

UNIVERSIDADE FEDERAL DE ALAGOAS  
CAMPUS A. C. SIMÕES  
INSTITUTO DE COMPUTAÇÃO  
SISTEMAS DE INFORMAÇÃO

LEONARDO ULISSES LIMA DE HOLANDA

**R-PARTNER:** Um sistema para apoio à realização de experimentos com design do tipo pré-teste e pós-teste em ambientes online de aprendizagem

Maceió, Alagoas

2022

LEONARDO ULISSES LIMA DE HOLANDA

**R-PARTNER:** Um sistema para apoio à realização de experimentos com design do tipo pré-teste e pós-teste em ambientes online de aprendizagem

Trabalho de Conclusão de Curso apresentado ao curso de Sistemas de Informação da Universidade Federal de Alagoas, como requisito parcial à obtenção do título de Bacharelado em Sistemas de Informação.

Orientador: Prof. Dr. Ranilson Paiva.

Maceió, Alagoas

2022

**Catálogo na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**  
**Divisão de Tratamento Técnico**

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 - 1767

H722r Holanda, Leonardo Ulisses Lima de.  
R-PARTNER : um sistema para apoio à realização de experimentos com design do tipo pré-teste e pós-teste em ambientes *online* de aprendizagem / Leonardo Ulisses Lima de Holanda. – 2022.  
80 f. : il.

Orientador: Ranilson Paiva.  
Monografia (Trabalho de conclusão de curso em Sistemas de Informação) – Universidade Aberta do Brasil. Universidade Federal de Alagoas, Instituto de Computação. Maceió, 2022.

Bibliografia: f. 78-80.

1. Experimentos científicos. 2. Práticas baseadas em evidências. 3. Pesquisa educacional - *Software*. I. Título.

CDU: 004.4:37



**UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL**  
**Instituto de Computação – IC**

Campus A. C. Simões - Av. Lourival de Melo Mota, BL 12  
Tabuleiro do Martins, Maceió/AL - CEP: 57.072-970  
Telefone: (082) 3214-1401



## Trabalho de Conclusão de Curso – TCC

### Formulário de Avaliação

Curso: \_\_\_\_\_

Nome do Aluno

L	e	o	n	a	r	d	o		U	l	i	s	s	e	s				
L	i	m	a		d	e			H	o	l	a	n	d	a				

Nº de Matrícula

1	7	2	1	0	2	3	2					-	
---	---	---	---	---	---	---	---	--	--	--	--	---	--

Título do TCC (Tema)

R-PARTNER: Um sistema para apoio à realização de experimentos com design do tipo pré-teste e pós-teste em ambientes online de aprendizagem.

Banca Examinadora

Ranilson Oscar Araújo Paiva  
Nome do Orientador

\_\_\_\_\_

Assinatura

Dr. Rafael de Amorim Silva  
Nome do Professor

\_\_\_\_\_

Assinatura

André Luís Alice Raabe  
Nome do Professor

\_\_\_\_\_

Assinatura

Data da Defesa

19/ 12/ 2022

Nota Obtida

10,0 (Dez)

Observações

O estudante deverá realizar os ajustes propostos pelos professores avaliadores.

**Coordenador do Curso**  
**De Acordo**

\_\_\_\_\_

Assinatura

Dedico este trabalho a Deus e a todos que persistem em fazer ciência apesar das crescentes dificuldades para isso em nosso país.

## **AGRADECIMENTOS**

Agradeço a Deus por todas as bênçãos em minha vida que me guiam e me dão força para perseverar.

Agradeço aos meus pais, Ulisses e Fátima, por todo o apoio e amor durante toda minha vida, por estarem sempre ao meu lado e me ajudarem nessa difícil jornada da graduação.

Agradeço ao meu orientador, professor Ranilson Paiva, pela paciência e por sua orientação de suma importância para a realização deste trabalho, bem como seu refinado senso de humor que tornou essa jornada muito mais leve.

Agradeço aos companheiros de equipe Márcio Henrique, João Victor Falcão e Michael Miller, que participaram do desenvolvimento do sistema proposto no presente trabalho, pela ajuda valiosa e pelo desvendar dos mistérios ocultos de Laravel e PHP.

Agradeço aos amigos Otávio, Joclean, Handrik, Laudemi e Geraldo pela amizade e pelo tempo compartilhado durante todos esses anos. Que nossos caminhos se encontrem uma vez mais!

Agradeço ao corpo docente do curso de Sistemas de Informação por todos os ensinamentos e pelo empenho durante esses anos de graduação mesmo com todas as dificuldades existentes.

Agradeço a todos que participaram do experimento realizado no presente trabalho por suas valiosas percepções e sugestões.

Agradeço, por fim, a todos aqueles que contribuíram para a realização deste presente trabalho.

## RESUMO

O uso de práticas educacionais baseadas em evidência vem ganhando relevância considerável em países no exterior. Essa tendência vem acompanhada do desenvolvimento de um aparato de pesquisa responsável por gerar e avaliar evidências científicas. Para a geração de evidências, um dos métodos mais convincentes são os experimentos controlados. Entretanto, tais experimentos são caros e difíceis de realizar. No contexto brasileiro, a pesquisa educacional que utiliza experimentos controlados é pouco expressiva, o que enfraquece a geração de evidências capazes de basear novas práticas educacionais. Neste cenário, o presente trabalho propõe um sistema capaz de apoiar a realização de experimentos com design do tipo pré-teste e pós-teste em ambientes online de aprendizagem, buscando oferecer melhores condições para a realização destes experimentos. Um experimento foi conduzido para avaliar a percepção dos usuários sobre o sistema. Os resultados foram analisados estatisticamente e sugerem percepções positivas dos usuários em relação às métricas de facilidade de uso, utilidade e baixo custo de tempo do sistema. Conclui-se que o sistema é capaz de apoiar a realização de experimentos com design do tipo pré-teste e pós-teste.

**Palavras-chave:** Experimentos; Práticas Baseadas em Evidências; Software de Pesquisa Educacional; Design Pré-teste e Pós-teste;

## **ABSTRACT**

The use of evidence-based practices in education is gaining traction in foreign countries. This trend has been accompanied by the development of a research apparatus responsible for generating and evaluating scientific evidence. To generate evidence, one of the most convincing methods is controlled experiments. However, such experiments are expensive and difficult to carry out. In the Brazilian context, educational research that uses controlled experiments is modest, which weakens the generation of evidence on which to base new educational practices. In this scenario, the present work proposes a system to manage experiments with pretest and posttest designs in online learning environments, seeking to offer better conditions for conducting these experiments. An experiment was conducted to evaluate the user's perception of the system. The results were statistically analyzed and suggest positive user perceptions regarding the system's metrics of ease of use, usefulness, and low time cost. It is concluded that the system is capable of supporting pre-test and post-test design experiments.

**Keywords:** Experiments; Evidence-based practices; Educational Research Software; Pretest-Posttest Design.

## LISTA DE FIGURAS

Figura 1	- Recorte do trecho da etapa de criação do processo.....	36
Figura 2	- Recorte do trecho da etapa de execução do processo.....	37
Figura 3	- Imagem do processo completo.....	38
Figura 4	- Menu principal do sistema.....	39
Figura 5	- Etapa “Nome e descrição”, primeira na criação do experimento.....	40
Figura 6	- Etapa “Escolha de turmas”, segunda na criação do experimento.....	41
Figura 7	- Etapa “Convite e permissão”, terceira na criação do experimento....	42
Figura 8	- Etapa “Prazo das respostas”, quarta na criação do experimento.....	43
Figura 9	- Etapa “Criação do pré-teste”, quinta na criação do experimento.....	44
Figura 10	- Etapa “Criação da intervenção”, sexta na criação do experimento...	45
Figura 11	- Etapa “Criação do pós-teste”, sétima na criação do experimento.....	46
Figura 12	- Etapa “Atribuição randômica”, oitava na criação do experimento de nome.....	48
Figura 13	- Etapa de resultados do experimento.....	50
Figura 14	- Início do relatório estatístico.....	51
Figura 15	- Continuação do primeiro capítulo do relatório estatístico.....	52
Figura 16	- Capítulo de premissas estatísticos do relatório.....	54
Figura 17	- Capítulo de escolha do teste estatístico do relatório.....	55
Figura 18	- Continuação do capítulo de escolha do teste estatístico do relatório.....	56
Figura 19	- Capítulos finais do relatório.....	57

## LISTA DE GRÁFICOS

Gráfico 1	- Respostas por métrica acumulada dos questionários das etapas	71
Gráfico 2	- Resultados por pergunta do questionário de comparação.....	73
Gráfico 3	- Resultados por métrica do questionário de comparação.....	74

## LISTA DE TABELAS

Tabela 1	- Comparação dos trabalhos relacionados com o presente trabalho.....	31
Tabela 2	- Atividades por etapa do questionário do experimento.....	58
Tabela 3	- Afirmações utilizadas nas questões por etapa onde n equivale a qualquer uma das 8 etapas da Tabela 1.....	59
Tabela 4	- Afirmações utilizadas no conjunto final de questões.....	60
Tabela 5	- Dados demográficos dos participantes.....	66
Tabela 6	- Medida do coeficiente de Alfa de Cronbach em relação às métricas avaliadas.....	68
Tabela 7	- Resultados do questionário por etapa.....	69
Tabela 8	- Resultados do Teste de Shapiro-Wilk por métrica.....	75
Tabela 9	- Resultados dos Testes de Mann-Whitney U das métricas.....	76

## LISTA DE ABREVIATURAS E SIGLAS

IES	Institute of Education Sciences
WCC	What Works Clearinghouse
PaaS	Platform-as-a-service
API	Application Programming Interface

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>15</b>
<b>1.1</b>	<b>Contextualização</b> .....	<b>15</b>
<b>1.2</b>	<b>Motivação</b> .....	<b>17</b>
<b>1.3</b>	<b>Objetivos</b> .....	<b>18</b>
<b>1.4</b>	<b>Estrutura do trabalho</b> .....	<b>19</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>20</b>
<b>2.1</b>	<b>Método científico</b> .....	<b>20</b>
<b>2.2</b>	<b>Ensaio Clínico Randomizado (ECR)</b> .....	<b>20</b>
<b>2.3</b>	<b>Experimento do tipo pré-teste e pós-teste</b> .....	<b>21</b>
<b>2.4</b>	<b>Teste A/B</b> .....	<b>22</b>
<b>2.5</b>	<b>Technology Acceptance Model (TAM)</b> .....	<b>23</b>
<b>3</b>	<b>TRABALHOS RELACIONADOS</b> .....	<b>25</b>
<b>3.1</b>	<b>The Assistments Ecosystem: Building A Platform That Brings Scientists And Teachers Together For Minimally Invasive Research On Human Learning And Teaching</b> .....	<b>25</b>
<b>3.2</b>	<b>Lioness Lab: A Free Web-Based Platform For Conducting Interactive Experiments Online</b> .....	<b>26</b>
<b>3.3</b>	<b>Psiturk: An Open-Source Framework For Conducting Replicable Behavioral Experiments Online</b> .....	<b>27</b>
<b>3.4</b>	<b>Upgrade: An Open Source Tool To Support A/B Testing In Educational Software</b> .....	<b>28</b>
<b>3.5</b>	<b>Terracotta - Tool For Educational Research With Randomized Controlled Trials</b> .....	<b>29</b>
<b>3.6</b>	<b>SES: Um Sistema Baseado na Web para o Design e a Condução de Experimentos Educacionais</b> .....	<b>30</b>
<b>3.7</b>	<b>Comparação dos trabalhos relacionados com a proposta do presente trabalho</b> .....	<b>30</b>
<b>4</b>	<b>PROPOSTA</b> .....	<b>33</b>
<b>4.1</b>	<b>Integração com o T-Partner e S-Partner</b> .....	<b>34</b>

<b>4.2</b>	<b>Sobre o processo</b> .....	35
4.2.1	Etapa de criação.....	35
4.2.2	Etapa de execução.....	36
<b>4.3</b>	<b>Detalhamento do processo</b> .....	39
4.3.1	Menu Principal.....	39
4.3.2	Criar experimento e atribuir nome e detalhes.....	40
4.3.3	Selecionar turmas por curso.....	40
4.3.4	Escrever pedido de permissão e convites.....	41
4.3.5	Criar pré-teste e definir seu prazo.....	43
4.3.6	Criar intervenção e definir seu prazo.....	44
4.3.7	Criar pós-teste e definir seu prazo.....	45
4.3.8	Selecionar método de atribuição randômica.....	46
4.3.9	Enviar pedidos de permissão aos professores.....	48
4.3.10	Enviar convites aos alunos de turmas permitidas.....	48
4.3.11	Atribuir alunos aos grupos de pesquisa.....	48
4.3.12	Requisitar cumprimento do pré-teste.....	49
4.3.13	Requisitar cumprimento da intervenção.....	49
4.3.14	Requisitar cumprimento do pós-teste.....	49
4.3.15	Produzir relatório estatístico.....	49
<b>5</b>	<b>MATERIAIS E MÉTODOS</b> .....	58
<b>5.1</b>	<b>Objetivo do experimento</b> .....	58
<b>5.2</b>	<b>Instrumento de coleta de dados</b> .....	58
<b>5.3</b>	<b>Materiais</b> .....	62
<b>5.4</b>	<b>Participantes</b> .....	62
<b>5.5</b>	<b>Hipóteses</b> .....	62
<b>6</b>	<b>RESULTADOS E DISCUSSÃO</b> .....	66
<b>6.1</b>	<b>Análise das participações</b> .....	66
<b>6.2</b>	<b>Análise de confiabilidade</b> .....	67
<b>6.3</b>	<b>Análise descritiva dos resultados das etapas</b> .....	68
<b>6.4</b>	<b>Análise descritiva por métrica acumulada</b> .....	71
<b>6.5</b>	<b>Resultados do questionário de comparação</b> .....	72
<b>6.6</b>	<b>Análise inferencial dos resultados</b> .....	74
6.6.1	Análise de normalidade.....	74

6.6.2	Correlações.....	75
6.6.3	Teste de Hipóteses.....	76
<b>7</b>	<b>CONCLUSÃO.....</b>	<b>77</b>
<b>7.1</b>	<b>Limitações e ameaças à validade.....</b>	<b>77</b>
<b>7.2</b>	<b>Trabalhos futuros.....</b>	<b>78</b>
	<b>REFERÊNCIAS .....</b>	<b>79</b>

# 1 INTRODUÇÃO

## 1.1 Contextualização

É possível perceber, nas últimas décadas, um aumento da publicação de artigos de pesquisa educacional com métodos experimentais ou quasi-experimentais em países estrangeiros. Uma revisão sistemática conduzida por (CONNOLLY; KEENAN; URBANSKA, 2018) identificou 1017 experimentos, do tipo randomizado controlado, realizados entre 1980 e 2016 em terras majoritariamente norte-americanas e europeias, onde a maior parte ocorreu entre 2007 e 2016. (GOPALAN; ROSINGER; AHN, 2020, p. 221), por exemplo, se refere ao crescimento de artigos de pesquisa quasi-experimental como “exponencial”. O autor afirma, de acordo com os resultados de sua pesquisa sistemática, que o número de artigos publicados anualmente passou de menos de 20 em 2008 para mais de 100 em 2018, num conjunto de 15 respeitados jornais estadunidenses de pesquisa educacional.

Para (GOPALAN; ROSINGER; AHN, 2020), um dos pontos que possivelmente explica esse crescente interesse, no contexto estadunidense, é uma maior importância dada pelo governo às evidências científicas nas práticas educacionais. Como exemplo, é citada uma lei, promulgada em 2015, chamada *Every Student Succeeds Act* que, segundo o autor, faz recomendações de seleção e uso de práticas baseadas em evidências em instituições educacionais. Outro exemplo dado pelo autor ocorre na concessão de investimentos para pesquisas através de instituições governamentais, como o *Institute of Education Sciences* do Departamento de Educação dos Estados Unidos, cuja aprovação está condicionada à presença, nas pesquisas, de evidências que atendem padrões rigorosos da *What Works Clearinghouse*, uma iniciativa do IES responsável pela avaliação de estudos.

O interesse governamental por tais evidências não é infundado. (SLAVIN, 2020) afirma a importância de tais políticas ao citar que o desenvolvimento visto em outros setores da sociedade que utilizam práticas baseadas em evidências, como medicina, por exemplo, poderia ocorrer também na educação. (SLAVIN, 2020, p. 11, tradução nossa) justifica tal possibilidade afirmando que “soluções que atendem a rigorosas normas de evidência substituem os produtos menos eficazes”. (MURPHY et al., 2020), por exemplo, realizou um experimento randomizado controlado para

avaliar os impactos de uma ferramenta de aprendizado online, chamada ASSISTments, que auxilia professores e alunos do ensino primário e secundário nos EUA (HEFFERNAN, N.; HEFFERNAN, C., 2014). Os resultados encontrados por (MURPHY et al., 2020) indicaram que escolas que utilizaram a ferramenta ASSISTments obtiveram resultados significativamente melhores em testes de matemática que escolas que utilizavam práticas já existentes. Por conta desse estudo, a ferramenta recebeu o maior grau de robustez de evidência pela WWC (WHAT WORKS CLEARINGHOUSE, 2019), sendo atualmente utilizada por mais de 1 milhão de alunos e 30 mil professores em mais de 20 países (ASSISTMENTS FOUNDATION; WORCESTER POLYTECHNIC INSTITUTE, 2013).

Para alcançar as evidências rigorosas o suficiente para basear as práticas educacionais, os métodos de pesquisa experimentais se apresentam como uma opção convincente, apesar de não ser a única. A lei *Every Student Succeeds Act*, como cita (SLAVIN, 2020, p. 2, tradução nossa), formula 3 níveis associados à robustez de uma evidência, onde o maior nível somente pode ser obtido “através de pelo menos um estudo experimental bem projetado e bem implementado”, e o nível abaixo deste por “pelo menos um estudo quasi-experimental bem projetado e bem implementado”. Essa distinção também é percebida no documento *Procedures and Standards Handbook* que detalha os padrões mais recentes utilizados pela WWC para avaliação de estudos (WHAT WORKS CLEARINGHOUSE, 2022).

Entretanto, há de se notar que existem dificuldades significativas que podem inviabilizar a realização de tais experimentos. (GOPALAN; ROSINGER; AHN, 2020; SULLIVAN, 2011; SLAVIN, 2002; SCHANZENBACH, 2012) citam os custos financeiros como exemplo, talvez o maior deles. (SCHANZENBACH, 2012, p. 225, tradução nossa) alerta para a “necessidade de mais experimentação em uma maior variedade de ambientes”, através da replicação de experimentos, o que tornaria o processo ainda mais custoso.

No contexto brasileiro, (SANDIM; MACHADO, 2020, p.45) avaliou, com base nos resultados de uma busca bibliográfica, que a produção acadêmica sobre políticas públicas baseadas em evidência, ou PPBEs, é “pouco expressiva”. Isto não se limitando somente às políticas educacionais, pois estas, de acordo com o trabalho do autor, são ainda menos expressivas. Segundo (FARIA; SANCHES, 2022), que realizou mapeamentos de teses, dissertações e artigos científicos

publicados no Brasil nas últimas décadas, há um crescente interesse por parte dos pesquisadores brasileiros pelo movimento das políticas públicas baseadas em evidências, apesar de tal interesse ainda ser tímido.

Apesar da produção acadêmica diminuta em relação à de outros países, algumas iniciativas vêm sendo tomadas para explorar o uso de evidências em políticas educacionais. (SANDIM; MACHADO, 2020, p. 41) cita, como um “exemplo de estratégia pioneira”, a criação da Assessoria Estratégica de Evidências pelo Ministério da Educação em 2018, entre outras iniciativas em diferentes áreas. Além disso, como apontado por (FARIA; SANCHES, 2022), há também a criação do Escritório de Evidências em 2020 pela Secretaria de Educação de São Paulo. É perceptível que há um esforço sendo feito para incentivar o uso de evidências em políticas educacionais no Brasil.

## **1.2 Motivação**

O uso de práticas educacionais baseadas em evidências se mostra bastante desenvolvido em países como os Estados Unidos. A criação de todo um aparato em volta da necessidade de pesquisar e avaliar evidências e o incentivo para adoção das práticas baseadas em tais evidências, que trazem benefícios comprovados para os estudantes, chama a atenção. Entretanto, essa evolução só se mostrou possível graças a uma melhora na produção científica por parte dos pesquisadores educacionais, tanto em quantidade quanto em qualidade, além do interesse e apoio governamental.

Para a geração de evidências capazes de basear as práticas educacionais também no contexto brasileiro, é necessária uma produção acadêmica constante e de qualidade. Entretanto, como demonstrado num mapeamento sistemático realizado por (MAGALHÃES et al., 2013), apenas 8,7% dos 835 estudos publicados entre 2001 e 2012 no Simpósio Brasileiro de Informática na Educação foram identificados como “experimentos formais ou quasi-experimentos”, sendo estes considerados métodos convincentes para a geração de evidências.

Sabendo das conhecidas dificuldades que inviabilizam a realização de experimentos, há de se constituir então um cenário onde a geração de evidências é enfraquecida. Por consequência, a adoção de práticas educacionais baseadas em evidências também se enfraquece pela ausência de evidências robustas. E há de se

perder, enfim, os possíveis benefícios que tais práticas trariam para os estudantes a quem elas se aplicariam.

Entretanto, é possível perceber, no contexto da pesquisa comportamental, que a tecnologia se mostrou uma ferramenta valiosa para suprir a demanda por experimentos com auxílio da *Internet* e de softwares específicos para este fim. Como mostra (WOLFE, 2017), várias vantagens associadas a experimentos baseados na web no contexto da pesquisa comportamental se mostraram válidas mesmo após décadas de suas conjecturas. O autor cita o acesso facilitado aos participantes sem barreiras geográficas, além da facilidade de participação do experimento, através até mesmo de um *smartphone*, a possibilidade de obter amostras de tamanho grande o suficiente para conferir significância estatística e a redução de custos por dispensar uso de infraestrutura física.

Sendo assim, questiona-se se uma ferramenta tecnológica não poderia também desempenhar um papel ativo no cenário de pesquisa educacional com experimentos, apresentando então uma oportunidade de propor uma ferramenta capaz de facilitar a realização de experimentos por parte dos pesquisadores educacionais.

### **1.3 Objetivos**

Nesse cenário de baixa produção científica para geração de evidências capazes de basear políticas educacionais e de conhecidas dificuldades na realização de experimentos capazes de produzir tais evidências, o presente trabalho tem como objetivo propor e desenvolver um sistema baseado na web que visa apoiar a realização de experimentos com design do tipo pré-teste e pós-teste em ambientes online de aprendizagem de forma a facilitar o trabalho do pesquisador para este fim, proporcionando, possivelmente, melhores condições e menos impedimentos para a realização destes experimentos.

Além disso, também se faz necessário avaliar a percepção dos usuários em relação ao sistema de forma a compreender se a solução proposta é adequada para o que se propõe fazer. Com esse objetivo, será conduzido um experimento onde os participantes irão, através do sistema, simular a criação e condução de um experimento, compartilhando suas percepções através de questionários. Dessa forma, espera-se obter dados que, ao serem analisados estatisticamente, permitirão

uma melhor compreensão da percepção dos participantes em relação ao sistema. Isso nos ajudará, portanto, a validar a proposta do presente trabalho, direcionando o desenvolvimento futuro do sistema para o caminho que melhor atender os pesquisadores.

#### **1.4 Estrutura do trabalho**

Os capítulos restantes que compõem a estrutura do trabalho são: fundamentação teórica, trabalhos relacionados, proposta, materiais e métodos, resultados e discussão e, por fim, conclusão. No capítulo de fundamentação teórica, os principais conceitos do presente trabalho são apresentados de forma a criar um embasamento teórico a ser consultado. No capítulo de trabalhos relacionados, listam-se trabalhos que, de alguma forma, se relacionam com o presente trabalho, comparando sua proposta com a de tais trabalhos. No capítulo de proposta, é detalhado a concepção e funcionamento do sistema, fornecendo informações sobre o processo de criação e condução de experimentos e verificação do relatório estatístico gerado a partir de seus resultados. No capítulo de materiais e métodos, são detalhados o objetivo do experimento, bem como o instrumento de coleta de dados utilizado, seu público-alvo de participantes e quais materiais foram utilizados. No capítulo de resultados e discussão, os dados obtidos através do experimento são analisados com uso de estatística descritiva e inferencial. Por fim, no último capítulo, são apresentadas as conclusões alcançadas pelo presente trabalho, bem como suas limitações e ameaças, além de indicar possíveis trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo de fundamentação teórica, serão tratados os principais conceitos que se relacionam com o tema do presente trabalho, possibilitando a consulta facilitada de informações sobre os mesmos por parte do leitor para complementar eventuais lacunas.

### 2.1 Método Científico

Uma definição básica do método científico é dada por (CAREY, 2011, p. 3, tradução nossa) como “um processo simples e de três etapas através do qual os cientistas investigam a natureza”. Segundo o autor, a primeira delas é a observação da natureza, onde é possível revelar aspectos dela cuja compreensão não é clara. Então, sucede-se a etapa de explicação com o objetivo de esclarecer o aspecto não compreendido através de especulações. Por fim, a etapa de testes, onde cria-se maneiras de testar o que se especula.

Um passo a passo mais detalhado do método científico, apesar de ser algo amplamente difundido na *Internet*, não aparenta ser um consenso na comunidade acadêmica. (MCCOMAS, 2002, p. 57, tradução nossa) detalha esse passo a passo da seguinte forma: “Definir o problema, buscar informações, formular uma hipótese, fazer observações relevantes, testar a hipótese, formular conclusões e comunicar os resultados.” Entretanto, (MCCOMAS, 2002, p. 58, tradução nossa) discorda desse passo a passo, afirmando, com base em estudos de outros autores, que “nenhum método de pesquisa é aplicado universalmente”.

### 2.2 Estudo Clínico Randomizado (ECR)

Segundo (KENDALL, 2003), um *Randomized Controlled Trial*, ou Ensaio Clínico Randomizado, é um experimento onde os participantes são randomicamente atribuídos em dois grupos: o grupo experimental, onde os indivíduos serão submetidos a uma intervenção experimental, e o grupo de controle, onde os indivíduos receberão uma intervenção padrão. De acordo com o autor, ambos os grupos serão acompanhados de forma que as diferenças nos resultados das

intervenções sejam detectadas. Os resultados e a análise feita em cima deles são utilizados para avaliar a efetividade da intervenção experimental em questão.

Além de definir, (KENDALL, 2003) ainda detalha os passos necessários para conduzir um ECR. Primeiramente, se faz necessário definir uma hipótese a ser testada, além de um protocolo que documenta a preparação do experimento, detalhando de que forma será executada cada etapa do mesmo para garantir que a hipótese será verdadeiramente testada. Posteriormente, deve-se escolher a população ideal para o experimento, definindo que tipo de indivíduo participará e como se dará seu recrutamento. O autor cita também a importância de definir os critérios de inclusão e exclusão dos indivíduos da amostra do estudo para que os resultados obtidos pelo experimento possam ser aplicados a uma população geral. Depois, é importante atribuir os indivíduos de forma randômica aos grupos do experimento, além de garantir que as variáveis dos indivíduos, como idade e sexo, por exemplo, estejam distribuídas igualmente entre os grupos. Por fim, o grupo experimental receberá a intervenção experimental, e o grupo de controle receberá a intervenção padrão. Os grupos então serão acompanhados de forma que os efeitos da intervenção experimental possam ser avaliados através da análise dos resultados obtidos no experimento. O autor, nesta etapa, cita a importância de que os participantes de ambos os grupos desconheçam a qual grupo pertencem e, da mesma forma, que o indivíduo responsável pela análise dos resultados também desconheça esse fator.

Para (SIBBALD; ROLAND, 1998), um ensaio clínico randomizado é a forma mais rigorosa de determinar uma relação de causa e efeito entre uma intervenção e o resultado. O autor também cita que outros projetos de estudos podem detectar associações entre uma intervenção e um resultado, mas não conseguem descartar a possibilidade de que a associação tenha sido causada por um fator externo.

### **2.3 Experimento do tipo pré-teste e pós-teste**

Uma das formas de avaliar os efeitos de uma intervenção através de um experimento controlado é utilizando o pré-teste e o pós-teste. Neste cenário, há a presença de pelo menos dois grupos de pesquisa, o experimental e o controle. Tal método consiste, para ambos os grupos, na aplicação do pré-teste, seguido pela aplicação de intervenções e, posteriormente, a aplicação do pós-teste. Compara-se

então os resultados do grupo experimental com o grupo controle de forma a identificar diferenças significativas. Como o experimento é planejado para que a única diferença entre os dois grupos seja a intervenção a eles aplicada, é possível concluir que a intervenção causou as diferenças identificadas (LEVY; ELLIS, 2011).

(STRATTON, 2019, p. 573, tradução nossa) cita tal método como sendo “rápido e conveniente para avaliar um grupo no qual uma intervenção foi aplicada”. Segundo o autor, esse tipo de método possui a vantagem de atribuir direcionalidade na pesquisa de forma que apenas uma variável seja testada antes e depois de uma intervenção.

O autor ainda detalha métodos utilizados para melhorar a validade dos estudos realizados com pré-teste e pós-teste. Para ele, a atribuição randômica dos participantes nos grupos do experimento é essencial. Além disso, também é sugerido a realização de um pós-teste tardio para acompanhamento dos efeitos da intervenção a longo prazo.

## **2.4 Teste A/B**

Segundo (SIROKER; KOOMEN, 2013, p. 8, tradução nossa), um de teste A/B pode ser descrito como “mostrar diferentes variações de seu website para diferentes pessoas e medir qual variação é a mais eficaz em virar transformar pessoas em clientes”. Para o autor, a ideia é a mesma de um ensaio controlado, onde as variações representam um papel semelhante aos grupos de pesquisa, nos quais os usuários são atribuídos randômicamente.

Ao realizar o teste A/B, há de se definir qual versão será considerada a versão A, normalmente atribuída a versão original, e qual versão será considerada a versão B, normalmente atribuída a versão experimental. Há várias possibilidades que podem diferenciar uma versão da outra como, por exemplo, diferenças na interface de uma página web ou diferentes algoritmos de recomendação de ofertas. Quando um usuário acessar o serviço em que o teste A/B ocorre, ele será randomicamente atribuído a uma das duas versões. Durante o uso do serviço pelo usuário, métricas de interesse dos pesquisadores, como tempo gasto pelo usuário, ou compras realizadas, serão coletadas e posteriormente analisadas de forma a apontar qual das duas versões se saiu melhor em relação à métrica avaliada (KOHAVI; THOME, 2017).

(KOHAVI; THOME, 2017) cita vários exemplos de testes A/B que foram realizados e trouxeram descobertas valiosas para grandes empresas como Microsoft e Amazon, onde mudanças na interface ou no comportamento das páginas, como abrir links em abas novas no navegador ao contrário de abrir na mesma aba ou o local na interface de uma oferta especial, por exemplo, são capazes de melhorar o engajamento dos usuários e os lucros financeiros das empresas.

(KOHAVI et al., 2020) reforça a relevância dos testes A/B ao citar que seu uso é feito de forma constante por grandes empresas de tecnologia como Google e Facebook, utilizando-se de sua imensa base de usuários para testar versões experimentais de interfaces, algoritmos de busca e anúncios e demais tecnologias.

## **2.5 Technology Acceptance Model (TAM)**

*Technology Acceptance Model*, ou Modelo de Aceitação Tecnológica, é um modelo proposto por Fred D. Davis para explicar o que leva um usuário a aceitar ou rejeitar um sistema (DAVIS, 1989). O autor propõe que há dois fatores determinantes para a aceitação ou rejeição do sistema por parte do usuário: Facilidade de Uso Percebida e Utilidade Percebida. Define-se ambos os conceitos da seguinte forma: Facilidade de Uso Percebida como “o grau em que uma pessoa acredita que a utilização de um determinado sistema seria livre de esforço” e Utilidade Percebida como “o grau em que uma pessoa acredita que a utilização de um determinado sistema melhoraria seu desempenho do trabalho” (DAVIS, 1989, p. 320, tradução nossa).

Como forma de avaliar ambos os fatores, (DAVIS, 1989) propõe uso de duas escalas psicométricas de 6 itens. A primeira escala trata da Utilidade Percebida, contendo os itens: Maior Velocidade de trabalho, Desempenho de Trabalho, Aumento de Produtividade, Eficácia, Facilidade de Trabalho e Utilidade. Já para a segunda escala que trata de Facilidade de Uso Percebida, os itens são: Facilidade de aprender, Controlável, Limpo e compreensível, Flexível, Fácil de se tornar habilidoso e Fácil de usar. Cada item é usado para formular uma afirmação na qual o indivíduo deve atribuir uma nota de 1 a 7, onde cada número representa um grau de concordância ou discordância em relação à afirmação.

De acordo com (LEE; KOZAR; LARSEN, 2003), a presença numerosa do TAM em estudos científicos na área de Sistemas de Informação lhe garante uma

popularidade significativa. Segundo o autor, a ampla aplicação do modelo numa diversidade de tecnologias e situações lhe confere relevância no meio acadêmico.

### 3 TRABALHOS RELACIONADOS

Nesta seção de trabalhos relacionados, listam-se 6 trabalhos obtidos através de consultas ao Google Scholar e ao Portal de Periódicos da CAPES, sob os critérios de: (1) abordar o tema de pesquisa do presente trabalho, isto é, software de apoio à pesquisa experimental educacional, ou (2) abordar o tema de software de apoio à pesquisa experimental além da área educacional e (3) ter sido publicado nos últimos 10 anos.

Para a busca, foram utilizadas strings como “internet based experiments”, “conducting web based experiments”, “educational a/b test” e “educational research software controlled experiments internet based”, bem como suas respectivas traduções para atingir artigos nacionais, além do uso da busca por artigos relacionados.

Para cada trabalho listado, apresenta-se um resumo do mesmo, seguido por uma descrição do seu objetivo e proposta, encerrando com os resultados que foram obtidos pelo trabalho. Ao final da seção, apresenta-se uma tabela comparativa dos trabalhos relacionados com o presente trabalho.

#### **3.1 The Assistments Ecosystem: Building A Platform That Brings Scientists And Teachers Together For Minimally Invasive Research On Human Learning And Teaching**

Neste artigo, (HEFFERNAN, N.; HEFFERNAN, C., 2014) trata do processo de construção da plataforma ASSISTments, demonstrando como as escolhas feitas durante esse processo contribuíram para o sucesso e a relevância da plataforma no contexto educacional tanto para estudantes e professores quanto para pesquisadores do meio acadêmico.

A plataforma citada, que é baseada na web, tem como objetivo fornecer uma melhor assistência no processo de aprendizado dos alunos na disciplina de matemática, além de auxiliar professores com relatórios de dados sobre o desempenho de suas turmas. Ela é capaz de auxiliar os alunos durante a resolução de atividades através da funcionalidade de feedback imediato, podendo ele ser de correção, isto é, se o aluno acertou ou não, texto e imagem, além do fornecimento de dicas. O professor possui acesso em tempo real aos dados gerados na resolução

das atividades, como o tempo gasto para resolver as questões, o desempenho do aluno em cada questão, a quantidade de dicas requisitadas, entre outros. Tais dados são agrupados em relatórios nos quais o professor pode utilizar para tomar decisões sobre a abordagem utilizada no ensino de forma a tratar das necessidades da turma (HEFFERNAN, N.; HEFFERNAN, C., 2014).

Como forma de testar a plataforma, (MURPHY et al., 2020, p. 235, tradução nossa) conduziu um ensaio clínico randomizado utilizando a ASSISTments como intervenção, onde os resultados indicaram que “os alunos das escolas designadas a utilizar ASSISTments aprenderam mais e o impacto foi maior para os alunos com menor aproveitamento anterior em matemática”. Além disso, outro ponto levantado pelo estudo foi de que a correção automática das atividades pela plataforma possibilitou aos professores mais tempo para preparar e adaptar suas instruções para as necessidades da turma (MURPHY et al., 2020).

Segundo (HEFFERNAN, N.; HEFFERNAN, C., 2014), a plataforma também pode ser usada como uma ferramenta de pesquisa auxiliando na criação de Ensaio Controlados Randomizados, sendo utilizada em mais de 18 publicações revisadas por pares que, em sua maioria, tratam da capacidade de diferentes feedbacks de influenciar o aprendizado do aluno. Além disso, o autor também acrescenta que os dados da plataforma são utilizados para pesquisa relacionada à mineração de dados, estando presente em mais de 50 publicações revisadas por pares.

### **3.2 Lioness Lab: A Free Web-Based Platform For Conducting Interactive Experiments Online**

Neste artigo, (GIAMATTEI et al., 2020) apresenta a plataforma web LIONESS Lab como uma ferramenta capaz de auxiliar pesquisadores com pouco conhecimento de programação na realização de experimentos interativos online no contexto da pesquisa comportamental, detalhando o processo de condução de experimentos através da plataforma.

Diversos pesquisadores comportamentais utilizam experimentos online para obter mais conhecimento sobre o processo decisório humano. Segundo o autor, tais experimentos geralmente são projetados de forma que não há interação entre participantes, sendo esta a característica o objeto de interesse do autor. Há softwares disponíveis que auxiliam nesse sentido ao permitir a realização de

experimentos com a interação entre participantes, mas os mesmos, para o autor, possuem uma barreira de conhecimento técnico de programação significativa. Com isso, a viabilidade de experimentos interativos é prejudicada (GIAMATTEI et al., 2020).

De acordo com o autor, a plataforma proposta no artigo é capaz de auxiliar a realização dos experimentos interativos sem que o pesquisador tenha conhecimentos avançados de programação, além de contar com o auxílio de diversos métodos para solucionar obstáculos presentes na condução de experimentos, como “redução no tempo de espera dos participantes” e “tratamento dos casos onde os participantes abandonam o experimento”, por exemplo (GIAMATTEI et al., 2020, p. 1, tradução nossa).

Para facilitar a realização dos experimentos, o foco da plataforma é voltado para o desenvolvimento de uma interface intuitiva e de um processo de criação de experimentos dividido em etapas. Apesar de permitir o uso de JavaScript para aprimorar os experimentos, todo o processo pode ser realizado através de cliques na interface, sem a necessidade de inserir códigos de programação (GIAMATTEI et al., 2020).

Segundo o site do LIONESS Lab, a plataforma conta com 1.311 usuários em 66 países, estando presente em 568 instituições e tendo auxiliado a realização de 6.068 experimentos.

### **3.3 Psiturk: An Open-Source Framework For Conducting Replicable Behavioral Experiments Online**

No artigo citado, (GURECKIS et al., 2016) reconhece as dificuldades técnicas presentes na realização de experimentos online no contexto de pesquisa comportamental, em especial com o auxílio do serviço de crowdsourcing Amazon Mechanical Turk. O autor então propõe um projeto open source com foco na facilitação dos processos necessários para conduzir tais experimentos, com a possibilidade de integração com o Amazon Mechanical Turk para o recrutamento de participantes.

Para realizar os experimentos online no contexto apresentado no artigo, é preciso superar alguns obstáculos. Segundo o autor, o experimento necessita de um servidor web, para que o participante possa acessar o experimento do seu browser,

e um banco de dados, para armazenar os dados coletados durante o experimento. Além disso, o autor também cita a necessidade de haver uma integração com o Amazon Mechanical Turk para que seja possível recrutar e, posteriormente, pagar os participantes que aceitaram fazer parte do experimento. Por fim, o próprio experimento deve ser desenvolvido, utilizando HTML, CSS e JavaScript, de forma que o participante, em seu browser, possa realizar as atividades necessárias para completar o experimento (GURECKIS et al., 2016).

A proposta em questão oferece recursos para facilitar o pesquisador a superar tais obstáculos. O autor cita como recurso uma Command Line Tool, capaz de executar comandos no terminal para auxiliar tarefas como controle do servidor do experimento, por exemplo, ou executar comandos que interagem com o Amazon Mechanical Turk como a aprovação de participantes. A CLI também é capaz de criar um projeto base de experimento, onde os arquivos do Web application server são gerados. O pesquisador então pode customizar os arquivos do Web application server para criar o experimento que melhor lhe servir, bem como realizar testes (GURECKIS et al., 2016). Por fim, o autor também disponibiliza o psiTurk Secure Ad Server, criado para solucionar problemas de integração com o Amazon Mechanical Turk relacionados à criptografia SSL.

De acordo com o autor, 24.203 participantes recrutados através do Amazon Mechanical Turk realizaram atividades em experimentos que utilizaram psiTurk, além de impedir que 9.107 participantes refizessem o experimento. Por fim, o autor cita um número de 42.648 execuções da CLI do psiTurk.

### **3.4 Upgrade: An Open Source Tool To Support A/B Testing In Educational Software**

Neste artigo, (RITTER et al., 2020) apresenta a UpGrade, uma ferramenta open source que auxilia na realização de estudos controlados randomizados em softwares educacionais através de uma interface web, bem como demonstra ideias futuras a serem implementadas na ferramenta.

Para demonstrar o funcionamento da ferramenta, o autor cita um exemplo de experimento onde os participantes devem realizar 5 atividades educacionais. No grupo experimental, duas das atividades são distintas do grupo de controle. No fim, busca-se medir o impacto das duas intervenções no aprendizado dos participantes.

Segundo o autor, a ferramenta fornece meios para garantir que restrições do experimento escolhidas pelo pesquisador sejam respeitadas, como a atribuição condicional de intervenções, onde um participante que recebeu uma intervenção experimental continuará recebendo intervenções desse tipo até o fim do experimento (RITTER et al., 2020).

O autor também foca no ponto da consistência, reforçando a ideia de que a ferramenta é capaz de direcionar as intervenções de forma individual e coletiva, no sentido de ser uma classe ou escola, por exemplo. Individualmente, as intervenções são atribuídas para cada aluno, “mesmo que essa experiência seja diferente do resto da classe“, enquanto em grupo todos os indivíduos daquele grupo recebem a mesma intervenção, sem variação (RITTER et al., 2020, p. 4, tradução nossa).

O artigo é concluído com o autor apresentando ideias para expandir o uso da ferramenta, como a possibilidade de realizar diferentes projetos de experimentos, além da integração com plataformas de geração de evidências.

### **3.5 Terracotta - Tool For Educational Research With Randomized Controlled Trials**

Terracotta é uma ferramenta web, atualmente em fase de desenvolvimento na Universidade de Indiana, caracterizada como um *experiment-builder*. Com ela, é possível realizar experimentos randomizados controlados no contexto educacional, facilitando a condução desses experimentos através da automatização de atividades normalmente custosas sem o auxílio de tecnologia (MOTZ et al., 2020).

Segundo o site, a facilidade se dá através de funcionalidades como a exportação de dados do experimento ao final do mesmo contendo todas as informações necessárias para que análises possam ser feitas sobre os resultados do experimento. Além disso, a atribuição de intervenções nos grupos de participantes é feita de maneira automática, podendo ter seus critérios customizados pelo pesquisador. (MOTZ et al., 2020) também cita que a apresentação do Termo de Consentimento também é feita de maneira automática e de forma integrada com o sistema de gestão educacional CANVAS, facilitando o recrutamento de participantes.

O objetivo da plataforma, segundo (MOTZ et al., 2020), é tornar a realização de experimentos randomizados controlados mais acessível para pesquisadores no contexto educacional através da facilitação da criação e condução dos mesmos com

o auxílio de uma plataforma web que está integrada com o sistema de gestão educacional CANVAS.

### **3.6 SES: Um Sistema Baseado na Web para o Design e a Condução de Experimentos Educacionais**

Neste artigo, (BOMBASAR; RAABE, 2017) apresenta a proposta de uma ferramenta capaz de realizar experimentos educacionais na web. Os autores realizaram um experimento para avaliar a ferramenta numa situação de uso real, onde foi possível perceber várias vantagens na realização de experimentos na web através da ferramenta.

Segundo (BOMBASAR; RAABE, 2017, p. 38), o SES é uma “ferramenta de apoio aos pesquisadores para a realização de experimentos educacionais de maneira automatizada, colaborativa e supervisionada”. A ferramenta se dá na forma de um sistema baseado na web com funcionalidades como testes de hipóteses e gerenciamento de experimentos, participantes e testes.

O experimento realizado para o artigo, cujo design contemplava apenas o pós-teste, contou com a participação de seis turmas em duas escolas distintas, totalizando 107 alunos que foram avaliados através de testes sobre modelos de computação utilizando jogos. Com base na condução do experimento, os autores informaram diversas vantagens na utilização da ferramenta, como controle satisfatório de ameaças à validade e facilidade relatada por supervisores nas escolas.

### **3.7 Comparação dos trabalhos relacionados com a proposta do presente trabalho**

De forma a posicionar o presente trabalho em relação aos trabalhos relacionados, uma comparação será realizada utilizando os seguintes critérios:

#### **A. Ferramenta para realização de experimentos**

O trabalho relacionado propõe uma ferramenta ou analisa uma já existente cujo objetivo principal é apoiar pesquisadores a realizarem experimentos em suas áreas de conhecimento.

### B. Ferramenta nacional

A ferramenta proposta ou analisada pelo trabalho foi feita no contexto nacional.

### C. Contexto educacional

O trabalho relacionado dá ênfase a conceitos e propostas que estão inseridas no contexto da pesquisa educacional.

### D. Integração com ambiente virtual de ensino

O trabalho relacionado, ao propor uma ferramenta, o faz de forma a integrar tal ferramenta a um ambiente virtual de ensino

### E. Ferramenta baseada na web

A proposta do trabalho se dá através de uma ferramenta baseada na web que pode ser acessada por um browser sem a necessidade de instalação de pacotes no computador do usuário.

**Tabela 1** - Comparação dos trabalhos relacionados com o presente trabalho

Trabalhos relacionados	Critérios				
	A	B	C	D	E
3.1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
3.2	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3.3	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.4	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
3.5	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
3.6	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Presente trabalho



Fonte: elaborado pelo autor (2022)

Dessa forma, é possível perceber que há um espaço faltante em relação aos trabalhos analisados. É esse espaço, composto pela ausência de funcionalidades e características valiosas para uma ferramenta que apoia a realização de experimentos, que a proposta do presente trabalho se propõe a preencher.

## 4 PROPOSTA

A possibilidade de utilizar um sistema para apoiar a realização de experimentos não é inédita. Algumas iniciativas, como as demonstradas em (RITTER et al., 2020), (MOTZ et al., 2020) e (HEFFERNAN, N.; HEFFERNAN, C., 2014) utilizam sistemas baseados na web para facilitar a realização de experimentos, sendo esta última a mais desenvolvida até então, tendo colaborado com a publicação de dezenas de artigos. Entretanto, a utilização de tais ferramentas se dá mediante a submissão de pedidos de pesquisa para universidades estrangeiras, cuja aprovação não é uma certeza. Além disso, uma ferramenta desenvolvida dentro do contexto nacional possivelmente será capaz de atender melhor às demandas e necessidades dos pesquisadores locais, sem a necessidade de concessão de um meio de pesquisa por terceiros.

A ferramenta proposta por (BOMBASAR; RAABE, 2017) atende tais requisitos, mas não foi concebida com a ideia de integração com um ambiente virtual de ensino, característica essa que pode facilitar o recrutamento de participantes através do acesso ao banco de dados dos estudantes cadastrados, sendo necessário apenas o envio de convites através de notificações. Além disso, a análise estatística dos resultados do experimento pode ser considerada limitada pois possui apenas um teste estatístico do tipo ANOVA *One Way* e sem a presença de um teste para avaliar a normalidade dos dados, apesar da promessa do autor de futura inclusão na ferramenta, cuja necessidade é apontada por (MISHRA et al., 2019). Por fim, para a condução do experimento que avaliou a ferramenta, foi relatado o desenvolvimento de páginas HTML para abrigar os materiais utilizados no experimento, o que consiste numa barreira técnica para pesquisadores sem habilidades de desenvolvimento para a web.

Dessa forma, se faz necessário propor uma solução que esteja acessível ao contexto brasileiro. Para isso, o presente trabalho propõe um sistema, chamado de R-Partner (*Researcher's Partner* - Parceiro do Pesquisador), capaz de apoiar pesquisadores do contexto educacional a realizarem experimentos com design do tipo pré-teste e pós-teste em ambientes online de aprendizagem, facilitando o processo ao automatizar tarefas custosas e reduzindo o esforço necessário para conduzir tais experimentos.

Para atingir tal objetivo, o *R-Partner* possui as funcionalidades de recrutamento facilitado de participantes para o experimento através da integração com um ambiente virtual de aprendizado composto por dois módulos: o *S-Partner* (*Student's Partner*), o módulo dos estudantes, e o *T-Partner* (*Teacher's Partner*), o módulo dos professores. O recrutamento se dá pelo envio de pedidos de permissão para participação de turmas aos professores através do *T-Partner* e o envio de convites aos alunos das turmas permitidas através do *S-Partner*. Além disso, o sistema realiza a atribuição automática e randômica dos participantes nos grupos de controle e experimental, realização de pré-testes, intervenções e pós-testes e, por fim, a análise estatística dos resultados do experimento. Ainda é possível exportar os dados anonimizados do experimento para arquivos CSV para análise própria.

A escolha do design de pré-teste e pós-teste se deu pela intenção de comparar intervenções educacionais. No experimento que o sistema se propõe a realizar, uma intervenção tradicional, ou tratamento, será aplicada em um determinado grupo de participantes, enquanto uma intervenção experimental será aplicada em outro grupo de participantes. O intuito é avaliar as diferenças que a intervenção experimental provocou no seu grupo de participantes em comparação com o grupo controle. Para isso, o design de pré-teste e pós-teste se mostra uma opção adequada de acordo com (LEVY; ELLIS, 2011).

#### **4.1 Integração com o T-Partner e S-Partner**

Para atingir seus objetivos, o R-Partner se integra a dois outros sistemas, S-Partner e T-Partner, ambos desenvolvidos através de um projeto de extensão do Laboratório NEES da UFAL. O T-Partner é um sistema de gestão de aprendizagem onde os professores cadastrados podem criar e gerenciar cursos com seus respectivos materiais de ensino e atividades. O S-Partner, por outro lado, é um ambiente online de aprendizagem onde estudantes podem ingressar em tais cursos, realizando as atividades e consumindo os conteúdos pertinentes aos cursos.

Para o participante cadastrado no S-Partner, o experimento se dará na forma de um curso como qualquer outro. Entretanto, há uma diferenciação pelo fato deste curso, que representa o experimento, ter sido criado por um pesquisador e sua estrutura ser definida pelo design de pré-teste e pós-teste. Logo, o curso contemplará sequencialmente um teste inicial, um conjunto de materiais compondo a

intervenção e um teste final, cujos resultados são visíveis apenas pelo pesquisador através do R-Partner.

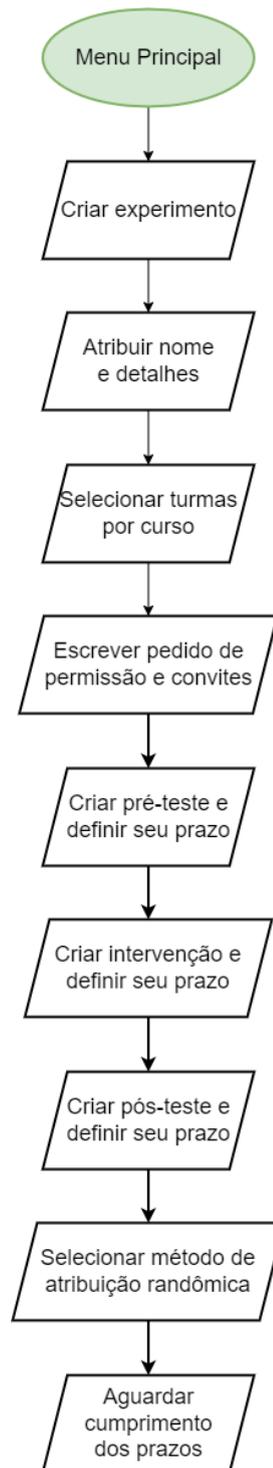
A integração entre sistemas se dá através de uma API no lado do servidor dos sistemas integrados que é acessível pelo R-Partner. A integração se dá em duas formas: compartilhamento de dados e acesso a funcionalidades. O compartilhamento de dados permite ao R-Partner acesso a lista de turmas e seus respectivos professores e estudantes. O acesso a funcionalidades, por outro lado, permite ao R-Partner a criação do curso do experimento, o controle de convites e pedidos de permissão e a realização dos testes do experimento.

## **4.2 Sobre o processo**

Como forma de especificar o caminho a ser percorrido pelo pesquisador desde a criação do experimento até a visualização de um relatório estatístico de seus resultados, foi feito um mapeamento do processo de condução de experimentos no sistema, pensado através das definições sobre experimentos encontradas em (KENDALL, 2003). O processo resultante pode ser dividido em 2 etapas, sendo elas: etapa de criação e etapa de execução.

### **4.2.1 Etapa de criação**

Na etapa de criação, o pesquisador deve fornecer informações que identifiquem o seu experimento dentro do sistema, como nome e detalhes. Além disso, o pesquisador deverá especificar o funcionamento do experimento, isto é, definir quais turmas farão parte do experimento, como se dará o convite aos alunos e o pedido de permissão ao professor, como se dará o pré-teste, as intervenções e o pós-teste e os seus respectivos prazos e, por fim, qual o método a ser utilizado na atribuição randômica de participantes.

**Figura 1** - Recorte da etapa de criação do processo

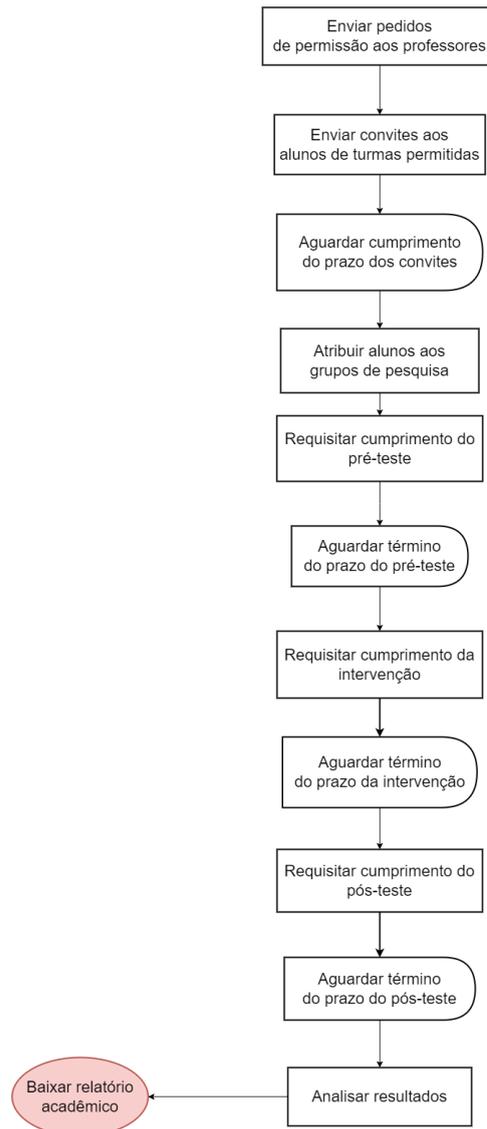
Fonte: elaborado pelo autor (2022)

#### 4.2.2 Etapa de execução

Na etapa de execução, o experimento acontecerá com o auxílio do sistema, que lidará com questões como a aplicação da atribuição randômica dos

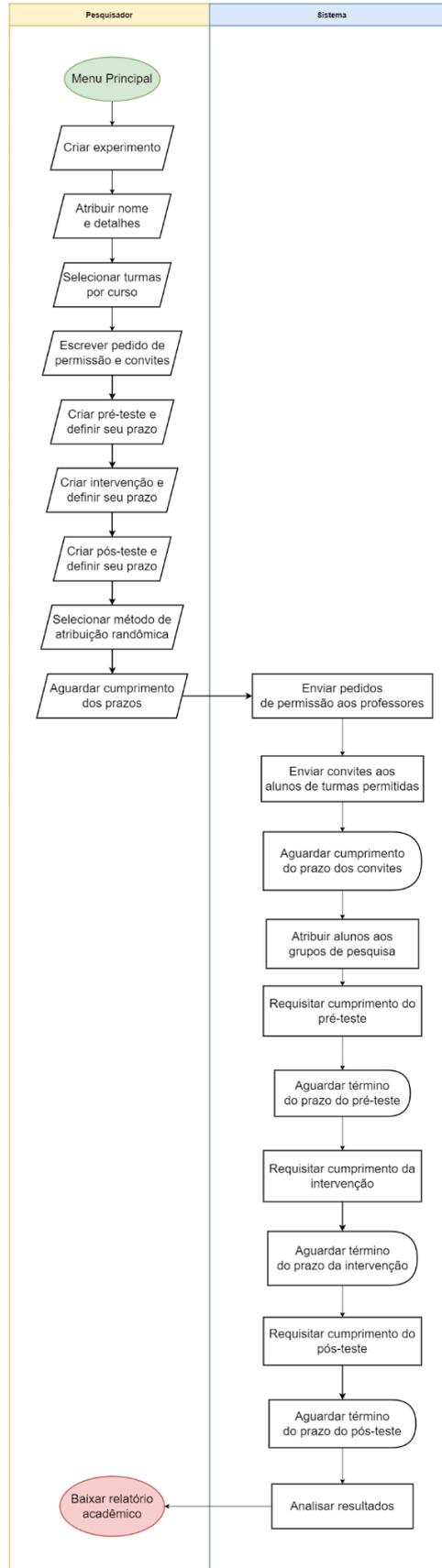
participantes, o envio e a coleta de respostas dos pedidos de permissão, convites, pré-teste e pós-teste e aplicação da intervenção dentro dos prazos especificados. Por fim, os dados obtidos serão analisados e apresentados ao pesquisador na forma de um relatório estatístico, além da possibilidade de baixar um arquivo .CSV com os dados obtidos pelo experimento para análise própria.

**Figura 2 - Recorte da etapa de execução do processo**



**Fonte:** elaborado pelo autor (2022)

Figura 3 - Imagem do processo completo



Fonte: elaborado pelo autor (2022)

### 4.3 Detalhamento do processo

Para cada etapa desse processo, será detalhado abaixo os requisitos que o sistema cumpre e de que forma ocorre a implementação dos mesmos. Um vídeo apresentando o sistema pode ser encontrado em <https://youtu.be/W-KUcC6A9hl>.

#### 4.3.1 Menu Principal

O sistema deve ter como sua página inicial o menu principal, onde o pesquisador deve poder acompanhar, editar e arquivar experimentos em andamento, bem como criar novos. A implementação dessa etapa se dará através de uma tabela listando os experimentos criados pelo pesquisador, com a possibilidade de visualizar, editar e arquivar os experimentos já existentes e criar experimentos novos.

Na figura 4, é possível visualizar o botão de criação do experimento e a tabela citada, com informações como o título, a etapa corrente e a data de criação do experimento. Além disso, uma coluna de ações agrupa botões que permitem ao pesquisador acompanhar o andamento do experimento, bem como editá-lo e arquivá-lo.

Figura 4 - Menu principal do sistema

The screenshot shows the main menu of the R-PARTNER system. At the top, there is a navigation bar with 'R-PARTNER' and 'Experimentos'. Below this, a yellow banner displays a message: 'O R-Partner ainda está em fase de desenvolvimento. Esta versão foi lançada para utilização no experimento que você está participando.' A blue button labeled 'Novo Experimento' is positioned above a table. The table has four columns: 'Titulo', 'Etapa', 'Data de criação', and 'Ações'. It lists four experiments with their respective stages and creation dates. The 'Ações' column contains three buttons for each row: 'Acompanhar' (green), 'Editar' (blue), and 'Arquivar' (red).

Titulo	Etapa	Data de criação	Ações
Teste	Recrutamento	07/12/2022	Acompanhar Editar Arquivar
Teste	Cadastro pendente	26/10/2022	Acompanhar Editar Arquivar
Teste	Cadastro pendente	26/10/2022	Acompanhar Editar Arquivar
Teste	Recrutamento	25/10/2022	Acompanhar Editar Arquivar

2022 R-Partner ©. Todos os Direitos Reservados.

Fonte: elaborado pelo autor (2022)

#### 4.3.2 Criar experimento e atribuir nome e detalhes

O pesquisador deve criar um experimento, onde será requisitado que o mesmo insira o nome e detalhes do experimento para que seja possível identificá-lo e diferenciá-lo de outros experimentos. A implementação dessa etapa se dará através de uma tela de formulário que possuirá dois campos, sendo um para o nome e outro para os detalhes do experimento.

Na figura 5, é possível visualizar a tela da primeira etapa com os dois campos citados, além dos seletores de etapa à esquerda que são liberados de acordo com o avançar das etapas durante a criação do experimento.

Figura 5 - Etapa “Nome e descrição”, primeira na criação do experimento

R-PARTNER Experimentos

1 Nome e descrição

2 Escolha de turmas

3 Convide e permissão

4 Prazo das respostas

5 Criação do pré-teste

6 Criação da intervenção

7 Criação do pós-teste

8 Atribuição randômica

9 Resumo final

Atribua um título e uma descrição ao experimento

Dessa forma, poderemos diferenciá-lo de outros experimentos

Título

Experimento Teste

17/75

Descrição

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam porttitor facilisis massa, id eleifend nulla imperdiet quis. Nullam tellus purus, fermentum in erat vitae, placerat aliquam lectus. Phasellus facilisis et velit a gravida. Aliquam tempor dolor vehicula orci suscipit, nec ultrices purus pulvinar.

309/2000

Cancelar Salvar e avançar

2022 R-Partner ©. Todos os Direitos Reservados.

Fonte: elaborado pelo autor (2022)

#### 4.3.3 Selecionar turmas por curso

O pesquisador deve selecionar quais turmas ele gostaria de convidar para fazer parte do experimento. O sistema implementará uma tabela capaz de filtrar turmas por critérios de inclusão como nome do curso, nome do professor do curso, quantidade mínima e máxima de alunos e datas de início e fim do curso. Cada linha

da tabela terá uma checkbox que diferenciará as turmas selecionadas para as próximas etapas do experimento.

A figura 6 apresenta a tela da segunda etapa de criação do experimento onde se encontra a tabela citada, com colunas para cada dado citado nos filtros, juntamente com os filtros que podem ser expandidos ou retraídos.

**Figura 6** - Etapa “Escolha de turmas”, segunda na criação do experimento

Escolha os cursos que você quer que participem do experimento

Selecione os cursos pela caixa de seleção na tabela abaixo

Curso	Professor	Nº de alunos	Data de início	Data de fim
<input type="checkbox"/> Curso Seed GIAX	Mrs. Addie Goldner	20	26/09/2022	26/09/2023
<input type="checkbox"/> Curso Seed oxRq	Anabel Mante V	20	26/09/2022	26/09/2023
<input type="checkbox"/> Curso Seed LDS4	Angelica Zulauf	20	26/09/2022	26/09/2023
<input type="checkbox"/> Curso Seed QLpy	Dr. Isabel O'Keefe Sr.	20	26/09/2022	26/09/2023
<input type="checkbox"/> Curso Seed P1ex	Chad Gutmann PhD	20	26/09/2022	26/09/2023

2022 R-Partner ©. Todos os Direitos Reservados.

**Fonte:** elaborado pelo autor (2022)

#### 4.3.4 Escrever pedido de permissão e convites

O pesquisador deve escrever tanto o pedido de permissão aos professores quanto o convite aos alunos, além de fornecer em anexo o TCLE (Termo de Comprometimento Livre e Esclarecido) que será apresentado ao aluno juntamente com o convite. Além disso, o pesquisador deve poder definir os prazos para os pedidos de permissão e convite aos alunos. Nesta etapa, o sistema implementará uma tela de formulário com três campos obrigatórios que irão receber o texto do pedido de permissão, do convite e o anexo do TCLE, além de seletores de data para os prazos.

A figura 7 apresenta a tela da terceira etapa de criação do experimento com os campos citados acima, exceto os seletores de datas.

**Figura 7 -** Etapa “Convite e permissão”, terceira na criação do experimento

**R-PARTNER** Experimentos

1 Nome e descrição  
2 Escolha de turmas  
3 **Convite e permissão**  
4 Prazo das respostas  
5 Criação do pré-teste  
6 Criação da intervenção  
7 Criação do pós-teste  
8 Atribuição randômica  
9 Resumo final

**Para os professores**  
Escreva abaixo um texto requisitando ao professor de cada turma selecionada a participação da mesma no experimento.  
Texto do pedido de permissão  
0/2000

**Para os alunos**  
Escreva um texto convidando os alunos das turmas com permissão de participação concedida a fazerem parte do experimento.  
Texto do convite  
0/2000

**TCLE**  
Anexe abaixo o TCLE (Termo de Comprometimento Livre e Esclarecido) do experimento. Ele será enviado aos alunos juntamente com o convite.  
Apenas arquivos PDF. Tamanho máximo: 25MB  
Choose File No file chosen

Voltar para o menu Voltar Salvar e avançar

2022 R-Partner ©. Todos os Direitos Reservados.

**Fonte:** elaborado pelo autor (2022)

A figura 8 apresenta os seletores dos prazos para os pedidos de permissão e convite aos alunos, juntamente com uma tabela das turmas selecionadas pelo usuário com os dados de data de início e fim do curso de cada turma para que o pesquisador possa consultar tais dados para decidir os prazos sem retroceder as etapas.

**Figura 8 -** Etapa “Prazo das respostas”, quarta na criação do experimento

**R-PARTNER** Experimentos

1 Nome e descrição  
2 Escolha de turmas  
3 Convite e permissão  
4 Prazo das respostas  
5 Criação do pré-teste  
6 Criação da intervenção  
7 Criação do pós-teste  
8 Atribuição randômica  
9 Resumo final

### Defina os prazos para os pedidos de permissão e convites

Turmas selecionadas

Curso	Data de início	Data de fim
Curso Seed GIAX	26/09/2022	26/09/2023
Curso Seed oxRq	26/09/2022	26/09/2023
Curso Seed LDS4	26/09/2022	26/09/2023
Curso Seed QLpy	26/09/2022	26/09/2023
Curso Seed P1ex	26/09/2022	26/09/2023

### Prazos

Prazo final para pedidos de permissão  
12/14/2022

Prazo final para convites aos alunos  
12/21/2022

Voltar para o menu Voltar Salvar e avançar

2022 R-Partner ©. Todos os Direitos Reservados.

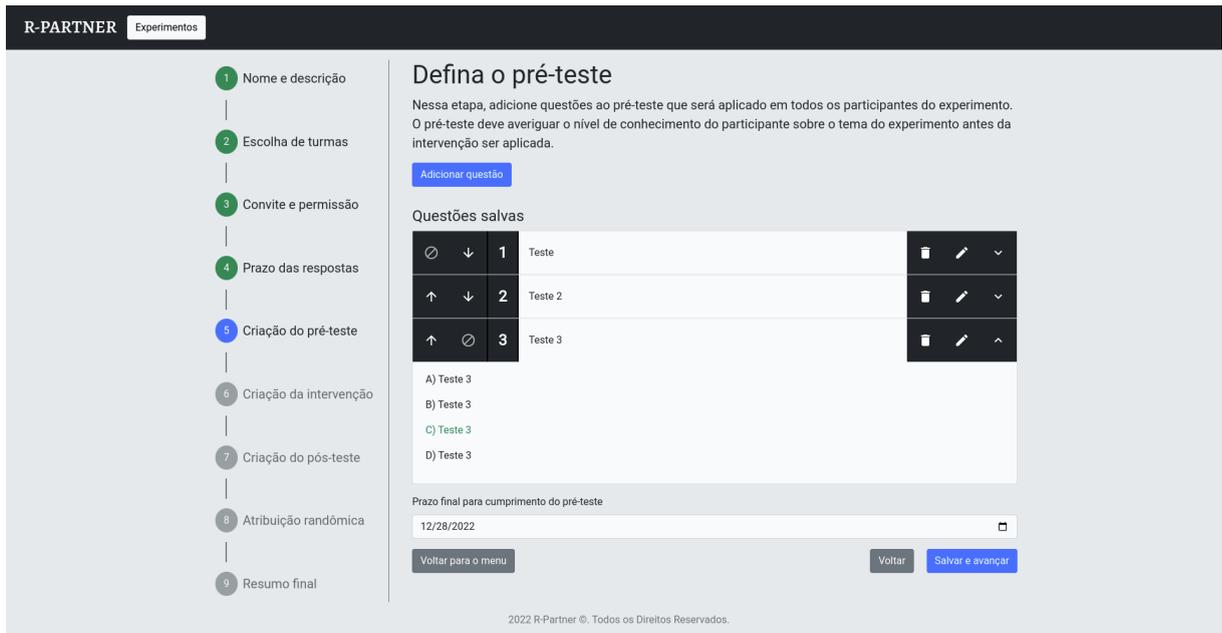
Fonte: elaborado pelo autor (2022)

#### 4.3.5 Criar pré-teste e definir seu prazo

O pesquisador deve criar questões para compor o pré-teste, cuja resolução será posteriormente requisitada aos participantes, além de definir o prazo de cumprimento do pré-teste. A implementação dessa etapa se dará através de um formulário apropriado para a criação das questões. Haverá ainda um campo onde o pesquisador fornecerá a data final permitida para o cumprimento do pré-teste.

A figura 9 apresenta a tela da quinta etapa de criação do experimento onde encontra-se o formulário citado que está acessível através do botão “Adicionar questão”. Abaixo do botão, encontra-se a lista de questões adicionadas, onde o usuário pode definir a ordem das questões, bem editar, remover e expandir as alternativas das questões. Por fim, encontra-se o campo para o seletor do prazo do pré-teste.

**Figura 9** - Etapa “Criação do pré-teste”, quinta na criação do experimento



Fonte: elaborado pelo autor (2022)

#### 4.3.6 Criar intervenção e definir seu prazo

De forma semelhante ao pré-teste, o pesquisador deve tanto criar quanto reusar materiais existentes para compor as intervenções de controle e experimental, além de definir o prazo de cumprimento das intervenções. A implementação dessa etapa se dará através de um seletor onde o pesquisador poderá escolher se deseja reusar ou criar materiais. Caso queira reusar, haverá outro seletor para escolher os materiais já existentes. Caso queira criar, haverá um formulário apropriado para a criação de materiais, suportando formatos de texto, áudio e vídeo. Haverá ainda um campo onde o pesquisador fornecerá a data final permitida para o cumprimento das intervenções.

A figura 10 apresenta a tela da sexta etapa de criação do experimento com o seletor para o prazo das intervenções. Abaixo dele, o botão que abre o formulário de criação de materiais num diálogo. Por fim, para cada intervenção, há um seletor capaz de selecionar materiais criados e botões para adicionar e remover os materiais das intervenções.

**Figura 10** - Etapa “Criação da intervenção”, sexta na criação do experimento

R-PARTNER Experimentos

1 Nome e descrição  
2 Escolha de turmas  
3 Convite e permissão  
4 Prazo das respostas  
5 Criação do pré-teste  
6 Criação da intervenção  
7 Criação do pós-teste  
8 Atribuição randômica  
9 Resumo final

### Defina os materiais das intervenções

Nessa etapa, você deverá criar o material das intervenções que serão aplicadas aos participantes, onde cada grupo de pesquisa possuirá uma intervenção.

Prazo final para cumprimento das intervenções

01/04/2023

Criar conteúdo

#### Intervenção experimental

Selecione abaixo um conteúdo

Adicionar

Teste Remove

Teste Remove

#### Intervenção de controle

Selecione abaixo um conteúdo

Adicionar

Teste Remove

Voltar para o menu Voltar Salvar e avançar

2022 R-Partner ©. Todos os Direitos Reservados.

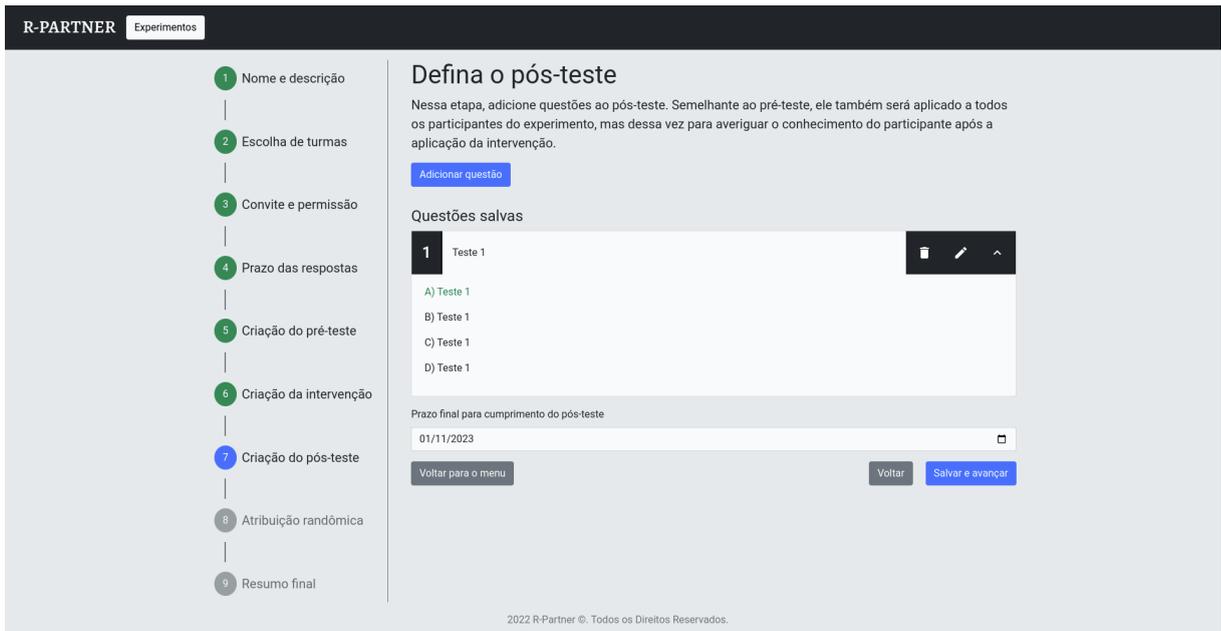
Fonte: elaborado pelo autor (2022)

#### 4.3.7 Criar pós-teste e definir seu prazo

De forma idêntica ao pré-teste, o pesquisador deve tanto criar quanto reusar questões existentes para compor o pós-teste, cuja resolução será requisitada aos participantes após o prazo de cumprimento das intervenções. A implementação desta etapa é idêntica à etapa 4.2.7, com a diferenciação das questões e o prazo escolhido serem destinados ao pós-teste.

A figura 11 apresenta a mesma tela da figura 9. A única diferença é que a figura 11 se refere à tela da sétima etapa de criação do experimento onde o pós-teste é especificado.

**Figura 11** - Etapa “Criação do pós-teste”, sétima na criação do experimento



**Fonte:** elaborado pelo autor (2022)

#### 4.3.8 Selecionar método de atribuição randômica

Inicialmente, apenas a opção de randomização simples estará disponível na primeira versão do sistema. Haverá também, em trabalhos futuros, as opções de randomização por bloco e estratificada. Entretanto, ambas estarão impossibilitadas de escolhas, sendo acompanhadas de uma mensagem informando que tais métodos estarão disponíveis em outras versões. Deve haver um breve texto introduzindo o conceito de cada opção ao pesquisador.

A implementação dessa etapa se dará através de uma página com textos explicativos sobre os métodos de atribuição randômica, possibilitando a seleção de apenas uma das três opções citadas através de um seletor. (SURESH, 2011, p. 9, tradução nossa) define a randomização simples como uma “sequência única de atribuições aleatórias”, citando como exemplo o ato de jogar uma moeda, onde o lado da moeda, cara ou coroa, define se um participante fará parte do grupo experimental ou de controle, de acordo com o critério definido pelo pesquisador. O autor alerta para o uso de randomização simples em amostras pequenas sob o risco de desequilíbrio no tamanho dos grupos de pesquisa. Por exemplo, o grupo de controle ter um número significativamente maior de participantes do que o grupo experimental.

(SURESH, 2011) também define a randomização por bloco como uma solução para o problema do desequilíbrio do tamanho dos grupos de pesquisa. Nele, o autor cita a necessidade de escolher um tamanho para o bloco, sendo este um múltiplo do número de grupos de pesquisa. Depois de escolhido o tamanho, calcula-se todas as combinações balanceadas de atribuição de participantes por bloco. Por fim, o autor cita a necessidade de escolher-se randomicamente os blocos e, através deles, atribuir os participantes aos grupos. O autor faz um alerta sobre o descontrole das covariantes ao utilizar esse método, citando um exemplo num contexto médico onde um grupo de pesquisa pode possuir participantes com mais doenças do que outro, apesar de possuírem um tamanho adequado, cujo desequilíbrio pode prejudicar os resultados do experimento.

Por fim, (SURESH, 2011) também define a randomização estratificada como um método capaz de controlar as covariantes dos participantes durante a atribuição dos grupos de pesquisa. Segundo o autor, os blocos de atribuição são gerados de acordo com as covariantes identificadas, onde os participantes são atribuídos de acordo com suas covariantes individuais nos blocos apropriados onde serão, posteriormente, atribuídos aos grupos de pesquisa através de uma randomização simples. O autor ainda alerta para a complexidade de utilizar esse método com muitas covariantes, além de citar a dificuldade de utilizar esse método num cenário onde as covariantes de todos os participantes não são identificadas previamente, já que os participantes são cadastrados ao longo do tempo e não de uma vez só.

A figura 12 apresenta a tela da oitava etapa de criação do experimento onde encontram-se os seletores para definição do método de atribuição randômica. Acima dos seletores, há um texto explicativo para cada método.

**Figura 12** - Etapa “Atribuição randômica”, oitava na criação do experimento de nome

**R-PARTNER** Experimentos

1 Nome e descrição

2 Escolha de turmas

3 Convite e permissão

4 Prazo das respostas

5 Criação do pré-teste

6 Criação da intervenção

7 Criação do pós-teste

8 **Atribuição randômica**

9 Resumo final

### Escolha o tipo de randomização

Essa randomização será aplicada durante a atribuição dos participantes nos grupos de pesquisa

#### Randomização simples

A randomização simples se dá através de uma sequência de atribuições aleatórias. Imagine que para cada participante será lançado uma moeda e o que definirá o grupo de pesquisa do participante será o lado em que a moeda cair. Por exemplo, se for coroa, será o grupo de controle e, se for cara, o experimental. Esse é o método, como o próprio nome diz, mais simples de randomização.

#### Randomização em bloco

Na randomização em bloco, são gerados blocos de um tamanho múltiplo da quantidade de grupos de pesquisa, onde cada bloco possui uma sequência de atribuição própria definida por permutação. Por exemplo, se temos 2 grupos de pesquisa, o tamanho do bloco pode ser 4 e as atribuições podem ser as seguintes: 'CECE', 'CCEE', 'EECC' e 'ECEC', onde E é o grupo experimental e C o grupo de controle. Será escolhido um bloco aleatoriamente onde os participantes serão atribuídos aos grupos de acordo com a sequência de atribuição do bloco. No primeiro bloco do exemplo anterior, cuja sequência é 'CECE', o primeiro e o terceiro participantes seriam atribuídos ao grupo de controle, e o segundo e quarto ao grupo experimental. Novos blocos são escolhidos aleatoriamente até que todos os participantes sejam atribuídos a um grupo. A randomização em bloco auxilia no balanceamento dos tamanhos dos grupos de pesquisa, onde tanto o grupo de controle quanto o grupo experimental terão uma quantidade de participantes semelhante, sendo igual.

#### Randomização estratificada

Nesse método, são gerados blocos a partir das características dos participantes, como idade ou sexo. Pode-se ter, por exemplo, um bloco para pessoas abaixo de 20 anos, outro bloco para pessoas entre 21 e 29 anos e outros blocos para as demais faixas etárias. Cada participante estará presente num bloco correspondente à sua característica. Os grupos de pesquisa receberão então um determinado número de participantes de cada bloco de forma randômica. A randomização estratificada auxilia no balanceamento das características dos participantes nos grupos de pesquisa.

Randomização simples

Randomização em bloco Em breve

Randomização estratificada Em breve

[Voltar para o menu](#) [Voltar](#) [Salvar e avançar](#)

2022 R-Partner ©. Todos os Direitos Reservados.

Fonte: elaborado pelo autor (2022)

#### 4.3.9 Enviar pedidos de permissão aos professores

O sistema deve enviar os pedidos de permissão aos professores das turmas selecionadas através do *T-Partner*, o módulo dos professores. A implementação dessa etapa se dará através da integração com o módulo dos professores, onde uma notificação será enviada para o professor de cada turma escolhida informando-o do pedido e requisitando sua resposta.

#### 4.2.10 Enviar convites aos alunos de turmas permitidas

O sistema deve enviar o convite aos alunos das turmas permitidas através do *S-Partner*, o módulo dos estudantes. A implementação dessa etapa se dará através da integração com o módulo dos alunos, onde uma notificação será enviada para cada aluno de cada turma permitida informando-o do convite e requisitando sua resposta.

#### 4.3.11 Atribuir alunos aos grupos de pesquisa

O sistema deve atribuir os alunos aos grupos de pesquisa de acordo com o método de atribuição randômica selecionado pelo pesquisador. A implementação

dessa etapa se dará através de um algoritmo que, dado uma lista de participantes, ou seja, alunos que aceitaram o convite do experimento, retorne dois objetos, onde cada um representará um grupo de pesquisa com seus respectivos participantes atribuídos.

#### 4.3.12 Requisitar cumprimento do pré-teste

O sistema deve requisitar o cumprimento do pré-teste através da integração com o *S-Partner*, o módulo dos estudantes. A implementação dessa etapa se dará através do envio de notificações aos participantes requisitando o cumprimento do pré-teste. É reaproveitada a funcionalidade de aplicação e resolução de exercícios do *S-Partner* para o cumprimento do pré-teste, já que o mesmo é composto por um conjunto de questões fechadas, modelo compatível com os exercícios do *S-Partner*.

#### 4.3.13 Requisitar cumprimento da intervenção

O sistema deve requisitar o cumprimento da intervenção aos participantes, de acordo com o grupo de pesquisa de cada participante. A implementação dessa etapa se dará através da integração com o *S-Partner*, o módulo dos estudantes, onde uma notificação será enviada para cada participante requisitando o cumprimento da intervenção adequada a seu grupo de pesquisa. De forma semelhante à etapa anterior, a funcionalidade de visualização de materiais de ensino do *S-Partner* é reaproveitada para permitir o acesso do material das intervenções pelo participante.

#### 4.3.14 Requisitar cumprimento do pós-teste

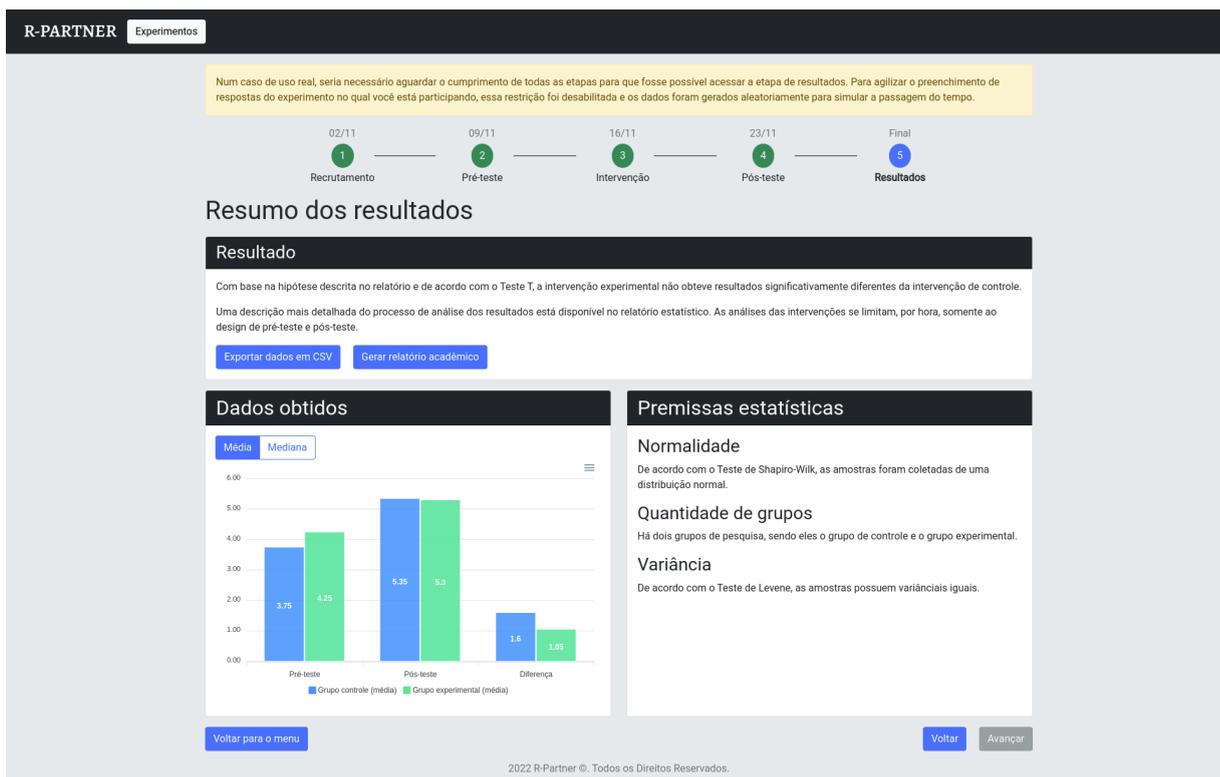
O sistema deve requisitar o cumprimento do pós-teste aos participantes. A implementação dessa etapa se dará da mesma forma que a etapa 4.2.12, com a diferenciação do pós-teste.

#### 4.3.15 Produzir relatório estatístico

O sistema deve apresentar ao pesquisador o resultado de uma análise estatística sobre os dados do experimento através de um relatório estatístico que pode ser baixado através de um botão, em conjunto com a possibilidade de exportar tais dados em arquivos CSV. A implementação dessa etapa se dará da seguinte forma: o sistema irá comparar os resultados do pré-teste e pós-teste entre os grupos de controle e experimental e fazer uma análise utilizando o teste estatístico mais apropriado aos dados.

Tais funcionalidades podem ser acessadas na página de resultados do experimento demonstrada na figura 13, acessível após o término dos prazos definidos pelo usuário. Antes, somente as páginas de acompanhamento do experimento são acessíveis, de acordo com os prazos definidos para cada etapa.

**Figura 13** - Etapa de resultados no acompanhamento do experimento



**Fonte:** elaborado pelo autor (2022)

A figura 14 mostra a primeira página do relatório estatístico gerado pelo sistema. Nela, são descritos os tópicos que o relatório irá tratar, iniciando com a análise dos participantes do experimento para cada grupo de pesquisa separadamente.

Figura 14 - Início do relatório estatístico

## Relatório estatístico

Este é o relatório estatístico do seu experimento intitulado Teste. Nele, os resultados obtidos no pré-teste e pós-teste são utilizados para avaliar o desempenho da intervenção experimental em relação à intervenção de controle. As análises das intervenções se limitam, por hora, ao design de pré-teste e pós-teste.

### Tópicos do relatório

1. **Dados da amostra**  
Detalhamento dos dados obtidos durante o experimento
2. **Preparação dos dados**  
Detalhamento do uso dos dados para as demandas dos testes estatísticos
3. **Premissas estatísticas**  
Realização dos testes de avaliação das premissas necessárias
4. **Teste estatístico**  
Realização do teste estatístico final
5. **Conclusão**  
Definição e teste das hipóteses

### 1. Dados da amostra

Inicialmente, a amostra contou com 40 participantes registrados após a etapa de recrutamento do experimento que ocorreu entre 25/10/2022 e 02/11/2022. O tamanho da amostra sugerido, de acordo com a fórmula simplificada de Yamane ([Link](#)) e utilizando um intervalo de confiança de 95%, é de 55 participantes. É necessário atentar à uma possível ameaça à validade por conta da quantidade de participantes ser inferior ao tamanho sugerido da amostra.

Os participantes foram separados em dois grupos de pesquisa através do método de randomização simples. O grupo de controle recebeu 20 participantes, enquanto o grupo experimental recebeu 20 participantes.

#### 1.1 Participação do grupo de controle

Dos 20 participantes do grupo controle, 20 (100%) completaram o pré-teste e 0 (0%) falharam em completar. Dos 20 participantes restantes do grupo **controle**, 20 (100%) completaram a intervenção e 0 (0%) falharam em completar. Por fim, 20 (100%) participantes completaram o pós-teste e 0 (0%) falharam em completar.

Nenhum participante do grupo controle foi excluído do experimento por falta de participação. A taxa de resposta do grupo de controle foi de 100%, enquanto a taxa de mortalidade foi de 0%.

**Fonte:** elaborado pelo autor (2022)

A figura 15 mostra a sequência ao início do relatório estatístico. As mesmas informações apresentadas sobre os participantes em ambos os grupos de pesquisa é apresentada sobre todos os participantes, sem distinção de grupo. Depois, um texto explicativo sobre a preparação dos dados é apresentado para informar ao pesquisador sobre como os dados foram tratados antes de serem utilizados nos próximos capítulos.

**Figura 15** - Continuação do primeiro capítulo do relatório estatístico

### 1.3 Participação geral

Dos 40 participantes do experimento, 40 (100%) completaram do pré-teste e 0 (0%) falharam em completar. Dos 40 participantes restantes, 40 (100%) completaram a intervenção e 0 (0%) falharam em completar. Por fim, 40 (100%) participantes completaram o pós-teste e 0 (0%) falharam em completar.

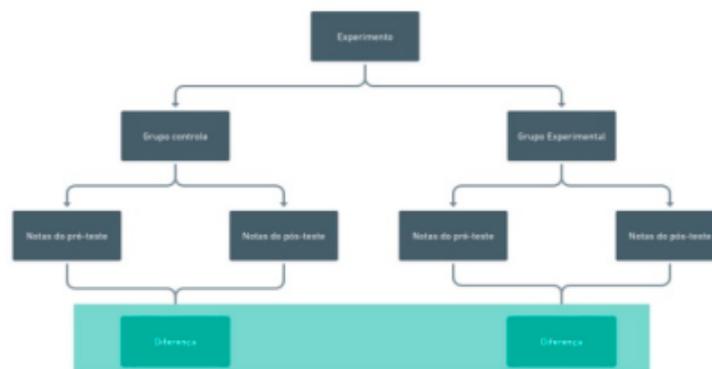
Nenhum participante foi excluído do experimento por falta de participação. A taxa de resposta foi de 100%, enquanto a taxa de mortalidade foi de 0%.

#### 1.3.1 Dados demográficos

Dado	Quantidade	Porcentagem
<b>Idade</b>		
18-20	20	50%
21-25	12	30%
26-29	6	15%
30+	2	5%
<b>Gênero</b>		
Masculino	18	45%
Feminino	22	55%
Outros	0	0%

## 2. Preparação dos dados

Foram reunidos os dados das pontuações dos participantes no pré-teste e no pós-teste. Em seguida, foi calculada a diferença entre a pontuação do pós-teste e do pré-teste de cada participante, onde o resultado é armazenado no conjunto de dados do grupo de pesquisa no qual o participante faz parte. Esses dois conjuntos, que refletem a variação das notas dos testes de cada grupo, é a fonte de dados utilizada nesse relatório.



**Fonte:** elaborado pelo autor (2022)

Para verificar qual o teste estatístico mais apropriado, (NAYAK; HAZRA, 2011, p. 85, tradução nossa) apresenta a questão do pareamento dos grupos de dados onde “grupos ou conjuntos de dados são considerados como não pareados se não houver possibilidade de que os valores em um conjunto de dados estejam relacionados ou sejam influenciados pelos valores nos outros conjuntos de dados.” Logo, podemos assumir que os grupos de pesquisa são não-pareados, pois não influenciam um ao outro.

Outro ponto levantado por (NAYAK; HAZRA, 2011) é a questão da distribuição dos dados. Segundo o autor, dados que possuem uma distribuição normal recebem testes paramétricos, caso contrário, os testes não-paramétricos são utilizados. Para avaliar tal fator, o autor recomenda o uso de testes de normalidade, citando como exemplo o teste de Shapiro-Wilk e o teste de Kolmogorov-Smirnov.

Na figura 16, é demonstrado o capítulo do relatório estatístico referente às premissas estatísticas. Nele, são detalhados quais as premissas avaliadas, quais os testes utilizados pelo sistema e suas respectivas hipóteses e resultados.

**Figura 16** - Capítulo de premissas estatísticas do relatório

### 3. Premissas estatísticas

Antes de realizar os testes estatísticos, se faz necessário avaliar as premissas estatísticas para determinar qual teste será usado, além de validar os resultados alcançados pelo mesmo. As premissas a serem avaliadas serão: normalidade, variância dos dados e quantidade de grupos.

Todos os testes são realizados com auxílio da ferramenta SciPy. Por conta disso, as hipóteses são definidas de antemão pela ferramenta.

#### 3.1. Normalidade

Para avaliar a normalidade, foi realizado um teste de Shapiro-Wilk. Esse teste é usado para verificar se uma determinada amostra vem de uma distribuição normal. Nesse teste, temos duas hipóteses:

**Hipótese nula (H0)**

Os dados foram coletados de uma distribuição normal.

**Hipótese alternativa (H1)**

Os dados NÃO foram coletados de uma distribuição normal.

Se o p-valor for menor ou igual a 0.05, a hipótese nula pode ser rejeitada. Caso contrário, a hipótese nula NÃO pode ser rejeitada.

#### Resultado

O p-valor encontrado no teste foi de ~0.263. Logo, a hipótese nula não pode ser rejeitada.

#### 3.2. Variância dos dados

Para avaliar a variância dos dados, foi realizado um Teste de Levene. Esse teste é usado para verificar se as variâncias de duas amostras são iguais. Nesse teste, temos duas hipóteses:

**Hipótese nula (H0)**

As amostras possuem variâncias iguais.

**Hipótese alternativa (H1)**

As amostras NÃO possuem variâncias iguais.

Se o p-valor for menor ou igual a 0.05, a hipótese nula pode ser rejeitada. Caso contrário, a hipótese nula NÃO pode ser rejeitada.

#### Resultado

O p-valor encontrado no teste foi de ~0.794. Logo, a hipótese nula não pode ser rejeitada.

#### 3.3. Quantidade de grupos

Atualmente, os experimentos realizados pelo R-Partner dão suporte a apenas 2 grupos de pesquisa, sendo eles o grupo controle e o grupo experimental.

**Fonte:** elaborado pelo autor (2022)

Para avaliar o teste estatístico mais apropriado, se faz necessário aplicar antes o teste de normalidade de Shapiro-Wilk, como também recomendado por (MISHRA et al., 2019). Caso a distribuição dos dados seja normal, utiliza-se o teste T. Caso a distribuição não seja normal, utiliza-se o teste de Mann-Whitney (NAYAK; HAZRA, 2011).

Na figura 17, é demonstrado o capítulo do relatório estatístico referente à escolha do teste estatístico. Nele, apresenta-se o processo decisório conduzido pelo

sistema até a escolha do teste. Em seguida, na Figura 18, o próximo capítulo apresenta as hipóteses e o resultado do teste aplicado.

**Figura 17** - Capítulo de escolha do teste estatístico do relatório

## 4. Escolha do teste estatístico

O propósito do teste estatístico neste relatório é avaliar estatisticamente a diferença entre os resultados obtidos no grupo de controle e no grupo experimental. Por conta disso, o relatório possui duas opções de teste estatístico: Teste T e Teste de Mann-Whitney.

A escolha entre ambos é feita mediante a avaliação das premissas estatísticas, tratada no capítulo 3. Os critérios para a escolha do teste se baseiam nas suposições estatísticas de cada teste, que estão descritas abaixo.

### 4.1 Suposições

#### 4.1.1 Suposições do Teste T

<b>Independência</b> As observações em uma amostra são independentes das observações na outra amostra.
<b>Normalidade</b> Ambas as amostras são distribuídas normalmente.
<b>Homogeneidade de Variâncias</b> Ambas as amostras têm aproximadamente a mesma variação. Caso não haja homogeneidade de variâncias, a adaptação do Teste T de Welch pode ser utilizada.
<b>Amostragem Randômica</b> Ambas as amostras foram obtidas usando um método de amostragem aleatória.

#### 4.1.2 Suposições do Teste de Mann-Whitney

<b>Independência</b> As observações em uma amostra são independentes das observações na outra amostra.
<b>Normalidade</b> Ambas as amostras não são distribuídas normalmente.
<b>Amostragem Randômica</b> Ambas as amostras foram obtidas usando um método de amostragem aleatória.

**Fonte:** elaborado pelo autor (2022)

**Figura 18** - Continuação do capítulo de escolha do teste estatístico do relatório

#### 4.2 Tabela comparativa

Abaixo, serão comparados os resultados obtidos nas análises das premissas estatísticas do capítulo 3 com as suposições descritas acima. A plataforma satisfaz a suposição de independência e amostragem randômica e, portanto, ambas não foram incluídas na tabela por simplificação.

Teste	Normalidade	Variância
Teste T	☑	☑
Teste T de Welch	☑	✘
Teste de Mann-Whitney	✘	☹
Dados do experimento	☑	☑

Legenda

- ☑ Requer (caso seja um teste) ou atende (caso sejam os dados) a suposição
- ✘ Não requer (caso seja um teste) ou não atende (caso sejam os dados) a suposição
- ☹ Independe da suposição

#### 4.3 Resultado da comparação

Com base na tabela de critérios acima, o teste escolhido para o experimento foi o Teste T.

### 5. Aplicação do teste estatístico

Para avaliar a diferença entre as médias dos dados do grupo controle e grupo experimental, será aplicado um Teste T. Nesse teste, temos duas hipóteses:

Todos os testes são realizados com auxílio da ferramenta SciPy. Por conta disso, as hipóteses são definidas de antemão pela ferramenta.

#### Hipótese nula (H0)

A média dos dados do grupo de intervenção é igual a média dos dados do grupo de controle.

#### Hipótese alternativa (H1)

A média dos dados do grupo de intervenção NÃO é igual a média dos dados do grupo de controle.

Se o p-valor for menor ou igual a 0.05, a hipótese nula pode ser rejeitada. Caso contrário, a hipótese nula NÃO pode ser rejeitada.

#### 5.1 Resultado

O p-valor encontrado no teste foi de ~0.262. Logo, a hipótese nula não pode ser rejeitada.

**Fonte:** elaborado pelo autor (2022)

Na figura 19, é demonstrado, por fim, os últimos capítulos do relatório estatístico. Nele, apresenta-se o resultado do teste de correlação de Pearson para avaliar a correlação dos dados entre ambos os grupos de pesquisa. Depois, a hipótese final do relatório é definida e, de acordo com o resultado do teste estatístico demonstrado no capítulo anterior, é apresentada a conclusão do relatório apontando se houve ou não diferença estatisticamente significativa entre as intervenções.

**Figura 19** - Capítulos finais do relatório

## 6. Dependência

Para contribuir com a análise final dos resultados, o relatório também busca avaliar a dependência entre os dados do grupo controle e experimental.

Com esse objetivo, foi realizado um teste de Correlação de Pearson. Esse teste é usado para obter uma medida de quão forte dois conjuntos de dados estão linearmente relacionados. Nesse teste, temos duas hipóteses:

Todos os testes são realizados com auxílio da ferramenta SciPy. Por conta disso, as hipóteses são definidas de antemão pela ferramenta.

### Hipótese nula (H0)

Não há relacionamento entre os dados de ambos os grupos de pesquisa.

### Hipótese alternativa (H1)

Há relacionamento entre os dados de ambos os grupos de pesquisa.

Se o p-valor for menor ou igual a 0.05, a hipótese nula pode ser rejeitada. Caso contrário, a hipótese nula NÃO pode ser rejeitada.

### 6.1 Resultado

O p-valor encontrado no teste foi de ~0.034. Logo, a hipótese nula foi rejeitada .

## 7. Conclusão

O relatório busca comparar as diferenças entre os dados do grupo de controle e experimental. O resultado dessa comparação está diretamente ligado ao resultado do teste estatístico realizado no capítulo 5. Dessa forma, temos também duas hipóteses finais:

### Hipótese nula (H0)

A intervenção experimental NÃO obteve resultados diferentes da intervenção de controle.

### Hipótese alternativa (H1)

A intervenção experimental obteve resultados diferentes da intervenção de controle.

Com a aceitação da hipótese nula do teste estatístico, não podemos rejeitar a hipótese nula final, concluindo que a intervenção experimental não obteve resultados diferentes, de forma estatisticamente significativa, da intervenção de controle.

**Fonte:** elaborado pelo autor (2022)

Para compor o CSV contendo os resultados do experimento, o sistema irá atribuir identificadores anônimos para os participantes, relacionando-os ao grupo de pesquisa que pertencem e incluindo, para cada um, suas respostas do pré-teste e pós-teste.

## 5 MATERIAIS E MÉTODOS

Neste capítulo, será detalhado o experimento aplicado durante a pesquisa, descrevendo seu objetivo, o instrumento de coleta de dados e a definição do público alvo para participantes.

### 5.1 Objetivo do Experimento

O experimento tem como objetivo avaliar a utilidade, a facilidade de uso e o custo de tempo do sistema proposto na realização de experimentos com *design* do tipo pré-teste e pós-teste. Com isso, espera-se avaliar se o sistema proposto de fato facilita a realização desses experimentos.

Inicialmente, se fez necessário definir o que seria facilidade. Para isso, foi utilizado o conceito de Facilidade de Uso Percebida que, segundo (DAVIS, 1989, p. 320, tradução nossa), significa “o grau com que uma pessoa acredita que a utilização de um determinado sistema se daria livre de esforço físico e mental”.

Da mesma forma, para definir o conceito de utilidade, utilizou-se o conceito de Utilidade Percebida que, segundo (DAVIS, 1989, p. 320, tradução nossa), significa “o grau com que uma pessoa acredita que a utilização de um determinado sistema melhoraria seu desempenho profissional”.

### 5.2 Instrumento de coleta de dados

Para obter os dados necessários para avaliar tais fatores, foi criado um experimento composto por 8 atividades a serem realizadas pelos participantes através do sistema proposto. As atividades, descritas na Tabela 1, consistem na criação de um experimento e na verificação do relatório estatístico dos resultados produzidos ao final do experimento.

**Tabela 2** - Atividades por etapa do questionário do experimento

<b>Etapa</b>	<b>Atividade</b>
1	Dar nome e detalhes ao experimento
2	Selecionar turmas participantes do experimento

- 3 Criar pedidos de permissão e convite
- 4 Criar pré-teste
- 5 Criar intervenção
- 6 Criar pós-teste
- 7 Selecionar método de atribuição randômica dos participantes nos grupos de pesquisa
- 8 Verificar o relatório estatístico produzido

---

**Fonte:** elaborado pelo autor (2022)

Para cada atividade, o participante é requisitado a compartilhar, em um questionário, suas percepções a partir das 3 afirmações descritas na Tabela 3. A escala Likert citada acima foi composta pelas seguintes alternativas: (1) Discordo fortemente, (2) Discordo, (3) Concordo, (4) Concordo fortemente. A escolha de uma escala de 4 alternativas foi motivada pela intenção de forçar o participante a dar uma resposta de concordância ou discordância, evitando uma resposta neutra.

**Tabela 3** - Afirmações utilizadas nas questões por etapa onde  $n$  equivale a qualquer uma das 8 etapas da Tabela 1

<b>Afirmação</b>	<b>Métrica</b>	<b>Tipo</b>	<b>Valores</b>
O sistema foi útil na realização da etapa $n$ do experimento	Utilidade Percebida	Qualitativo	Escala Likert de 4 pontos
O sistema foi fácil de utilizar na realização da etapa $n$ do experimento	Facilidade de Uso Percebida	Qualitativo	Escala Likert de 4 pontos
O sistema teve um baixo custo de tempo na realização da etapa $n$ do experimento	Custo de tempo	Qualitativo	Escala Likert de 4 pontos

---

**Fonte:** elaborado pelo autor (2022)

Ao finalizar todas as etapas, foi requisitado ao participante que respondesse um conjunto final de questões, adaptado do questionário utilizado por (SEK et al., 2010), no qual o mesmo deveria expressar a sua percepção, a partir de determinadas afirmações, sobre a utilidade e a facilidade de uso do sistema como

um todo em comparação com as ferramentas normalmente utilizadas pelo próprio participante na realização de experimentos.

**Tabela 4** - Afirmações utilizadas no conjunto final de questões

<b>Afirmiação</b>	<b>Identificador</b>	<b>Variável</b>	<b>Tipo</b>	<b>Valores</b>
Aprender a usar o sistema seria fácil para mim em comparação com as ferramentas que normalmente utilizo para a realização de experimentos	FUPC1	Facilidade de Uso Percebida Comparada	Qualitativo	Escala Likert de 5 pontos
Eu sei facilmente como proceder com o sistema para fazer o que eu quero em comparação com as ferramentas que normalmente utilizo para a realização de experimentos	FUPC2	Facilidade de Uso Percebida Comparada	Qualitativo	Escala Likert de 5 pontos
A maneira de interação com o sistema é clara e facilmente compreendida em comparação com as ferramentas que normalmente utilizo para a realização de experimentos	FUPC3	Facilidade de Uso Percebida Comparada	Qualitativo	Escala Likert de 5 pontos
Seria fácil ficar mais habilidoso no uso do sistema em comparação com as ferramentas que normalmente utilizo para a realização de experimentos	FUPC4	Facilidade de Uso Percebida Comparada	Qualitativo	Escala Likert de 5 pontos
Eu considero o sistema fácil de usar em comparação com as ferramentas que normalmente utilizo para a realização de	FUPC5	Facilidade de Uso Percebida Comparada	Qualitativo	Escala Likert de 5 pontos

## experimentos

Usando o sistema meu trabalho fica mais rápido em comparação com as ferramentas que normalmente utilizo para a realização de experimentos	UPC1	Utilidade Percebida Comparada	Qualitativo	Escala Likert de 5 pontos
Usando o sistema meu desempenho melhora em comparação com as ferramentas que normalmente utilizo para a realização de experimentos	UPC2	Utilidade Percebida Comparada	Qualitativo	Escala Likert de 5 pontos
Usando o sistema minha produtividade aumenta em comparação com as ferramentas que normalmente utilizo para a realização de experimentos	UPC3	Utilidade Percebida Comparada	Qualitativo	Escala Likert de 5 pontos
Meu trabalho fica mais efetivo usando o sistema em comparação com as ferramentas que normalmente utilizo para a realização de experimentos	UPC4	Utilidade Percebida Comparada	Qualitativo	Escala Likert de 5 pontos
Meu trabalho fica mais fácil usando o sistema em comparação com as ferramentas que normalmente utilizo para a realização de experimentos	UPC5	Utilidade Percebida Comparada	Qualitativo	Escala Likert de 5 pontos
O sistema é mais útil para o meu trabalho do que as ferramentas que normalmente utilizo para a realização de experimentos	UPC6	Utilidade Percebida Comparada	Qualitativo	Escala Likert de 5 pontos

A escala Likert citada acima foi composta pelas seguintes alternativas: (1) Discordo fortemente, (2) Discordo, (3) Não discordo nem concordo, (4) Concordo, (5) Concordo fortemente. A motivação para a escolha da escala Likert de 5 pontos, ao invés da escala de 4 pontos utilizada anteriormente, se deu pela possibilidade do participante utilizar uma ferramenta de capacidade semelhante ao sistema proposto. Logo, há a necessidade de uma resposta neutra.

Para cada etapa e para o conjunto final de questões, foi disponibilizado um espaço para que o participante pudesse escrever, se desejasse, comentários, sugestões de melhorias e críticas em relação às etapas individualmente ou ao sistema como um todo.

### **5.3 Materiais**

Foi utilizado o sistema proposto, que é baseado na web e utiliza as tecnologias PHP (Laravel), JavaScript (Vue.js) e MySQL. O questionário do experimento foi feito através do Google Forms. A hospedagem do sistema na web se deu através do Heroku, uma plataforma como serviço (PaaS) na nuvem, com auxílio financeiro obtido através do *GitHub Student Pack*.

### **5.4 Participantes**

O público alvo do experimento é caracterizado por membros do Instituto de Computação da Universidade Federal de Alagoas, apesar de um dos participantes ser de uma universidade estrangeira. Buscou-se atrair participantes que possuíam experiência com experimentos controlados como alunos em iniciação científica, de mestrado e doutorado, além de professores e pesquisadores.

A divulgação do experimento se deu através de listas de email institucional e convites diretos aos participantes. O acesso ao experimento esteve disponível por 14 dias.

### **5.5 Hipóteses**

Para formular as hipóteses a serem testadas pela análise inferencial dos resultados do experimento, os dados obtidos pelos questionários serão agrupados em duas categorias: concordância e discordância. Por concordância, entende-se todas as respostas cujas alternativas escolhidas foram “Concordo fortemente” ou “Concordo”. Da mesma forma, por discordância, entende-se todas as respostas cujas alternativas escolhidas foram “Discordo fortemente” ou “Discordo”.

Esta abordagem criará dois grupos que serão comparados de forma a avaliar a concordância ou discordância de cada métrica avaliada. Logo, de acordo com o objetivo definido para o experimento, pode-se definir as seguintes perguntas:

- “O sistema é útil ao pesquisador quando apoia a realização de experimentos com design do tipo pré-teste e pós-teste em ambientes online de aprendizagem?”, com as seguintes hipóteses:
  - H1-0: Não há diferença entre as percepções dos participantes em relação à utilidade do sistema para as diferentes etapas do experimento.
  - H1-1: Há diferença entre as percepções dos participantes em relação à utilidade do sistema para as diferentes etapas do experimento.
  
- “O sistema é fácil de usar quando apoia a realização de experimentos com design do tipo pré-teste e pós-teste em ambientes online de aprendizagem?”, com as seguintes hipóteses:
  - H2-0: Não há diferença entre as percepções dos participantes em relação à facilidade de uso do sistema para as diferentes etapas do experimento.
  - H2-1: Há diferença entre as percepções dos participantes em relação à facilidade de uso do sistema para as diferentes etapas do experimento.
  
- “A realização de experimentos com design do tipo pré-teste e pós-teste em ambientes online de aprendizagem através do sistema possui um baixo custo de tempo?”, com as seguintes hipóteses:
  - H3-0: Não há diferença entre as percepções dos participantes em relação ao custo de tempo do sistema para as diferentes etapas do experimento.

- H3-1: Há diferença entre as percepções dos participantes em relação ao custo de tempo do sistema para as diferentes etapas do experimento.
- “O sistema é fácil de usar quando apoia a realização de experimentos com design do tipo pré-teste e pós-teste em ambientes online de aprendizagem em comparação com as ferramentas normalmente utilizadas pelos participantes para a realização de experimentos?”, com as seguintes hipóteses:
  - H4-0: Não há diferença entre as percepções dos participantes em relação à facilidade de uso do sistema em comparação com as ferramentas normalmente utilizadas pelos participantes para a realização de experimentos.
  - H4-1: Há diferença entre as percepções dos participantes em relação à facilidade de uso do sistema em comparação com as ferramentas normalmente utilizadas pelos participantes para a realização de experimentos.
- “O sistema é útil ao pesquisador quando apoia a realização de experimentos com design do tipo pré-teste e pós-teste em ambientes online de aprendizagem em comparação com as ferramentas normalmente utilizadas pelos participantes para a realização de experimentos?”, com as seguintes hipóteses:
  - H5-0: Não há diferença entre as percepções dos participantes em relação à utilidade do sistema em comparação com as ferramentas normalmente utilizadas pelos participantes para a realização de experimentos.
  - H5-1: Há diferença entre as percepções dos participantes em relação à utilidade do sistema em comparação com as ferramentas normalmente utilizadas pelos participantes para a realização de experimentos.

- “O sistema facilita o trabalho do pesquisador quando apoia a realização de experimentos com design do tipo pré-teste e pós-teste em ambientes online de aprendizagem?”
  - H6-0: H1-1 + H2-1 + H3-1 + H4-1 + H5-1
  - H6-1: H1-0 + H2-0 + H3-0 + H4-0 + H5-0

## 6 RESULTADOS E DISCUSSÃO

Neste capítulo, o conjunto de dados obtido como resultado do experimento será analisado de forma a apresentar as descobertas da pesquisa e relacioná-las com o objetivo da mesma.

### 6.1 Análise das participações

Um total de 17 participantes aceitaram o convite e completaram, de forma adequada, todas as etapas do experimento. Entretanto, como apresenta-se na Tabela 4, dois participantes afirmaram que não realizam experimentos controlados com frequência. Logo, ambos foram excluídos da análise, devido ao fato da experiência com experimentos ser um fator primordial para a composição da amostra. Portanto, totalizaram-se 15 participações válidas.

**Tabela 5** - Dados demográficos dos participantes

Dado	Quantidade	Porcentagem
<b>Gênero</b>		
Masculino	14	82,40%
Feminino	3	17,60%
Outros (a informar)	0	0,00%
Prefiro não dizer	0	0,00%
<b>Idade</b>		
18-20	1	5,88%
21-25	11	64,71%
26-29	2	11,76%
30+	3	17,65%
<b>Maior nível de graduação</b>		
Superior Incompleto	12	70,60%
Superior Completo	2	11,80%
Mestrado Incompleto	2	11,80%

Doutorado Incompleto	1	5,90%
<b>Experiência no uso de tecnologias (0 a 5)</b>		
1	0	0%
2	0	0%
3	1	5,9%
4	3	17,60%
5	13	76,50%
<b>Frequência com que realiza experimentos controlados</b>		
Nenhuma	2	11,80%
Raramente	7	41,20%
Às vezes	6	35,30%
Muitas vezes	1	5,90%
Sempre	1	5,90%
<b>Frequência com que utiliza software para apoiar a realização de experimentos controlados</b>		
Nenhuma	3	17,60%
Raramente	6	35,30%
Às vezes	5	29,40%
Muitas vezes	1	5,90%
Sempre	2	11,8%

---

**Fonte:** elaborado pelo autor (2022)

## 6.2 Análise de confiabilidade

Para avaliar a credibilidade dos dados obtidos pelo experimento, será necessário avaliar a confiabilidade do instrumento de coleta de dados utilizado. Por confiabilidade, entende-se a consistência do instrumento ao reproduzir resultados semelhantes em circunstâncias diferentes quando reaplicado (ROBERTS; PRIEST, 2006).

A medida de consistência interna foi utilizada na análise de confiabilidade. Esta medida significa o relacionamento entre todos os resultados obtidos através do instrumento de coleta. Para medir a consistência interna, foi utilizado o coeficiente *Alfa de Cronbach* (ROBERTS; PRIEST, 2006).

Os resultados demonstrados na Tabela 5 indicam que os valores do coeficiente em relação às métricas variou entre 0.90 e 0.91. De acordo com (ROBERTS; PRIEST, 2006), tais valores são considerados satisfatórios. Os coeficientes foram obtidos com auxílio do software SPSS, versão 29.0.0.0.

**Tabela 6** - Medida do coeficiente de Alfa de Cronbach em relação às métricas avaliadas

Métrica	Alfa de Cronbach
Utilidade Percebida	0.916
Facilidade de Uso Percebida	0.900
Baixo Custo de Tempo	0.911
Facilidade de Uso Percebida Comparada	0.931
Utilidade Percebida Comparada	0.951

**Fonte:** elaborado pelo autor (2022)

### 6.3 Análise descritiva dos resultados das etapas

Os resultados demonstrados na Tabela 5 sugerem que houve uma aceitação considerável dos participantes, em relação às métricas avaliadas, sobre o sistema em todas as etapas do fluxo de criação do experimento e verificação do relatório estatístico. Portanto, há uma indicação de que o sistema é útil, fácil de usar e possui baixo custo de tempo dentro do fluxo proposto.

Segundo (BOONE; BOONE, 2012), antes da análise de dados de escalas Likert, deve-se escolher a escala de medida a ser utilizada. Para escalas cujas respostas são usadas para representar uma atitude, como são as métricas avaliadas no experimento, o autor recomenda o uso da escala de medida intervalar. Para essa escala de medida, o autor instrui o uso de medidas estatísticas como média e desvio padrão para a análise dos dados.

Para a representação visual dos dados através de gráficos, (ROBBINS; HEIBERGER, 2011) recomenda o uso de gráficos de barras divergentes. Entretanto, a interpretação de tal gráfico se mostrou confusa no conjunto de dados do experimento devido ao baixo número de respostas discordantes. Por conta disso, optou-se pelo uso de gráficos de barras empilhadas.

**Tabela 7** - Resultados do questionário por etapa

Etapa	Métrica	Média	Desvio Padrão
<b>1. Dar nome e detalhes ao experimento</b>			
1.1	Percepção de Utilidade	3.53	0.74
1.2	Percepção de Facilidade de Uso	3.80	0.56
1.3	Baixo custo de tempo	3.80	0.56
<b>2. Selecionar turmas participantes do experimento</b>			
2.1	Percepção de Utilidade	3.73	0.59
2.2	Percepção de Facilidade de Uso	3.73	0.70
2.3	Baixo custo de tempo	3.73	0.59
<b>3. Criar pedidos de permissão e convite</b>			
3.1	Percepção de Utilidade	3.67	0.62
3.2	Percepção de Facilidade de Uso	3.73	0.59
3.3	Baixo custo de tempo	3.80	0.41
<b>4. Criar pré-teste</b>			
4.1	Percepção de Utilidade	3.73	0.59
4.2	Percepção de Facilidade de Uso	3.80	0.56

4.3	Baixo custo de tempo	3.60	0.74
<b>5. Criar intervenção</b>			
5.1	Percepção de Utilidade	3.67	0.72
5.2	Percepção de Facilidade de Uso	3.87	0.35
5.3	Baixo custo de tempo	3.80	0.56
<b>6. Criar pós-teste</b>			
6.1	Percepção de Utilidade	3.87	0.35
6.2	Percepção de Facilidade de Uso	3.87	0.35
6.3	Baixo custo de tempo	3.67	0.62
<b>7. Selecionar método de atribuição randômica dos participantes nos grupos de pesquisa</b>			
7.1	Percepção de Utilidade	3.60	0.74
7.2	Percepção de Facilidade de Uso	3.80	0.56
7.3	Baixo custo de tempo	3.87	0.52
<b>8. Verificar o relatório final produzido</b>			
8.1	Percepção de Utilidade	3.80	0.56
8.2	Percepção de Facilidade de Uso	3.87	0.52
8.3	Baixo custo de tempo	3.87	0.52

---

**Fonte:** elaborado pelo autor (2022)

Em especial, a etapa de criação do pós-teste foi percebida como a mais útil. Entretanto, é preciso atentar-se para o fato de que a etapa de pré-teste, que obteve

médias menores, é idêntica à etapa de pós-teste, que obteve médias maiores. Possivelmente, os participantes atribuíram melhores notas à etapa de pós-teste pela familiaridade adquirida na etapa de pré-teste. Caso omitíssemos ambas as etapas citadas devido ao fator de familiaridade citado, a etapa percebida como mais útil se torna a etapa de verificação do relatório final.

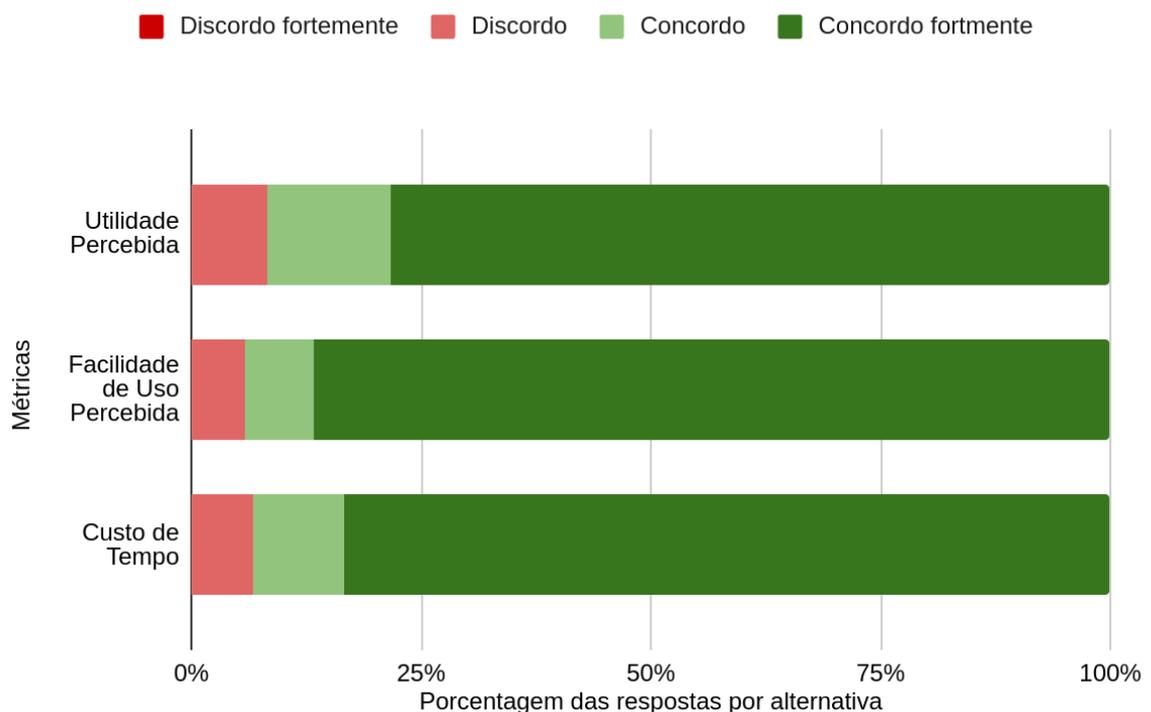
Em relação à métrica de Facilidade de Uso, as etapas de criação do pós-teste, criação da intervenção e verificação do relatório final foram percebidas como as mais fáceis de usar.

Já na métrica de Baixo Custo de Tempo, as etapas de seleção de método de atribuição randômica e verificação do relatório final foram percebidas como as de menor custo de tempo.

#### 6.4 Análise descritiva por métrica acumulada

Ao acumular as respostas de acordo com a métrica avaliada, obtém-se os resultados da Figura 1. Nela, os resultados complementam o tópico 5.2 ao sugerir, naturalmente, uma aceitação considerável dos participantes sobre o sistema em relação às métricas avaliadas.

**Gráfico 1** - Respostas por métrica acumulada dos questionários das etapas



**Fonte:** elaborado pelo autor (2022)

A métrica de Facilidade de Uso Percebida (Média = 3,83, Desvio Padrão = 0,5) obteve a melhor média, seguida por Custo de Tempo (Média = 3,79, Desvio Padrão = 0,53) e, por fim, Utilidade Percebida (Média = 3,74, Desvio Padrão = 0,59). Para a maior média na métrica de Facilidade de Uso Percebida, uma possível explicação pode ser o uso de recursos como formulários de multi-etapas com textos descritivos para cada etapa, da atenção às boas práticas na construção de interfaces e o uso, na construção da interface do sistema, do consolidado framework Bootstrap que oferece recursos valiosos como modais e estilos visuais para diversos elementos da interface, tornando-a mais intuitiva.

Tal resultado da métrica de Facilidade de Uso Percebida pode ter contribuído para a boa média da métrica de Baixo Custo de Tempo por conta da facilidade dos usuários em se familiarizar com a interface que, possivelmente, se converte em menos tempo gasto para realizar as atividades. Além disso, o sistema se encarrega de realizar diversas atividades custosas em relação ao tempo que, normalmente, seriam realizadas pelo pesquisador, como descrito no tópico 4.2 de Detalhamento do Processo. Isso contribui para uma economia considerável de tempo.

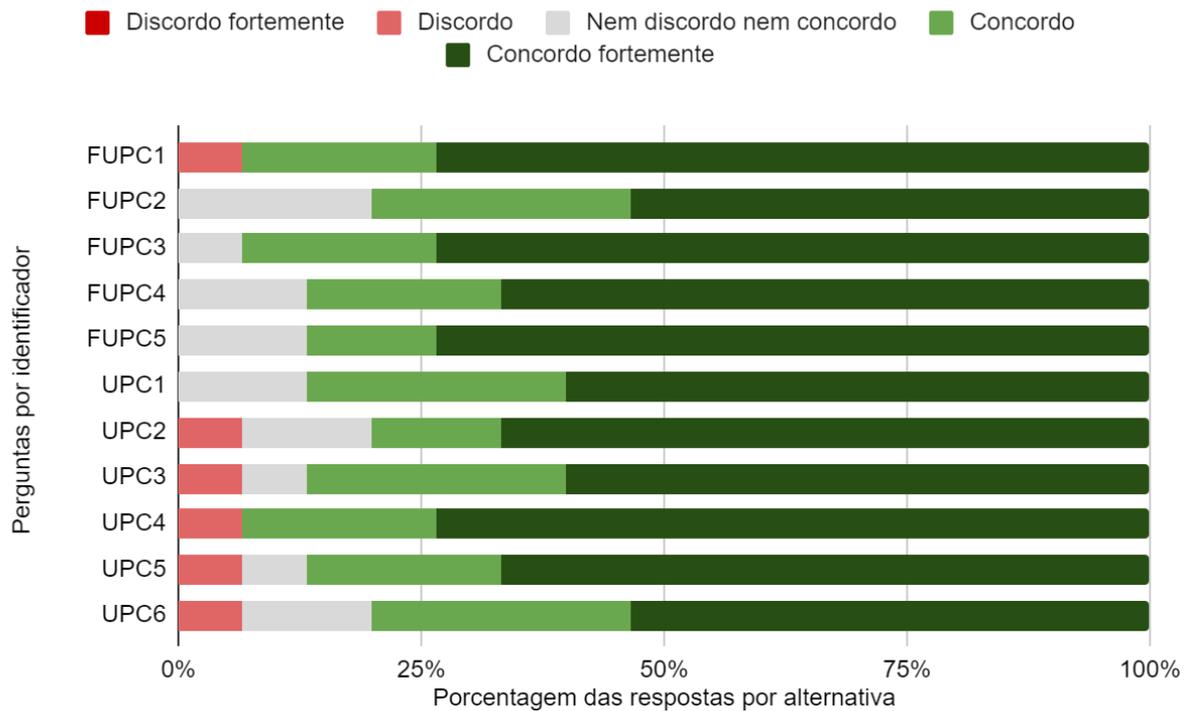
A métrica de Utilidade Percebida obteve a menor média, apesar de ainda poder ser considerada um resultado positivo. Uma possível explicação pode advir dos comentários deixados por alguns participantes no questionário sugerindo a inclusão de funcionalidades. Por exemplo, um dos participantes fez a seguinte sugestão: “O pré-teste poderia disponibilizar a opção de formular questões que avaliam construtos psicológicos por meio da escala de Likert de 1 a 4/ 1 a 5. E ainda, se possível, disponibilizar mais de um pré-teste, ou a possibilidade de criar duas ou mais etapas de pré-teste antes do início da intervenção”.

Este e outros comentários indicam que alguns participantes podem ter percebido o sistema como menos útil devido à ausência das funcionalidades sugeridas que não faziam parte do conjunto de funcionalidades essenciais do sistema proposto.

## **6.5 Resultados do questionário de comparação**

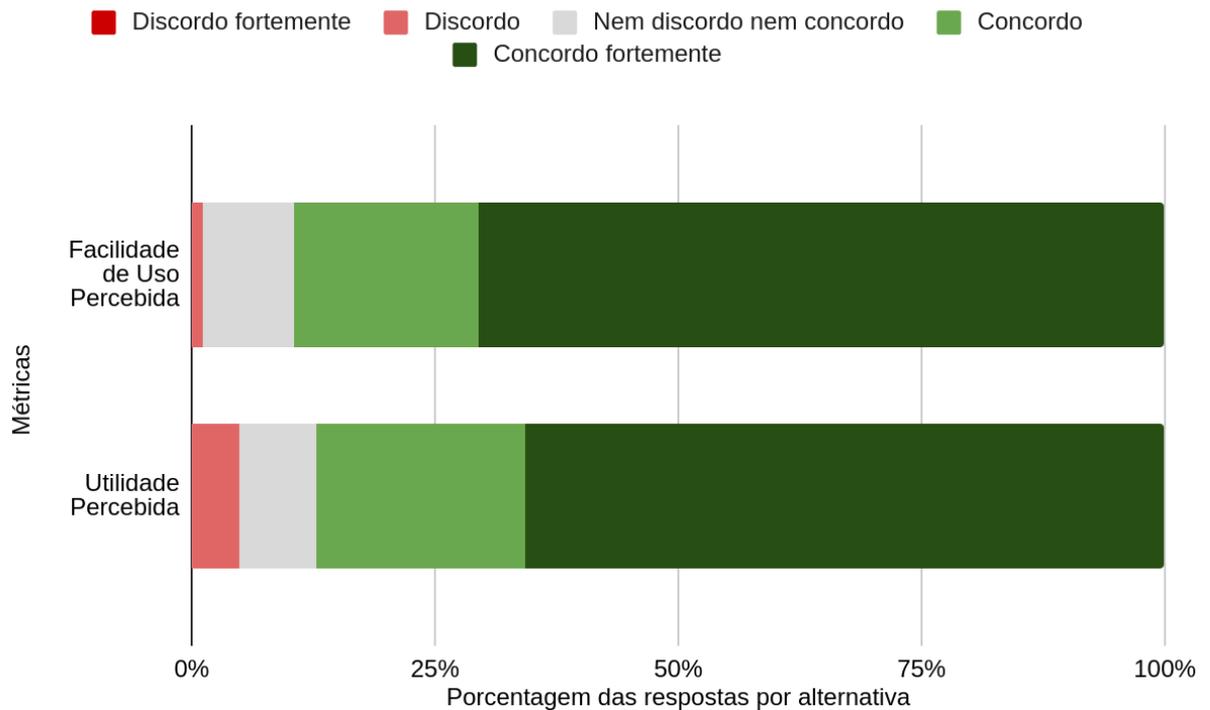
Os resultados demonstrados na Figura 2 sugerem que, em comparação com as ferramentas normalmente utilizadas pelos participantes para a realização de experimentos, o sistema se mostrou uma melhor alternativa em relação às métricas avaliadas. Os identificadores utilizados no gráfico estão presentes na Tabela 4.

**Gráfico 2** - Resultados por pergunta do questionário de comparação



Fonte: elaborado pelo autor (2022)

Ao acumular as respostas por métrica na Figura 3, os resultados se complementam, sugerindo que o sistema foi percebido como mais fácil de usar e mais útil se comparado com as ferramentas normalmente utilizadas pelos participantes para a realização de experimentos.

**Gráfico 3** - Resultados por métrica do questionário de comparação

Fonte: elaborado pelo autor (2022)

Uma possível explicação para estes resultados é a ausência de uma ferramenta criada especificamente para a realização de experimentos com design do tipo pré-teste e pós-teste que esteja disponível abertamente para pesquisadores. Isto implica no uso de ferramentas de propósito geral que, naturalmente, necessitam de um esforço maior para alcançar resultados comparáveis com os obtidos pelo sistema, que é uma ferramenta de propósito específico.

## 6.6 Análise inferencial dos resultados

### 6.6.1 Análise de normalidade

É necessário avaliar a normalidade dos dados antes de escolher os testes estatísticos necessários para avaliar as correlações e os testes de hipótese. Isto se dá porque tais testes fazem suposições acerca da normalidade dos dados. Para avaliar a normalidade, será utilizado o Teste de Shapiro-Wilk por ser mais indicado para amostras pequenas. (MISHRA et al., 2019).

A hipótese nula do Teste de Shapiro-Wilk afirma que os dados foram coletados de uma população normalmente distribuída. Se o valor-p obtido através do Teste de Shapiro-Wilk for menor que 0.05, a hipótese nula não pode ser rejeitada. Logo, conclui-se que os dados estão normalmente distribuídos (MISHRA et al., 2019).

Os resultados demonstrados na Tabela 6 indicam que todas as métricas possuem um valor-p inferior a 0.001. Logo, a hipótese nula é rejeitada e, portanto, conclui-se que os dados não estão normalmente distribuídos. Os resultados foram obtidos com auxílio do software SPSS, versão 29.0.0.0.

**Tabela 8** - Resultados do Teste de Shapiro-Wilk por métrica

Métrica	Teste de Shapiro-Wilk
Utilidade Percebida	<.001
Facilidade de Uso Percebida	<.001
Baixo Custo de Tempo	<.001
Utilidade Percebida Comparada	<.001
Facilidade de Uso Percebida Comparada	<.001

**Fonte:** elaborado pelo autor (2022)

### 6.6.2 Correlações

Para avaliar as associações entre as métricas analisadas, se faz necessário avaliar a correlação entre elas. Para isso, foi calculado o coeficiente de correlação Spearman. A escolha do coeficiente foi motivada pelo fato da suposição estatística da normalidade dos dados do coeficiente de Pearson não ter sido atendida, como recomendado em (GOGTAY; THATTE, 2017).

Houve uma correlação positiva, forte e significativa entre a métrica de Facilidade de Uso Percebida e a métrica de Utilidade Percebida (Coef. Correlação = 0.66,  $n = 120$ , valor-p < .001). Entre a métrica de Facilidade de Uso Percebida e Baixo Custo de Tempo, também houve uma correlação positiva, forte e significativa (Coef. Correlação = 0.61,  $n = 120$ , valor-p < .001). Por fim, entre a métrica de Utilidade Percebida e Baixo Custo de Tempo, houve uma correlação positiva,

moderada e significativa (Coef. Correlação = 0.54, n = 120, valor-p < .001). Os resultados foram obtidos com auxílio do software SPSS, versão 29.0.0.0.

Logo, pode-se afirmar que todas as métricas possuem correlações positivas, significativas e, no mínimo, moderadas entre si.

### 6.6.3 Teste de Hipóteses

Para avaliar as hipóteses definidas no capítulo 4.5, foi utilizado o Teste de Mann-Whitney U com base nas recomendações sugeridas em (NAYAK; HAZRA, 2011) onde considerou-se a natureza numérica, a não-normalidade e o não-pareamento dos dados. Para cada métrica, o teste foi realizado comparando dois grupos: respostas concordantes e respostas discordantes em relação à métrica, também definidos no capítulo 4.5. O nível de significância utilizado foi de 0.05. Os resultados foram obtidos com auxílio do software SPSS, versão 29.0.0.0.

Os resultados demonstrados na Tabela 7 indicam que houve uma diferença significativa entre as respostas concordantes e discordantes para todas as métricas. A quantidade de respostas concordantes foi maior do que a quantidade de respostas discordantes também em todas as métricas.

**Tabela 9** - Resultados dos Testes de Mann-Whitney U das métricas

Métricas	Respostas concordantes N	Respostas discordantes N	Teste de Mann-Whitney valor-p
Utilidade Percebida	110	10	<.001
Facilidade de Uso Percebida	113	7	<.001
Baixo Custo de Tempo	112	8	<.001
Utilidade Percebida Comparada	66	9	<.001
Facilidade de Uso Percebida Comparada	77	13	<.001

**Fonte:** elaborado pelo autor (2022)

## 7 CONCLUSÃO

A realização de experimentos controlados pode render evidências robustas o suficiente para basear práticas educacionais. Estas, por sua vez, podem trazer mais benefícios para a aprendizagem dos estudantes do que as práticas que não são baseadas em evidências. Entretanto, tais experimentos são custosos e difíceis de conduzir.

Existem ferramentas disponíveis no cenário nacional e internacional que buscam facilitar a realização de determinados experimentos, mas tais ferramentas, no caso das internacionais, possuem difícil acesso e, no caso das nacionais, não contemplam algumas funcionalidades pelo qual o presente trabalho preza. É nesse cenário que a presente pesquisa propôs um sistema que busca facilitar o trabalho do pesquisador na realização de experimentos com design do tipo pré-teste e pós-teste em ambientes online de aprendizagem.

Além disso, buscou-se avaliar se o sistema proposto é capaz de apoiar a realização de tais experimentos. Para isso, foi realizado um experimento onde 17 participantes percorreram um fluxo, através do sistema, para criação de um experimento e a verificação do relatório estatístico de seus resultados, compartilhando suas percepções sobre utilidade, facilidade de uso e custo de tempo para cada etapa do fluxo.

Os resultados do experimento indicam que **o sistema proposto é percebido, de forma estatisticamente significativa, como útil, fácil de usar e possuindo baixo custo de tempo quando apoia a realização de experimentos com design do tipo pré-teste e pós-teste em ambientes online de aprendizagem**, mesmo em comparação com ferramentas normalmente utilizadas por pesquisadores para o mesmo fim. Sendo assim, observa-se que os objetivos que o presente trabalho buscou alcançar foram alcançados ao propor, desenvolver e avaliar a ferramenta já descrita.

### 7.1 Limitações e ameaças à validade

O questionário utilizado no experimento de avaliação do sistema passou por diversas alterações com o objetivo de reduzir o esforço necessário do participante para completá-lo. Ao optar por avaliar cada etapa separadamente, sendo 8 ao total,

o questionário, naturalmente, se torna mais extenso. Esse cenário limita as possibilidades de construção de um questionário com maior escopo de avaliação devido à necessidade de atentar-se à ameaça de mortalidade dos participantes.

Outra limitação é a quantidade de participantes do experimento. Foram obtidas apenas 17 participações, onde 15 eram válidas. Esse número reduzido de participantes prejudica a análise dos resultados da presente pesquisa pela incapacidade da amostra de representar fielmente uma população de pesquisadores.

## **7.2 Trabalhos futuros**

A presente pesquisa buscou avaliar o sistema somente no cenário de realização de experimentos do tipo pré-teste e pós-teste. Entretanto, os designs de experimento são diversos. Novos trabalhos podem ser realizados para avaliar como o sistema se sai em diferentes cenários onde designs de experimentos além do design de pré-teste e pós-teste são utilizados. Isso pode aumentar o leque de possibilidades do sistema no apoio à realização de experimentos.

Além disso, outro ponto a ser notado é que o sistema foi proposto apenas para a construção de experimentos educacionais. Entretanto, um participante do experimento revelou possíveis funcionalidades que tornariam o sistema apto para uso em experimentos na área de psicologia, possivelmente para pesquisas comportamentais. A possibilidade de estender a capacidade de atuação do sistema além da pesquisa educacional pode ser explorada através de trabalhos futuros.

Por fim, após correções e aprimoramentos feitos com base no feedback dos participantes do experimento, o sistema deve ser utilizado para a realização de experimentos reais. Posteriormente, novos trabalhos podem ser realizados para avaliar como as descobertas realizadas por pesquisas feitas com auxílio do sistema impactam a produção científica da pesquisa educacional no contexto que o sistema se insere.

## REFERÊNCIAS

- ASSISTMENTS FOUNDATION; WORCESTER POLYTECHNIC INSTITUTE. ASSISTments. Disponível em: <<https://new.assistments.org/>>. Acesso em: 4 dez. 2022.
- BOMBASAR, J.; RAABE, A. SES: um sistema baseado na web para o design e a condução de experimentos educacionais. *Revista Brasileira de Informática na Educação*, v. 25, p. 30, 31 out. 2017.
- BOONE, H. N.; BOONE, D. A. Analyzing likert data. *Journal of extension*, v. 50, n. 2, p. 1–5, 2012.
- CAREY, S. S. *A beginner's guide to scientific method*. [s.l.] Cengage Learning, 2011.
- CONNOLLY, P.; KEENAN, C.; URBANSKA, K. The trials of evidence-based practice in education: A systematic review of randomised controlled trials in education research 1980–2016. *Educational Research*, v. 60, n. 3, p. 276–291, 2018.
- DAVIS, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, p. 319–340, 1989.
- FARIA, C. A. P. DE; SANCHES, A. E. *Mapeamento e caracterização do movimento das políticas públicas baseadas em evidências no Brasil*. 2022.
- GIAMATTEI, M. et al. LIONESS Lab: a free web-based platform for conducting interactive experiments online. *Journal of the Economic Science Association*, v. 6, n. 1, p. 95–111, 2020.
- GOGTAY, N. J.; THATTE, U. M. Principles of correlation analysis. *Journal of the Association of Physicians of India*, v. 65, n. 3, p. 78–81, 2017.
- GOPALAN, M.; ROSINGER, K.; AHN, J. B. Use of quasi-experimental research designs in education research: Growth, promise, and challenges. *Review of Research in Education*, v. 44, n. 1, p. 218–243, 2020.
- GURECKIS, T. M. et al. psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavior research methods*, v. 48, n. 3, p. 829–842, 2016.
- HEFFERNAN, N. T.; HEFFERNAN, C. L. The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, v. 24, n. 4, p. 470–497, 2014.
- KENDALL, J. M. Designing a research project: randomised controlled trials and their principles. *Emergency Medicine Journal*, v. 20, n. 2, p. 164, 1 mar. 2003.

KOHAVI, R. et al. Online randomized controlled experiments at scale: lessons and extensions to medicine. *Trials*, v. 21, n. 1, p. 150, 7 fev. 2020.

KOHAVI, R.; THOME, S. The Surprising Power of Online Experiments. *Harvard Business Review*, 1 set. 2017.

LEE, Y.; KOZAR, K. A.; LARSEN, K. R. T. The Technology Acceptance Model: Past, Present, and Future. *Communications of the Association for Information Systems*, v. 12, 2003.

LEVY, Y.; ELLIS, T. J. A guide for novice researchers on experimental and quasi-experimental studies in information systems research. *Interdisciplinary Journal of information, knowledge, and management*, v. 6, p. 151, 2011.

MAGALHÃES, C. V. et al. Caracterizando a pesquisa em informática na educação no Brasil: um mapeamento sistemático das publicações do SBIE. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). Anais...2013.*

MCCOMAS, W. F. The Principal Elements of the Nature of Science: Dispelling the Myths. Em: MCCOMAS, W. F. (Ed.). *The Nature of Science in Science Education: Rationales and Strategies*. Dordrecht: Springer Netherlands, 2002. p. 53–70.

MISHRA, P. et al. Descriptive statistics and normality tests for statistical data. *Annals of cardiac anaesthesia*, v. 22, n. 1, p. 67, 2019.

MOTZ, B. et al. Terracotta. Disponível em: <<https://www.terracotta.education/>>. Acesso em: 4 dez. 2022.

MURPHY, R. et al. Investigating efficacy, moderators and mediators for an online mathematics homework intervention. *Journal of Research on Educational Effectiveness*, v. 13, n. 2, p. 235–270, 2020.

NAYAK, B. K.; HAZRA, A. How to choose the right statistical test? *Indian journal of ophthalmology*, v. 59, n. 2, p. 85, 2011.

RITTER, S. et al. UpGrade: an open source tool to support A/B testing in educational software. *Proceedings of the First Workshop on Educational A/B Testing at Scale (at Learning@ Scale 2020)*. Anais...2020.

ROBBINS, N. B.; HEIBERGER, R. M. Plotting Likert and other rating scales. *Proceedings of the 2011 joint statistical meeting*. Anais...American Statistical Association, 2011.

ROBERTS, P.; PRIEST, H. Reliability and validity in research. *Nursing standard*, v. 20, n. 44, p. 41–46, 2006.

SANDIM, T. L.; MACHADO, D. A. O paradigma das políticas públicas baseadas em evidências na gestão pública brasileira: uma análise das publicações acadêmicas.

2020.

SCHANZENBACH, D. W. Limitations of experiments in education research. *Education Finance and Policy*, v. 7, n. 2, p. 219–232, 2012.

SEK, Y.-W. et al. Prediction of user acceptance and adoption of smart phone for learning with technology acceptance model. *Journal of Applied Sciences(Faisalabad)*, v. 10, n. 20, p. 2395–2402, 2010.

SIBBALD, B.; ROLAND, M. Understanding controlled trials. Why are randomised controlled trials important? *BMJ (Clinical research ed.)*, v. 316, n. 7126, p. 201, 17 jan. 1998.

SIROKER, D.; KOOMEN, P. *A/B testing: The most powerful way to turn clicks into customers*. [s.l.] John Wiley & Sons, 2013.

SLAVIN, R. E. *Evidence-Based Education Policies: Transforming Educational Practice and Research*. *Educational Researcher*, 2002. Disponível em: <<https://journals.sagepub.com/doi/abs/10.3102/0013189X031007015>>

SLAVIN, R. E. How evidence-based reform will transform research and practice in education. *Educational Psychologist*, v. 55, n. 1, p. 21–31, 2 jan. 2020.

STRATTON, S. J. Quasi-Experimental Design (Pre-Test and Post-Test Studies) in Prehospital and Disaster Research. *Prehospital and disaster medicine*, v. 34, n. 6, p. 573–574, dez. 2019.

SULLIVAN, G. M. Getting off the “gold standard”: randomized controlled trials and education research. *Journal of graduate medical education*, v. 3, n. 3, p. 285–289, set. 2011.

SURESH, K. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *Journal of human reproductive sciences*, v. 4, n. 1, p. 8–11, jan. 2011.

WHAT WORKS CLEARINGHOUSE. *WWC Review*. Disponível em: <<https://ies.ed.gov/ncee/wwc/Study/86375>>. Acesso em: 4 dez. 2022.

WHAT WORKS CLEARINGHOUSE. *Procedures and Standards Handbook, Version 5.0*. Disponível em: <<https://ies.ed.gov/ncee/wwc/handbooks>>. Acesso em: 4 dez. 2022.

WOLFE, C. R. Twenty years of Internet-based research at SCiP: A discussion of surviving concepts and new methodologies. *Behavior Research Methods*, v. 49, n. 5, p. 1615–1620, 1 out. 2017.