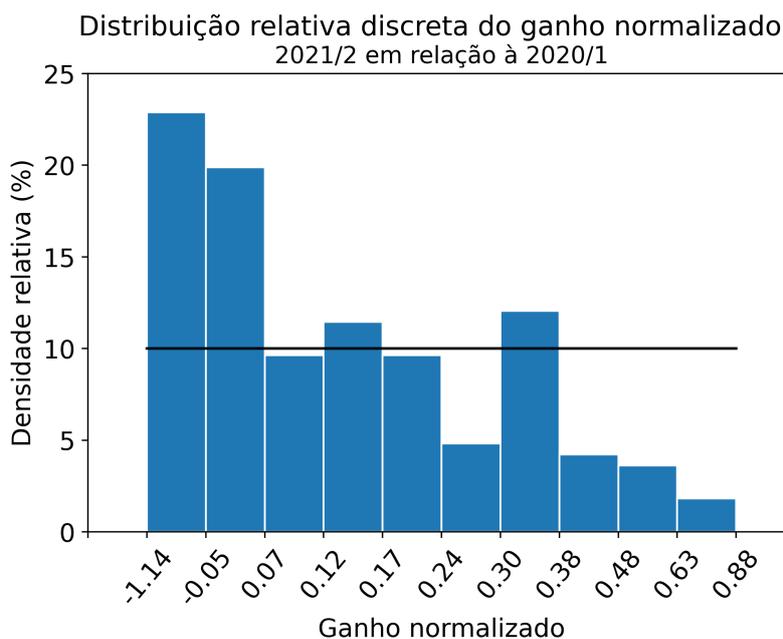
Método	Gênero	N	Nota final		$g$		$d$
			$m$	$\sigma$	$m$	$\sigma$	
2020/1	Masc.	111	88,6	8,3	0,270	0,351	0,938
	Fem.	63	91,6	6,2	0,200	0,214	1,170
	Ambos	174	89,7	7,7	0,245	0,310	0,943
2020/2	Masc.	189	81,9	13,0	0,258	0,276	1,107
	Fem.	112	83,9	13,0	0,181	0,215	0,950
	Ambos	301	82,7	13,0	0,230	0,256	1,017
2021/1	Masc.	202	81,4	11,9	0,231	0,262	0,772
	Fem.	70	78,2	12,3	0,164	0,248	0,746
	Ambos	272	80,5	12,1	0,211	0,260	0,745
2021/2	Masc.	110	77,3	20,0	0,158	0,217	0,625
	Fem.	56	76,7	17,6	0,012	0,203	0,111
	Ambos	166	77,1	19,2	0,109	0,222	0,466

Na tabela 3, vemos que o  $g$  médio decresce ao longo dos semestres, sendo muito menor no último semestre de ERE, 2021/2. Também vemos efeito parecido no  $d$ , com exceção do semestre 2020/2, onde o  $d$  é maior que no semestre anterior. Na Fig.10a,

vemos que, a cada semestre, a distribuição se desloca para a esquerda, indicando uma probabilidade maior de ganhos mais baixos. A Fig.10b mostra a distribuição relativa entre o



(a) Estimativa de densidade kernel do ganho normalizado no Ensino Remoto Emergencial, separado por semestre.



(b) Distribuição relativa discreta do ganho normalizado, com estudantes de 2020/1 como grupo de referência e estudantes de 2021/2 como grupo de comparação.

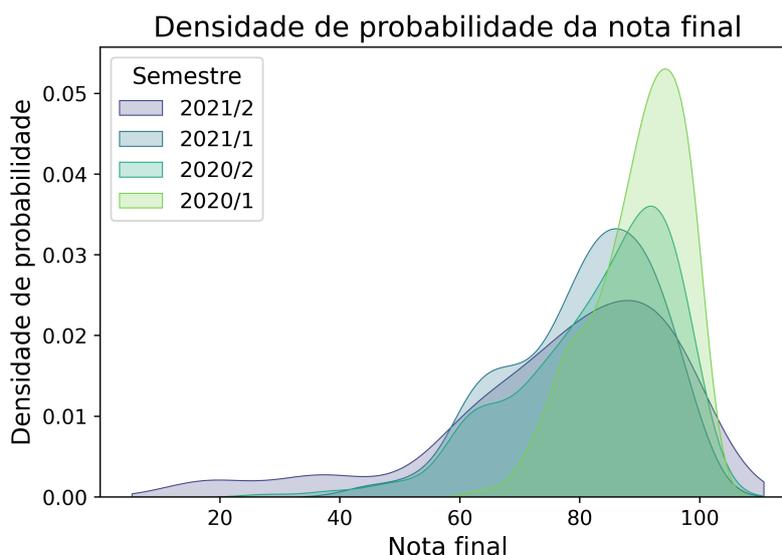
Figura 10 – Comparação do ganho normalizado entre os semestres de Ensino Remoto Emergencial.

primeiro e o último semestre de ERE. Nesse gráfico, vemos uma grande sobrerepresentação de estudantes do semestre 2021/2 em ganhos negativos. Já para decis com  $g > 0,38$ , há menos de 5% de estudantes do último semestre do ERE.

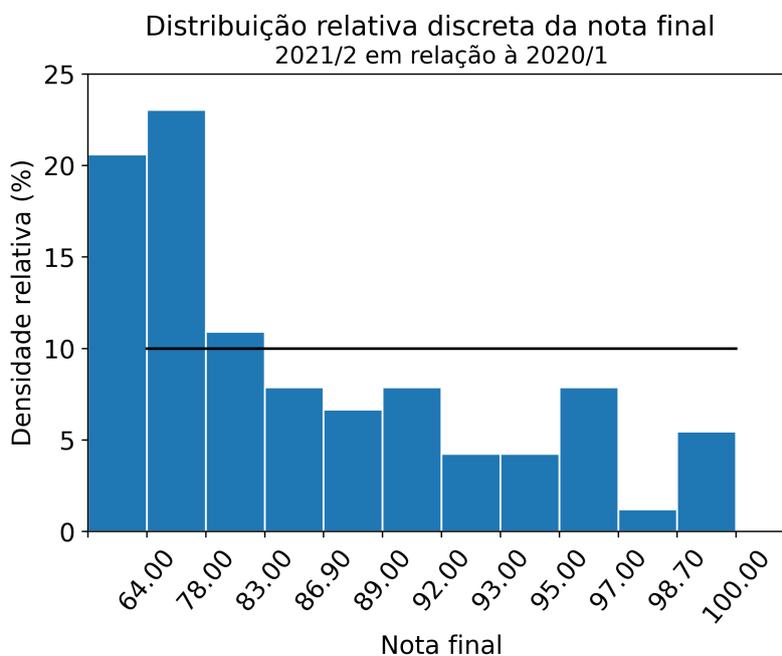
Os resultados do ganho por semestre indicam que a efetividade do ensino no período diminuiu ao longo dos semestres, sendo particularmente inferior no último semestre. O resultado pode ser reflexo da alta aprovação em disciplinas que são pré-requisito de

Fundamentos de Eletromagnetismo [29], bem como reflexo do esgotamento de estudantes e professores no período prolongado de atividades virtuais em disciplinas que não foram, a princípio, pensadas para este modelo.

Então, fazemos a análise das notas durante o ERE. Na tabela 3, vemos que a nota média é menor a cada semestre de ERE, com diferença de 12,6 entre o primeiro e o último semestre. Na Fig11, vemos a distribuição da densidade de probabilidade da nota por semestre e a distribuição relativa do último semestre em relação ao primeiro.



(a) Estimativa de densidade kernel da nota final no Ensino Remoto Emergencial, separado por semestre.



(b) Distribuição relativa discreta da nota final, com estudantes de 2020/1 como grupo de referência e estudantes de 2021/2 como grupo de comparação.

Figura 11 – Comparação da nota final entre os semestres de Ensino Remoto Emergencial.

Na Fig.11a, vemos que não só a densidade de probabilidade da nota decai ao

longo dos semestres, mas também as curvas ficam mais largas, mostrando uma maior variação na nota, indicando que, ao longo dos semestres, as avaliações foram ajustadas na tentativa de refletir melhor o conhecimento dos estudantes. A Fig11b mostra grande sobre-representação de estudantes de 2021/2 em notas abaixo de 78, comparado com estudantes de 2020/1. Neste gráfico, é interessante notar também os limites de nota dos decis dos estudantes de 2020/1: os 10% de estudantes com notas mais altas têm nota entre 98.7 e 100, o decil anterior têm notas entre 97 e 98.7, etc. Esses limites mostram a concentração dos estudantes de 2020/1 em notas altas.

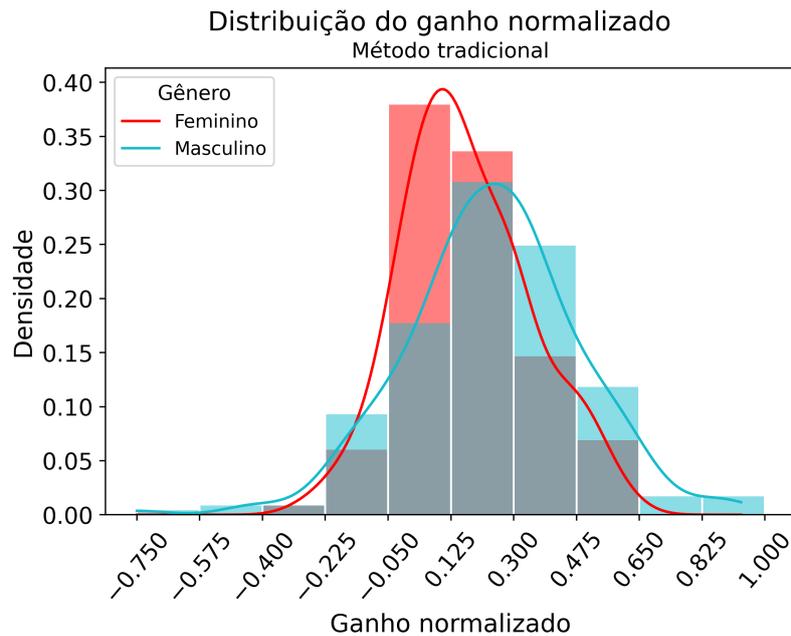
### 3.3 Gap de gênero

Analizamos a diferença de ganho normalizado e nota final entre homens e mulheres para cada método separadamente. Primeiramente, vemos na tabela 2 que, em todos os métodos, há um maior número de participantes homens.

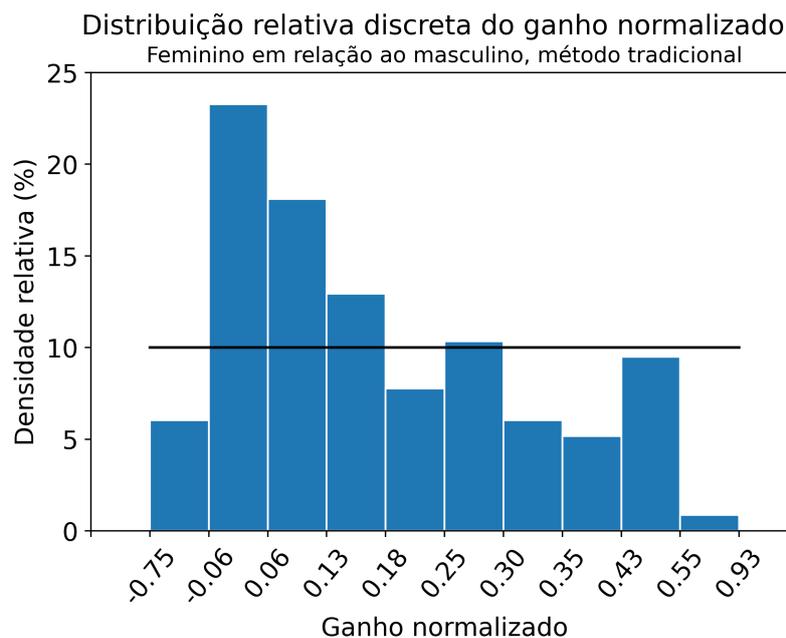
Para o método tradicional, vemos na tabela 2 que homens têm um ganho médio de 0,239 e mulheres têm um ganho médio de 0,166. A diferença de 0,073 representa um aumento de aproximadamente 44% de ganho dos homens em relação às mulheres. Também vemos diferença no  $d$ , com homens tendo  $d = 0,932$  e mulheres tendo  $d = 0,864$ . A Fig. 12 mostra a distribuição do ganho separado por gênero e a distribuição relativa do ganho. Na Fig. 12a, vemos que, apesar de homens terem maiores probabilidades de ganhos até  $-0,050$ , após esse valor eles apresentam probabilidades maiores de ter ganhos maiores que mulheres. Isso é refletido na distribuição relativa da Fig. 12b. Neste gráfico, vemos que, apesar de no primeiro decil de ganho dos estudantes do gênero masculino haver uma sub-representação de estudantes do gênero feminino, há uma tendência de sobre-representação das mulheres em ganhos mais baixos e sub-representação em ganhos mais altos, indicando que mulheres estão concentradas em ganhos mais baixos em relação aos homens.

Analizando a nota final, vemos na tabela 2 que estudantes do gênero masculino têm uma nota média de 70,0 e estudantes do gênero feminino têm uma nota média de 68,1, uma diferença na nota final de 1,9, representando um aumento relativo de aproximadamente 3% da nota de homens em relação à nota de mulheres. O teste-t para essa diferença apresenta um p-valor= 0.392, então falhamos em descartar a hipótese de que a média das notas sejam iguais.

Na Fig.13, vemos a distribuição da nota final no método tradicional separada por gênero e a distribuição relativa da nota final. Na Fig.13a a distribuição para mulheres é maior em notas entre 30 e 70 pontos, e para homens é maior em notas a partir de 70. No entanto, as distribuições são parecidas, ambas com pico próximo de 70 pontos. Na Fig. 13b, vemos que há uma grande sobre-representação de mulheres no segundo decil de nota dos homens, com faixa entre 42 e 60. No entanto, não há uma tendência crescente de



(a) Histograma e EDK do ganho normalizado para estudantes do método tradicional, separado por gênero. O ganho normalizado é separado em dez faixas de 0,175.

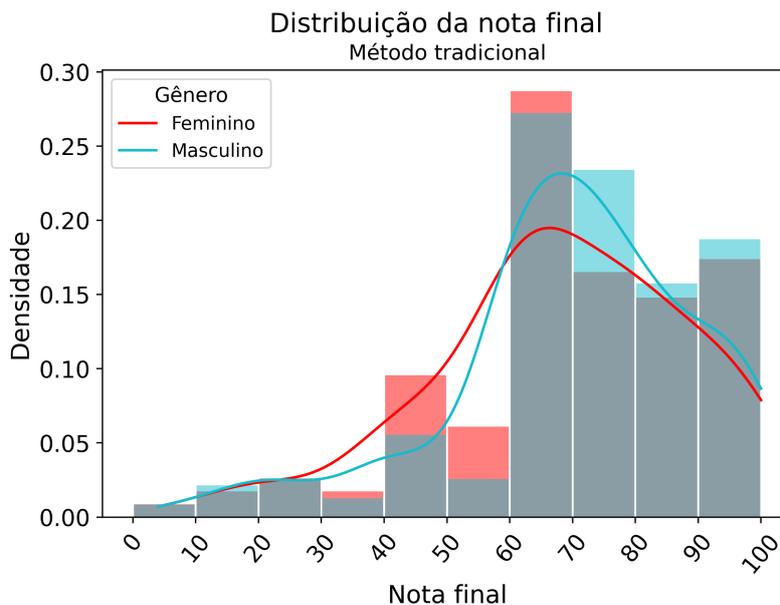


(b) Distribuição relativa discreta do ganho normalizado, com estudantes do gênero masculino como grupo de referência e estudantes do gênero feminino como grupo de comparação.

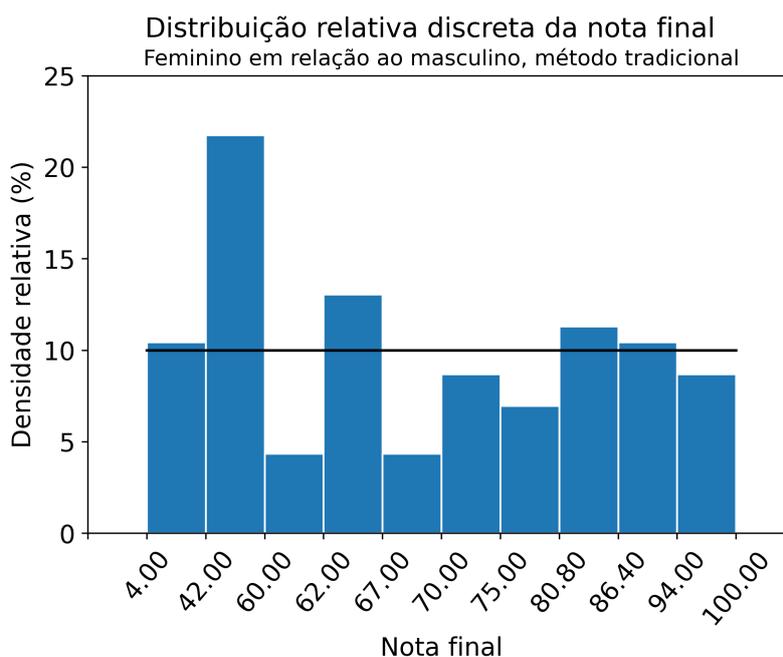
Figura 12 – Comparação do ganho normalizado entre estudantes do gênero masculino e estudantes do gênero feminino do método tradicional.

sub-representação ou sobre-representação entre os dois grupos. Os resultados da nota final reforçam a ideia de que a nota final não é um bom instrumento para medir e comparar a aprendizagem entre grupos.

Em seguida, analisamos as diferenças entre estudantes do gênero masculino e feminino no *Peer Instruction*. A tabela 2 mostra que homens têm um ganho médio de 0,318 e mulheres têm um ganho médio de 0,237. A diferença de 0,081 é maior que a diferença no



(a) Histograma e EDK da nota final para estudantes do método tradicional, separado por gênero. A nota final é separada em dez faixas de 10.

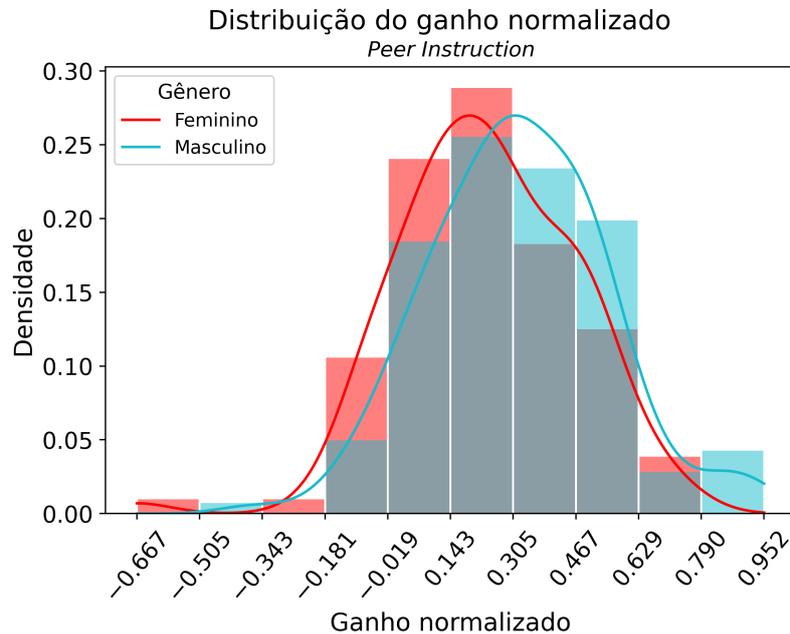


(b) Distribuição relativa discreta da nota final, com estudantes do gênero masculino como grupo de referência e estudantes do gênero feminino como grupo de comparação.

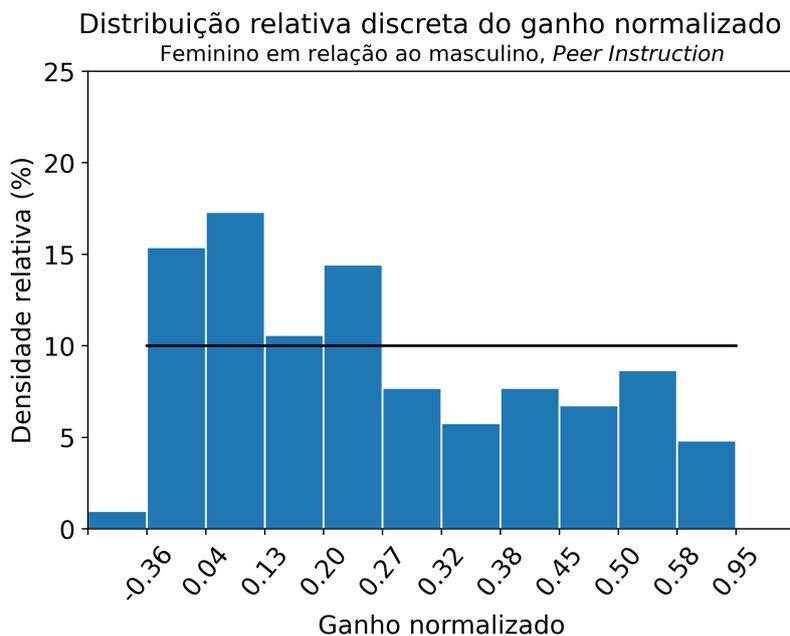
Figura 13 – Comparação da nota final entre estudantes do gênero masculino e estudantes do gênero feminino do método tradicional.

método tradicional, porém como a média geral é maior, esse valor representa um aumento de aproximadamente 34% de ganho dos homens em relação às mulheres, um aumento relativo um pouco menor, mas considerável. O teste t para essa diferença apresenta um p-valor= 0,006, então descartamos hipótese de as médias de  $g$  serem iguais. Para o  $d$  de cohen, vemos homens com  $d = 1,248$  e mulheres tendo  $d = 1,123$ , indicando um efeito maior do curso para homens do que para mulheres. Exploramos a diferença de  $g$  nos

gráficos de distribuição do ganho e distribuição relativa, mostrados na Fig.14. Na Fig.14a,



(a) Histograma e EDK do ganho normalizado para estudantes de *Peer Instruction*, separado por gênero. O ganho normalizado é separado em dez faixas de 0,162.



(b) Distribuição relativa discreta do ganho normalizado, com estudantes do gênero masculino como grupo de referência e estudantes do gênero feminino como grupo de comparação.

Figura 14 – Comparação do ganho normalizado entre estudantes do gênero masculino e estudantes do gênero feminino para *Peer Instruction*.

vemos que homens têm probabilidades maiores de ter ganhos maiores que mulheres no PI, mas, ao comparar com o método tradicional, vemos que as curvas aparentam estar mais próximas, indicando uma diferença menor entre os gêneros. Na Fig. 14b, vemos um efeito parecido, ainda com sobrerrepresentação de mulheres em decis de ganhos mais baixos e sub-representação de mulheres em decis de ganhos mais altos. No entanto, as proporções

são menos extremas que no método tradicional. Estes resultados indicam um *gap* de gênero menor no PI, mas não nulo.

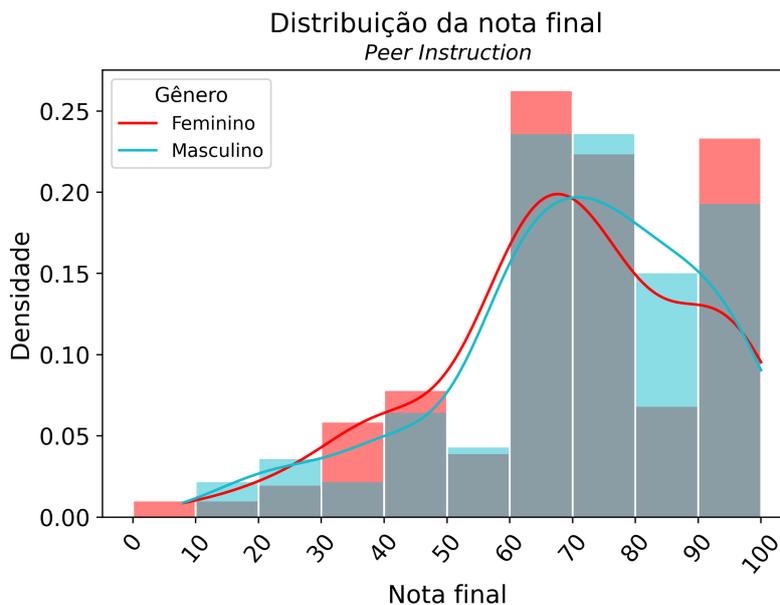
Então, comparando a nota final entre gêneros no PI, vemos na tabela 2 que a nota final média para homens é 70,2 e para mulheres é 69,0, uma diferença de 1,2 para a qual o teste-t apresenta um p-valor= 0,654, então falhamos em descartar a hipótese de as notas finais médias serem iguais. A distribuição da nota final separada por gênero e a distribuição relativa são mostradas na Fig.15. Na Fig.15a vemos que a distribuição de notas finais de homens e mulheres é muito parecida. A Fig.15b mostra que, assim como nos método tradicional, não há uma tendência de sub-representação ou sobrerrepresentação de mulheres em relação à homens nos decis de nota.

Por fim, analisamos o *gap* de gênero durante o Ensino Remoto Emergencial. Na tabela 2, vemos que homens têm um *g* médio de 0,233 e mulheres têm um *g* médio de 0,150. A diferença de 0,083 apresenta um aumento relativo de aproximadamente 55%, o maior aumento dentre os métodos. O teste t para essa diferença apresenta um p-valor < 0,001, então descartamos a hipótese de que as médias sejam iguais. Para o *d* de cohen, temos homens com  $d = 0,8665$  e mulheres com  $d = 0,782$ , indicando que, assim como nos outros métodos, os homens têm um aproveitamento maior do curso. Na Fig. 15 vemos a distribuição do ganho separado por gênero e a distribuição relativa.

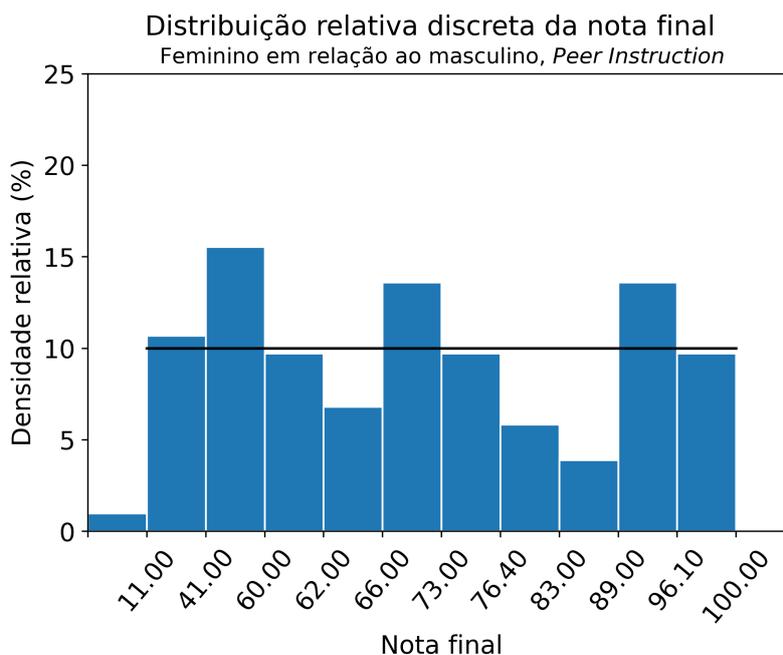
A Fig.16a mostra que, no ERE, homens têm probabilidades maiores de ter ganhos maiores que mulheres. No entanto, as curvas da distribuição não são tão distantes quanto se deduziria pelos resultados da média de *g*, sendo muito próximas entre 0 e 0,4. Contudo, há um extremo de ganhos mais altos para os quais homens apresentam uma probabilidade maior. Na Fig.16b, vemos que há, como nos outros métodos, uma sobrerrepresentação de mulheres em decis de ganho mais baixo e uma sub-representação de mulheres em decis de ganhos mais altos.

Para a nota final, vemos na tabela 2 que a diferença entre as médias de nota final para homens e mulheres é de apenas 0,6. O teste-t dessa diferença apresenta o p-valor= 0,489. Na Fig.17, vemos a distribuição da nota final e a distribuição relativa de mulheres em relação à homens. A Fig.17a mostra uma distribuição muito parecida entre homens e mulheres. A Fig.17b mostra que a distribuição de mulheres entre os decis de notas dos homens é equivalente, sem uma tendência de sobrerrepresentação ou sub-representação em ganhos altos ou baixos. Como nos outros métodos, não há evidências de uma diferença na nota final entre homens e mulheres. Ressaltamos aqui que as avaliações durante o ERE foram, em geral, muito homogêneas e não refletiram bem o conhecimento de estudantes.

Por fim, nos questionamos se o *gap* de gênero para o ganho mudava ao longo dos semestres de ensino remoto. Na Fig.18, mostramos a distribuição de ganho separado por gênero e a densidade relativa para cada semestre de ERE. Nas Figs. 18a, 18c, 18e, vemos que as distribuições do ganho nos três primeiro semestres de ERE são relativamente



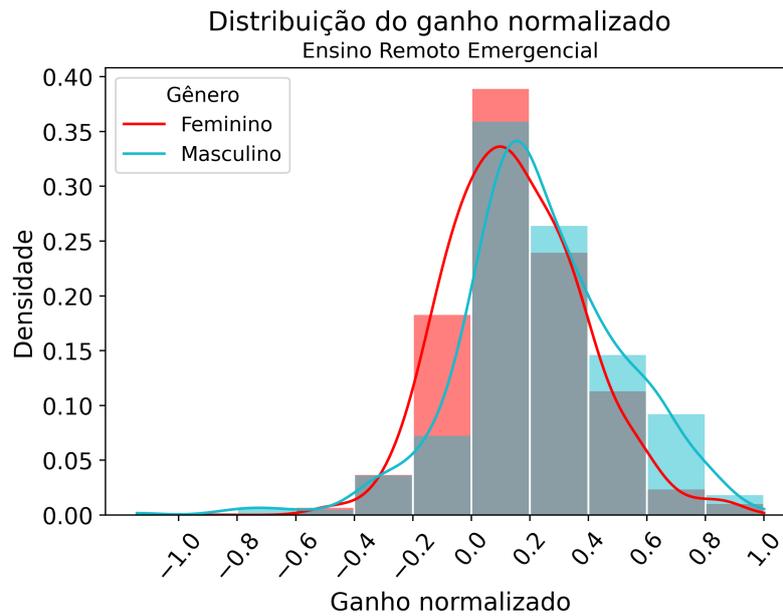
(a) Histograma e EDK da nota final para estudantes do *Peer Instruction*, separado por gênero. A nota final é separada em dez faixas de 10.



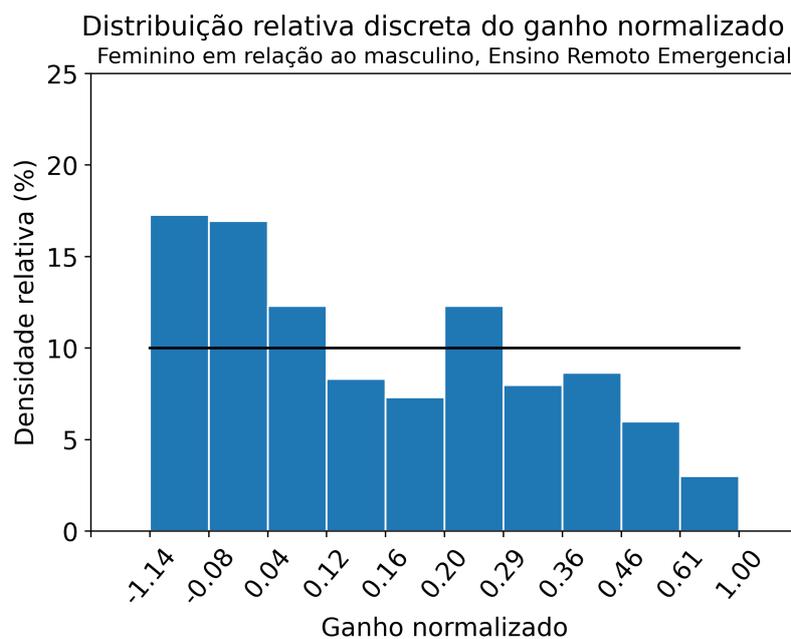
(b) Distribuição relativa discreta da nota final, com estudantes do gênero masculino como grupo de referência e estudantes do gênero feminino como grupo de comparação.

Figura 15 – Comparação da nota final entre estudantes do gênero masculino e estudantes do gênero feminino do *Peer Instruction*

semelhantes, mas homens apresentam uma cauda em ganhos mais altos, indicando que, nestes semestres, há um grupo maior de homens com alto aprendizado. Esse grupo é refletido nos gráficos de densidade relativa das Figs.18b, 18d, 18f, onde há uma grande sub-representação de mulheres dos últimos dois decis de ganho dos homens. Na tabela 3, vemos diferença na média de homens e mulheres nestes semestres. No entanto, notamos que, em 2020/1, mulheres apresentam um  $d$  maior que homens, sendo o único caso de



(a) Histograma e EDK do ganho normalizado para estudantes do Ensino Remoto Emergencial, separado por gênero. O ganho normalizado é separado em dez faixas de 0,20.

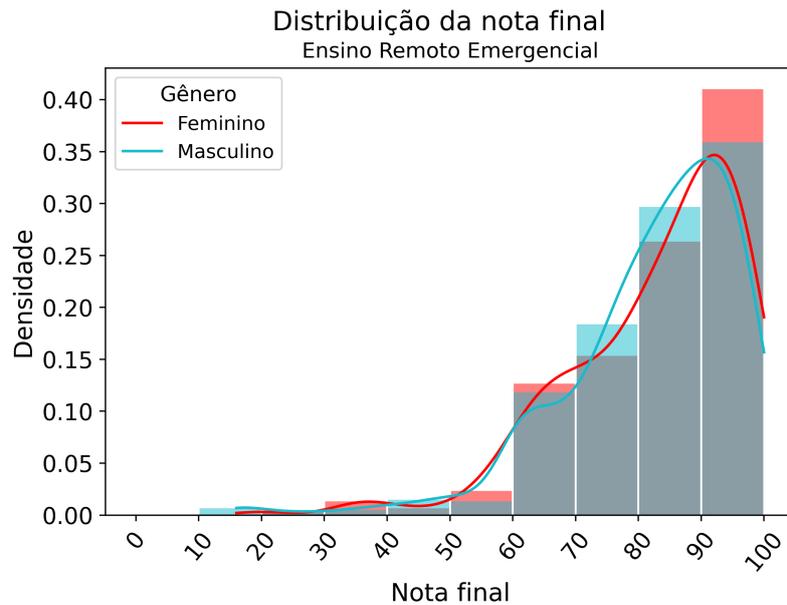


(b) Distribuição relativa discreta do ganho normalizado, com estudantes do gênero masculino como grupo de referência e estudantes do gênero feminino como grupo de comparação.

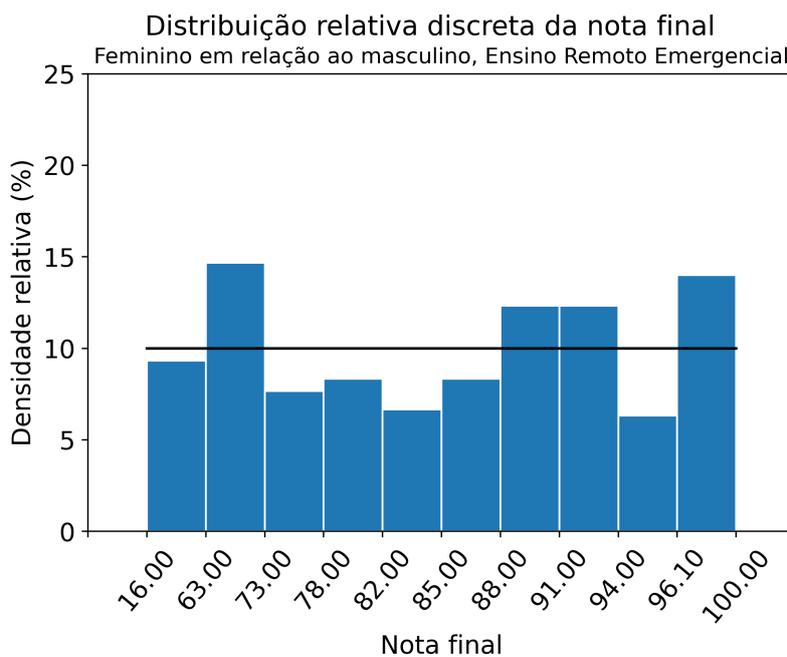
Figura 16 – Comparação do ganho normalizado entre estudantes do gênero masculino e estudantes do gênero feminino do Ensino Remoto Emergencial.

discrepância na diferença entre  $g$  e  $d$ . Em 2021/1 ainda há diferença entre os  $d$ , mas os valores são próximos. Fazendo o teste-t para a diferença entre  $g$  de homens e mulheres temos:

- 2020/1: o teste-t para a diferença entre  $g$  de homens e mulheres retorna um o-valor de 0,156



(a) Histograma e EDK da nota final para estudantes do Ensino Remoto Emergencial, separado por gênero. A nota final é separada em faixas de 10.



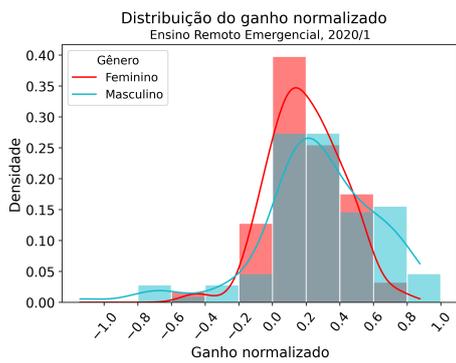
(b) Distribuição relativa discreta da nota final, com estudantes do gênero masculino como grupo de referência e estudantes do gênero feminino como grupo de comparação.

Figura 17 – Comparação da nota final entre estudantes do gênero masculino e estudantes do gênero feminino do Ensino Remoto Emergencial.

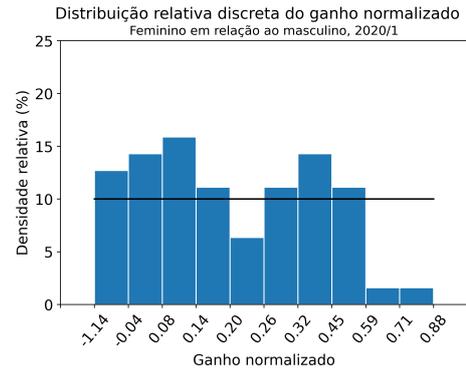
- 2020/2: o teste-t para a diferença entre  $g$  de homens e mulheres retorna um o-valor de 0,012
- 2021/1: o teste-t para a diferença entre  $g$  de homens e mulheres retorna um o-valor de 0,064

Então, falhamos em descartar a hipótese de que a diferença entre os semestres 2020/1

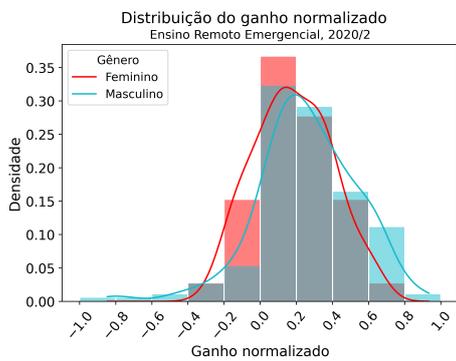
e 2021/1 seja nula. No último semestre, porém, vemos uma diferença acentuada entre homens e mulheres, com uma grande sobrerrepresentação de mulheres nos dois primeiros decis de ganho mais baixo dos homens, com faixas de ganho até 0, e sub-representação de mulheres em todos os decis seguintes. Na tabela 3, vemos também uma grande diferença no ganho, com mulheres tendo o ganho próximo de zero. O teste-t para essa diferença apresenta um p-valor de 0,000, então descartamos a hipótese de que a diferença seja nula. Há também uma grande diferença no  $d$ , e o  $d$  de mulheres em 2021/2 é o único considerado pequeno, indicando que a disciplina não influenciou o resultado do pós teste em relação ao pré teste.



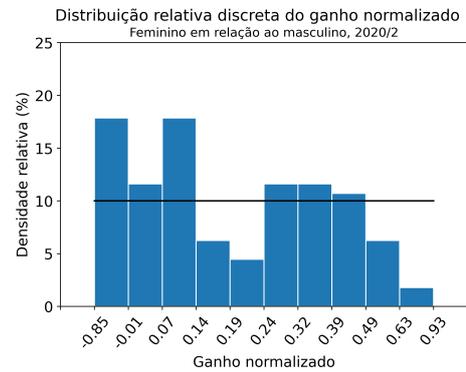
(a) Distribuição do ganho normalizado em 2020/1.



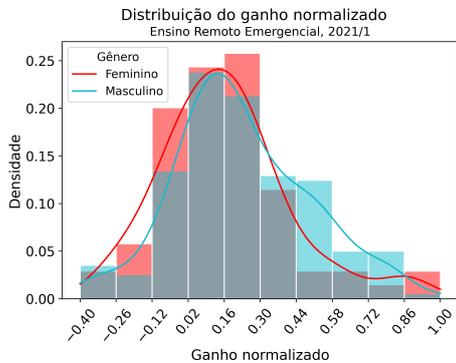
(b) Distribuição relativa discreta do ganho normalizado.



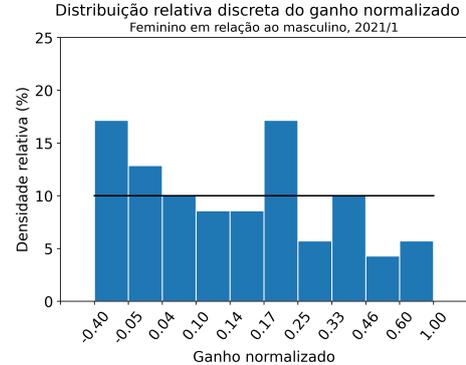
(c) Distribuição do ganho normalizado em 2020/2.



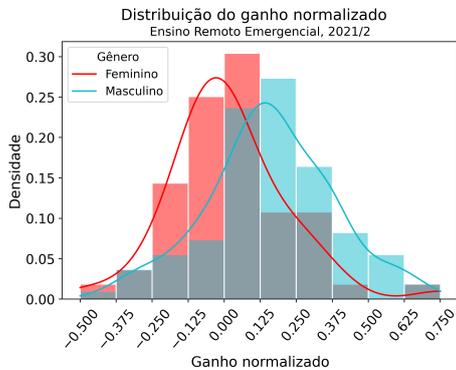
(d) Distribuição relativa discreta do ganho normalizado.



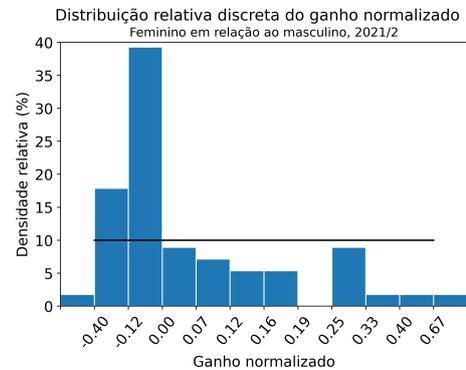
(e) Distribuição do ganho normalizado em 2021/1.



(f) Distribuição relativa discreta do ganho normalizado.



(g) Distribuição do ganho normalizado em 2021/2.



(h) Distribuição relativa discreta do ganho normalizado.

Figura 18 – Comparação do ganho normalizado separado por gênero para cada semestre do Ensino Remoto Emergencial.

## 4 CONCLUSÕES

Na pesquisa em ensino, as metodologias ativas são bem estabelecidas como uma alternativas ao método tradicional que promove maior aprendizagem conceitual. Dentre as metodologias ativas mais utilizadas no ensino de Física, está *Peer Instruction*, que consiste em atividades preparatórias pré aula e discussões em grupo durante a aula. Também é comum o uso do PI em conjunto com *Just in Time Teaching*, que consiste no planejamento das aulas a partir do *feedback* de atividades preparatórias pré aula. Nesse trabalho, estudamos o uso combinado destas duas metodologias na disciplina de Fundamentos de Eletromagnetismo na UFMG, comparando com a metodologia tradicional. Além disso, estudamos o Ensino Remoto Emergencial, modelo adotado no período da pandemia de COVID-19, comparando com as outras duas metodologias e analisando o resultado por semestre de ERE.

Para a análise, temos os pré teste, pós teste feito pelo teste BEMA e nota final dos estudantes que responderam ambos os pré e pós teste. A partir do pré e pós teste, calculamos o ganho normalizado  $g$  e o D de Cohen  $d$ , como forma de quantificar o aprendizado dos estudantes. Então, analisamos como a metodologia de ensino e o gênero dos estudantes afetam a média de  $g$  e da nota final, testando a significância com o teste t, e usamos o  $g$  e a nota final como variáveis para calcular a densidade de probabilidade e a distribuição relativa, comparando grupos.

Na análise de dados, vimos que há um aumento significativo do  $g$  para estudantes do PI. Além disso, vimos que estudantes do PI têm maiores probabilidades de ter ganhos maiores, e há uma sobrerrepresentação de estudantes do PI em ganhos mais altos em relação ao método tradicional. No entanto, não encontramos o mesmo efeito da metodologia de ensino sobre a nota final. Assumimos que  $g$  seja uma medida mais sensível ao aprendizado de estudantes que a nota final, que depende dos critérios individuais de cada professor. Então, sugerimos que a nota não seja um bom padrão para a avaliação e comparação de aprendizado de estudantes. Ressaltamos que há apenas dois professores que aplicam o PI, o que pode gerar viés de confusão com as habilidades destes professores. Contudo, os resultados do PI são consistentes com a literatura [2], [45].

Ao incluirmos o Ensino Remoto Emergencial como uma metodologia na análise, vemos que o resultado do ganho nesse período foi, em média, muito similar ao método tradicional, sem diferença significativa. No entanto, ao olharmos para a nota final, vemos um

aumento expressivo dessa medida comparado às outras duas metodologias. Ao separarmos o  $g$  para cada semestre de ERE, vemos que há uma queda no ganho a cada semestre, sendo consideravelmente menor no último semestre, indicando que o ensino durante o ERE foi cada vez menos efetivo. Este resultado pode ser reflexo da também alta aprovação nas disciplinas que são pré-requisito de Fund. Eletro ou do esgotamento do período pandêmico. A nota final também apresenta queda ao longo dos semestres de ERE, sugerindo que as avaliações foram ajustadas a cada semestre.

Por fim, procuramos entender o *gap* de gênero. Ao separar o ganho por gênero em cada metodologia, vemos que, em aulas tradicionais, homens têm ganhos médios maiores, com diferença significativa. Também vemos que homens têm probabilidades maiores de ter ganhos maiores e há uma sub-representação de mulheres em ganhos mais altos em relação a homens. No PI, vemos os mesmos efeitos na comparação de homens e mulheres, porém a diferença é um pouco menor. Este resultado traz evidências de que o *gap* de gênero pode ser menor no PI, mas ainda existe. Portanto, para extinguir o *gap* de gênero, apenas usar PI não é suficiente. Separando a nota final por gênero nestes métodos, vemos que não há diferença significativa, reforçando a ideia de que a nota final é um instrumento menos sensível. Para o ERE, também vemos um *gap* de gênero refletido tanto no  $g$  médio quanto na distribuição, mas, na distribuição de ganho dos homens, vemos uma cauda para ganhos mais altos, indicando que os melhores estudantes homens têm ganhos maiores no ERE. Ao separar o ganho no ERE por gênero e semestre, vemos que a diferença na distribuição de homens e mulheres é menos expressiva nos três primeiros semestres, e, no último semestre, temos uma grande diferença entre homens e mulheres. Estes dados devem ser analisados com cuidado, pois, ao separar por gênero e semestre, temos grupos menores, e os dados se tornam menos representativos. As razões para o aumento considerável do *gap* de gênero são incertas, mas o efeito pode ser resultado de maiores provas orais, exigindo uma prova de conhecimentos mais pública, frente a colegas.

Acreditamos que, ao promover maior participação ativa dos estudantes, promovemos uma maior aprendizagem conceitual, mas que, para a diminuição do *gap* de gênero, a mera adoção de uma metodologia ativa não é suficiente. Vemos, também, que a nota final é um instrumento com pouca variação, e deve ser tratada com cuidado para comparação do desempenho de grupos diferentes. Ainda que a análise da nota final seja relacionada à aprovação, recomendamos que análises mais aprofundadas sejam feitas para avaliar a retenção entre gêneros ou metodologias de ensino diferentes. Quanto ao ERE, ressaltamos que são necessários mais estudos para avaliar a piora do ganho ao longo dos semestres e do *gap* de gênero no último semestre. Também recomendamos uma análise cuidadosa dos resultados pós pandemia, afim de entender o impacto do período em gerações futuras de estudantes.

## REFERENCIAS

- [1] Halloun, Ibrahim Abou e David Hestenes: *The Initial Knowledge State of College Physics Students*. American Journal of Physics, 53(11):1043–1055, novembro 1985, ISSN 0002-9505. Citado na página 9.
- [2] Von Korff, Joshua, Benjamin Archibeque, K. Alison Gomez, Tyrel Heckendorf, Sarah B. McKagan, Eleanor C. Sayre, Edward W. Schenk, Chase Shepherd e Lane Sorell: *Secondary Analysis of Teaching Methods in Introductory Physics: A 50k-Student Study*. American Journal of Physics, 84(12):969–974, dezembro 2016, ISSN 0002-9505. Citado 3 vezes nas páginas 9, 23 e 44.
- [3] Prince, Michael: *Does Active Learning Work? A Review of the Research*. Journal of Engineering Education, 93(3):223–231, 2004, ISSN 2168-9830. Citado na página 9.
- [4] Quibao, Matheus Pinheiro, Angélica Carrillo Silva, Nicolý Soares de Almeida, Rosanna Maria Araújo Andrade Silva, Sérgio Ricardo Muniz e Fernando Fernandes Paiva: *Investigando a compreensão conceitual em física de alunos de graduação em cursos de ciências, engenharias e matemática*. Revista Brasileira de Ensino de Física, 41, novembro 2018, ISSN 1806-1117, 1806-9126. Citado na página 9.
- [5] Lopes, Antônio Martins: *Combinando Metodologia de Ensino Peer Instruction com Just-in-Time Teaching para o Ensino de Física*. Tese de Doutorado, Universidade Federal de Viçosa, janeiro 2016. Citado 2 vezes nas páginas 9 e 12.
- [6] Madsen, Adrian, Sarah B. McKagan e Eleanor C. Sayre: *Gender Gap on Concept Inventories in Physics: What Is Consistent, What Is Inconsistent, and What Factors Influence the Gap?* Physical Review Special Topics - Physics Education Research, 9(2):020121, novembro 2013. Citado 2 vezes nas páginas 9 e 15.
- [7] Ribeiro, Tais Bastani: *Análise quantitativa da aprendizagem conceitual e do gap de gênero em turmas de instrução por pares e tradicionais*, 2020. Citado na página 9.
- [8] Mazur, Eric: *Peer instruction: a revolução da aprendizagem ativa*. Penso Editora, 2015. Citado na página 10.

- [9] Crouch, Catherine H. e Eric Mazur: *Peer Instruction: Ten Years of Experience and Results*. American Journal of Physics, 69(9):970–977, setembro 2001, ISSN 0002-9505. Citado 2 vezes nas páginas 11 e 12.
- [10] Fagen, Adam P., Catherine H. Crouch e Eric Mazur: *Peer Instruction: Results from a Range of Classrooms*. The Physics Teacher, 40(4):206–209, abril 2002, ISSN 0031-921X. Citado na página 11.
- [11] Turpen, Chandra e Noah D. Finkelstein: *Not All Interactive Engagement Is the Same: Variations in Physics Professors' Implementation of Peer Instruction*. Physical Review Special Topics - Physics Education Research, 5(2):020101, agosto 2009. Citado na página 11.
- [12] Araujo, A. V. R. de, E. S. Silva, V. L. B. de Jesus e A. L. de Oliveira: *Uma associação do método Peer Instruction com circuitos elétricos em contextos de aprendizagem ativa*. Revista Brasileira de Ensino de Física, 39, novembro 2016, ISSN 1806-1117, 1806-9126. Citado na página 12.
- [13] Cid, Alberto Silva, Márcio Pizzi, Thiago Corrêa Lacerda e Erichardson Tarocco de Oliveira: *Proposta de Sequência Didática para Hidrostática: Aprendizagem Ativa em Destaque no Ensino de Física*. Caderno Brasileiro de Ensino de Física, 38(1):422–445, março 2021, ISSN 2175-7941. Citado na página 12.
- [14] Müller, Maykon Gonçalves, Ives Solano Araujo e Eliane Angela Veit: *Metodologias interativas de ensino na formação de professores de física: um estudo de caso com o método instrução pelos colegas (Peer Instruction)*. Alexandria: Revista de Educação em Ciência e Tecnologia, 10(2):171–195, novembro 2017, ISSN 1982-5153. Citado na página 12.
- [15] McDermott, Lillian C, Peter S Shaffer *et al.*: *Tutorials in introductory physics*. Prentice Hall Upper Saddle River, NJ, 2002. Citado na página 12.
- [16] Finkelstein, Noah D e Steven J Pollock: *Replicating and understanding successful innovations: Implementing tutorials in introductory physics*. Physical Review Special Topics-Physics Education Research, 1(1):010101, 2005. Citado na página 12.
- [17] Novak, Gregor M., Evelyn T. Patterson, Andrew D. Gavrin e Wolfgang Christian: *Just in Time Teaching*. American Journal of Physics, 67(10):937–938, outubro 1999, ISSN 0002-9505. Citado na página 12.
- [18] Araujo, Ives Solano e Eric Mazur: *Instrução pelos colegas e ensino sob medida: uma proposta para o engajamento dos alunos no processo de ensino-aprendizagem de Física*. Caderno Brasileiro de Ensino de Física, 30(2):362–384, abril 2013, ISSN 2175-7941. Citado na página 12.

- [19] Coqueiro, Naiara Porto da Silva e Erivan Coqueiro Sousa: *A educação a distância (EAD) e o ensino remoto emergencial (ERE) em tempos de Pandemia da Covid 19/ Distance education (Ed) and emergency remote education (ERE) in times of Pandemic Covid 19*. Brazilian Journal of Development, 7(7):66061–66075, julho 2021, ISSN 2525-8761. Citado na página 13.
- [20] White, Jonathon David: *Teaching General Physics in a COVID-19 Environment: Insights from Taiwan*. Physics Education, 55(6):065027, outubro 2020, ISSN 0031-9120. Citado na página 13.
- [21] Guo, Siming: *Synchronous versus Asynchronous Online Teaching of Physics during the COVID-19 Pandemic*. Physics Education, 55(6):065007, agosto 2020, ISSN 0031-9120. Citado na página 13.
- [22] Paula, Helder de Figueiredo e, Sérgio Luiz Talim, Cecília Siman Salema e Vinícius Reis Camillo: *ENGAJAMENTO DE ESTUDANTES EM UM ENSINO REMOTO E EMERGENCIAL DE FÍSICA*. Ensaio Pesquisa em Educação em Ciências (Belo Horizonte), 23, agosto 2021, ISSN 1983-2117. Citado na página 13.
- [23] Aguiar, Carlos Eduardo, Marcos Moura e Marta F. Barroso: *Ensino de física em tempos de pandemia: Instrução remota e desempenho acadêmico*. Revista Brasileira de Ensino de Física, 44, janeiro 2022, ISSN 1806-1117, 1806-9126. Citado na página 13.
- [24] *Coronavírus*. <https://ufmg.br/coronavirus>. Acesso: 2021-05-25. Citado na página 13.
- [25] Labudde, Peter, Walter Herzog, Markus P. Neuenschwander, Enrico Violi e Charlotte Gerber: *Girls and Physics: Teaching and Learning Strategies Tested by Classroom Interventions in Grade 11*. International Journal of Science Education, 22(2):143–157, fevereiro 2000, ISSN 0950-0693. Citado na página 15.
- [26] Lorenzo, Mercedes, Catherine H. Crouch e Eric Mazur: *Reducing the Gender Gap in the Physics Classroom*. American Journal of Physics, 74(2):118–122, fevereiro 2006, ISSN 0002-9505. Citado na página 15.
- [27] Pollock, Steven J., Noah D. Finkelstein e Lauren E. Kost: *Reducing the Gender Gap in the Physics Classroom: How Sufficient Is Interactive Engagement?* Physical Review Special Topics - Physics Education Research, 3(1):010107, junho 2007. Citado na página 15.
- [28] Freeman, Scott, Sarah L. Eddy, Miles McDonough, Michelle K. Smith, Nnadozie Okoroafor, Hannah Jordt e Mary Pat Wenderoth: *Active Learning Increases Student*

- Performance in Science, Engineering, and Mathematics*. Proceedings of the National Academy of Sciences, 111(23):8410–8415, junho 2014. Citado na página 16.
- [29] *Relatórios do setor de estatística*. <https://www.ufmg.br/prograd/relatorios/>. Acesso: 2021-05-25. Citado 3 vezes nas páginas 16, 31 e 33.
- [30] *Ementa das disciplinas do curso de Física*. <https://www.fisica.ufmg.br/graduacao/disciplinas-da-fisica/>. Acesso: 2021-05-25. Citado na página 17.
- [31] *Regras acadêmicas*. <https://ufmg.br/vida-academica/regras-academicas>. Acesso: 2021-05-25. Citado na página 18.
- [32] Hestenes, David, Malcolm Wells e Gregg Swackhamer: *Force Concept Inventory*. The Physics Teacher, 30(3):141–158, março 1992, ISSN 0031-921X. Citado na página 18.
- [33] Ramlo, Susan: *The Force and Motion Conceptual Evaluation*. outubro 2002. Citado na página 18.
- [34] Ding, Lin, Ruth Chabay, Bruce Sherwood e Robert Beichner: *Evaluating an Electricity and Magnetism Assessment Tool: Brief Electricity and Magnetism Assessment*. Physical Review Special Topics - Physics Education Research, 2(1):010105, março 2006. Citado 2 vezes nas páginas 18 e 26.
- [35] Thornton, Ronald K. e David R. Sokoloff: *Assessing Student Learning of Newton's Laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula*. American Journal of Physics, 66(4):338–352, abril 1998, ISSN 0002-9505. Citado na página 18.
- [36] Hake, Richard R.: *Interactive-Engagement versus Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses*. American Journal of Physics, 66(1):64–74, janeiro 1998, ISSN 0002-9505. Citado 2 vezes nas páginas 18 e 23.
- [37] Bao, Lei: *Theoretical Comparisons of Average Normalized Gain Calculations*. American Journal of Physics, 74(10):917–922, outubro 2006, ISSN 0002-9505. Citado na página 19.
- [38] Goulet-Pelletier, Jean Christophe e Denis Cousineau: *A Review of Effect Sizes and Their Confidence Intervals, Part I: The Cohen's d Family*. The Quantitative Methods for Psychology, 14:242–265, 2018, ISSN 2292-1354. Citado na página 20.
- [39] Cohen, Jacob: *Statistical power analysis for the behavioral sciences*. Academic press, 2013. Citado na página 20.

- [40] Nissen, Jayson M., Robert M. Talbot, Amreen Nasim Thompson e Ben Van Dusen: *Comparison of Normalized Gain and Cohen's d for Analyzing Gains on Concept Inventories*. Physical Review Physics Education Research, 14(1):010115, março 2018. Citado na página 20.
- [41] Coletta, Vincent P. e Jeffrey J. Steinert: *Why Normalized Gain Should Continue to Be Used in Analyzing Preinstruction and Postinstruction Scores on Concept Inventories*. Physical Review Physics Education Research, 16(1):010108, fevereiro 2020. Citado na página 20.
- [42] Chen, Yen Chi: *A Tutorial on Kernel Density Estimation and Recent Advances*. Biostatistics & Epidemiology, 1(1):161–187, janeiro 2017, ISSN 2470-9360. Citado na página 20.
- [43] Handcock, Mark S. e Martina Morris: *Relative Distribution Methods*. Sociological Methodology, 28(1):53–97, agosto 1998, ISSN 0081-1750. Citado na página 21.
- [44] Barros, J, Julie Remold, Glauco SF da Silva e José Roberto Tagliati: *Engajamento interativo no curso de Física I da UFJF*. Revista Brasileira de Ensino de Física, 26:63–69, 2004. Citado na página 23.
- [45] Müller, Maykon Gonçalves, Ives Solano Araujo, Eliane Angela Veit e Julie Schell: *Uma revisão da literatura acerca da implementação da metodologia interativa de ensino Peer Instruction (1991 a 2015)*. Revista Brasileira de Ensino de Física, 39, março 2017, ISSN 1806-1117, 1806-9126. Citado na página 44.

## APÊNDICE A — TABELAS

Tabela 4 – Média da nota final, pré-teste, pós-teste e ganho normalizado para o métodos tradicional (T), *Peer Instruction* (PI) e Ensino Remoto Emergencial (ERE) separados por gênero masculino (Masc.) e feminino (Fem.). Número de respondentes (N), média (m) e desvio padrão ( $\sigma$ ) da nota final e ganho normalizado ( $g$ ) e D de Cohen ( $d$ ) de estudantes que completaram o pré e o pós teste.

Método	Gênero	N	Nota final		Pré-teste		Pós-teste		$g$		$d$
			m	$\sigma$	m	$\sigma$	m	$\sigma$	m	$\sigma$	
T	Masc.	237	70,0	19,6	9,78	5,09	14,73	5,57	0,239	0,242	0,932
	Fem.	116	68,1	20,3	7,83	3,53	11,48	4,87	0,166	0,177	0,864
	Ambos	353	69,4	19,8	9,14	4,72	13,66	5,55	0,215	0,225	0,881
PI	Masc.	141	70,2	20,2	8,94	4,53	15,47	5,88	0,318	0,229	1,248
	Fem.	104	69,0	20,7	7,58	3,82	12,87	5,50	0,237	0,226	1,123
	Ambos	245	67,7	20,4	8,36	4,29	14,36	5,86	0,283	0,231	1,172
ERE	Masc.	612	82,2	13,9	9,30	4,62	14,10	6,35	0,233	0,279	0,866
	Fem.	301	82,8	13,8	7,36	3,18	10,82	5,41	0,150	0,229	0,782
	Ambos	917	82,4	13,9	8,655	4,307	13,01	6,25	0,205	0,224	0,812

Tabela 5 – Média da nota final, pré-teste, pós-teste e ganho normalizado para os Ensino Remoto Emergencial, separados por semestre e gênero masculino (Masc.) e feminino (Fem.). Número de respondentes (N), média (m) e desvio padrão ( $\sigma$ ) da nota final e ganho normalizado ( $g$ ) e D de Cohen ( $d$ ) de estudantes que completaram o pré e o pós teste.

Método	Gênero	N	Nota final		Pré-teste		Pós-teste		$g$		$d$
			m	$\sigma$	m	$\sigma$	m	$\sigma$	m	$\sigma$	
2020/1	Masc.	111	88,6	8,3	9,43	5,20	15,15	6,92	0,270	0,351	0,938
	Fem.	63	91,6	6,2	6,54	2,84	11,30	5,09	0,200	0,214	1,170
	Ambos	174	89,7	7,7	8,38	4,69	13,76	6,57	0,245	0,310	0,943
2020/2	Masc.	189	81,9	13,0	9,00	4,03	14,49	5,92	0,258	0,276	1,107
	Fem.	112	83,9	13,0	7,07	3,18	11,23	3,35	0,181	0,215	0,950
	Ambos	301	82,7	13,0	9,23	3,88	13,31	5,91	0,230	0,256	1,017
2021/1	Masc.	202	81,4	11,9	9,84	4,76	14,22	6,47	0,231	0,262	0,772
	Fem.	70	78,2	12,3	8,13	3,55	11,75	5,95	0,164	0,248	0,746
	Ambos	272	80,5	12,1	9,38	4,53	13,52	6,44	0,211	0,260	0,745
2021/2	Masc.	110	77,3	20,0	8,85	4,63	12,14	5,88	0,158	0,217	0,625
	Fem.	56	76,7	17,6	7,88	2,82	8,29	4,48	0,012	0,203	0,111
	Ambos	166	77,1	19,2	8,52	4,13	10,84	5,73	0,109	0,222	0,466