

Um estudo com o TRUEBLE: como mudanças de interface influenciam a percepção de estudantes sobre o apoio ao estudo de lógica proposicional

Eduarda M. Sifuentes², Rafael S. Da Silva¹,
João Pablo Silva da Silva², Ildevana Poltronieri¹, Alice Finger²

¹Universidade Federal do Pampa (UNIPAMPA), Campus Alegrete, Alegrete/RS

²Programa de Pós-Graduação em Engenharia de Software –
Universidade Federal do Pampa (UNIPAMPA)

{eduardasifuentes, rafaelstds3}.aluno@unipampa.edu.br

{joaosilva, ildevanarodrigues, alicefinger}@unipampa.edu.br

Abstract. *This study investigates the redesign of the “Learn” module of the TRUEBLE application as a tool to support the study of propositional logic. The original version was compared with a redesigned version featuring changes in content organization, navigation, and explanatory feedback. The evaluation involved eight Computing students through an A/B test with a crossover design, a post-task questionnaire, and semi-structured interviews. Results indicated more favorable perceptions of the redesigned version, particularly regarding card-based content organization, structured navigation, and feedback after errors. The findings suggest that design decisions can enhance the perceived support for studying, although they do not allow conclusions about actual learning gains.*

Resumo. *Este estudo investiga a reformulação do módulo “Aprender” do aplicativo TRUEBLE como apoio ao estudo de lógica proposicional. A versão original foi comparada a uma versão reformulada, com alterações na organização do conteúdo, na navegação e no feedback explicativo. A avaliação envolveu oito estudantes de Computação, por meio de teste A/B com delineamento cruzado, questionário pós-tarefa e entrevistas semiestruturadas. Os resultados indicaram percepção mais favorável da versão reformulada, com destaque para a organização em cards, a navegação estruturada e o feedback após erros. Os achados sugerem que decisões de design podem ampliar a percepção de apoio ao estudo, embora não permitam afirmar ganhos de aprendizagem efetiva.*

1. Introdução

O ensino de disciplinas fundamentais da Computação, como a lógica proposicional, permanece um desafio recorrente nos cursos de graduação [Silva et al. 2006]. A natureza abstrata dos conceitos envolvidos, especialmente na manipulação de estruturas formais e na construção de tabelas-verdade, tende a elevar o esforço cognitivo dos estudantes e a dificultar a consolidação dos fundamentos teóricos [Valente 2020]. Como consequência, observam-se, com frequência, baixo engajamento e dificuldades na compreensão conceitual desses conteúdos [Vasconcelos and Andrade 2018, Prietch and Pazeto 2010]. Nesse

cenário, ambientes digitais têm sido adotados como apoio ao estudo, com exemplos e exercícios interativos [Frezato et al. 2024, Lugli 2018].

Entretanto, a simples inserção de tecnologias no processo de ensino não garante ganhos pedagógicos [de Oliveira Martins et al. 2016]. Elementos como organização da informação, clareza de navegação e qualidade do *feedback* exercem papel decisivo na mediação da aprendizagem [de Oliveira Martins et al. 2016]. Nesse sentido, princípios de Interação Humano-Computador destacam que interfaces pouco intuitivas, fluxos confusos ou retornos restritos a certo ou errado podem aumentar o esforço necessário para interagir com o sistema, desviando a atenção do estudante do conteúdo conceitual [Barbosa and Silva 2010]. Sistemas interativos devem, portanto, favorecer clareza, consistência, reconhecimento dos elementos de interface e redução de barreiras de uso [Barbosa and Silva 2010, Norman 2013, Nielsen 2005].

O TRUEBLE é um aplicativo voltado ao ensino de lógica proposicional que inclui, entre outros recursos, o módulo “Aprender”, destinado à apresentação de conteúdos e à resolução de exercícios sobre semântica e tabelas-verdade. Avaliações anteriores identificaram limitações relacionadas à organização visual, à clareza de rótulos e à ausência de *feedback* explicativo após respostas incorretas, indicando oportunidades de aprimoramento do potencial pedagógico da ferramenta [Moura et al. 2024, Arrojo et al. 2024, Sifuentes et al. 2025]. A partir dessas evidências, foi realizada a reestruturação do fluxo de estudo desse módulo, envolvendo a reorganização dos tópicos, ajustes na hierarquia das informações e implementação de um mecanismo de *feedback* mais detalhado.

Diante desse contexto, este trabalho investiga em que medida a reformulação do módulo “Aprender” do TRUEBLE amplia seu potencial de apoio à aprendizagem de lógica proposicional, especialmente quanto à organização do conteúdo e ao *feedback* oferecido aos estudantes. Para isso, foi realizado um teste A/B com delineamento cruzado [Quin et al. 2024, Pavlovich and Kirynovich 2021], complementado por entrevistas semi-estruturadas, articulando dados quantitativos e qualitativos sobre a experiência de estudo nas duas versões [Rosala and Pernice 2023].

O estudo é orientado pelas seguintes questões de pesquisa: **QP1)** Em que medida a versão reformulada é percebida como mais adequada para apoiar o estudo de lógica proposicional em comparação à original? **QP2)** Quais aspectos da interface reformulada (organização do conteúdo, navegação e *feedback*) são percebidos como mais impactantes na experiência de estudo? **QP3)** Esse padrão de avaliação se mantém independentemente da ordem de utilização das versões?

O artigo está organizado da seguinte forma. A Seção 2 apresenta os trabalhos relacionados, com foco em estudos sobre usabilidade, design de interface e *feedback* em tecnologias educacionais. A Seção 3 descreve a abordagem metodológica e o protocolo de avaliação. A Seção 4 contextualiza o aplicativo TRUEBLE e apresenta a reformulação do módulo “Aprender”. A Seção 5 discute os resultados e sua interpretação. A Seção 6 apresenta as ameaças à validade. Por fim, a Seção 7 apresenta as considerações finais e os encaminhamentos para trabalhos futuros.

2. Trabalhos Relacionados

Esta seção revisa estudos sobre avaliação de tecnologias educacionais, com foco em usabilidade, design de interface e *feedback* no apoio à aprendizagem. A busca foi conduzida

de forma não sistemática na Biblioteca Digital da Sociedade Brasileira de Computação (SOL), com foco em trabalhos das áreas de Informática na Educação e Interação Humano-Computador que relacionassem interface, *feedback* e aprendizagem. Embora essa abordagem não garanta cobertura exaustiva da literatura, ela foi adotada intencionalmente para delimitar o escopo ao contexto nacional e complementar revisões sistemáticas mais amplas já disponíveis na área. A limitação decorrente da estratégia de busca é discutida na seção de ameaças à validade.

No contexto brasileiro, a revisão sistemática de [Santana et al. 2024] aponta a usabilidade como o critério técnico mais frequente na avaliação de softwares educacionais para o ensino de Computação, enquanto a eficácia da aprendizagem aparece como principal aspecto pedagógico. Os autores também destacam o uso recorrente de questionários e avaliações com usuários como instrumentos de coleta. Quanto aos métodos de avaliação, a literatura combina técnicas de inspeção e testes com usuários. O estudo de [Silva et al. 2021], por exemplo, utilizou avaliação heurística com base em Nielsen e testes de usabilidade para analisar um jogo sério na educação médica. Os resultados indicaram problemas de consistência visual, organização da informação e clareza do *feedback*, evidenciando que limitações de design podem comprometer tanto a experiência de uso quanto a interação do estudante com o conteúdo.

Estudos recentes também empregam métricas e instrumentos de usabilidade para examinar a qualidade da interação em sistemas educacionais. Em [Costa et al. 2024], abordagens estruturadas foram aplicadas à avaliação de um aplicativo voltado ao ensino de Libras, ressaltando a relevância de aspectos como simplicidade, segurança e atratividade da interface. Ainda que tratados sob a perspectiva da interação, esses elementos influenciam diretamente o acesso e a apropriação dos recursos de aprendizagem.

A literatura recente destaca que a qualidade de interfaces educacionais depende da redução de esforços desnecessários de interação e do estímulo ao engajamento. Enquanto [Macena et al. 2025] demonstram que a organização visual e textual influencia diretamente a usabilidade e a percepção de aprendizagem, [Andrade et al. 2014] reforçam que elementos interativos são fundamentais para despertar o interesse em conceitos científicos. Alinhando essas frentes, [Bandeira et al. 2019] propõem uma avaliação que integra dimensões técnicas e pedagógicas em conformidade com a BNCC, reiterando a necessidade de articular o design de interface aos objetivos educacionais.

Embora esses trabalhos contribuam para a compreensão da relação entre interface, usabilidade e aprendizagem, observa-se que a maioria das pesquisas concentra-se na avaliação de versões únicas de sistemas educacionais. Além disso, poucos estudos investigam explicitamente o papel do *feedback* como elemento pedagógico ou realizam comparações experimentais entre diferentes versões de uma mesma ferramenta.

A Tabela 1 evidencia que os trabalhos analisados concentram-se predominantemente na avaliação de interface, usabilidade e aceitação de versões únicas de sistemas educacionais. O termo “percepção de aprendizagem” refere-se à avaliação subjetiva dos participantes sobre sua própria aprendizagem, enquanto “apoio percebido ao estudo” corresponde à percepção de que a ferramenta auxilia a organização, a compreensão e a revisão dos conteúdos. Já “eficácia da aprendizagem” refere-se a evidências de aprendizagem reportadas pelos estudos primários sintetizados na revisão sistemática de literatura.

Tabela 1. Comparação entre os trabalhos relacionados e o presente estudo.

Trabalho	Avalia Interface	Avalia Feedback	Compara Versões	Foco Educacional Avaliado
[Macena et al. 2025]	Sim	Não	Não	Percepção de aprendizagem
[Santana et al. 2024]	Sim	Não	Não	Eficácia da aprendizagem (RSL)
[Silva et al. 2021]	Sim	Sim	Não	Aspectos didáticos
[Costa et al. 2024]	Sim	Não	Não	Não avaliado
[Andrade et al. 2014]	Sim	Não	Não	Motivação e engajamento
[Bandeira et al. 2019]	Sim	Sim	Não	Competências da BNCC
Este estudo	Sim	Sim	Sim	Apoio percebido ao estudo

Observa-se ainda que nenhum dos trabalhos analisados combina comparação entre versões, avaliação de *feedback* explicativo e delineamento experimental cruzado. Embora alguns estudos investiguem interface, *feedback* ou aspectos pedagógicos de forma isolada, não foram identificados trabalhos que analisassem experimentalmente como alterações na organização do conteúdo, na navegação e no *feedback* de um mesmo software educacional influenciam a percepção dos estudantes durante o processo de estudo.

Essa lacuna motiva o presente trabalho, que emprega um teste A/B com delineamento cruzado para comparar duas versões do módulo “Aprender” do TRUEBLE. Diferentemente dos estudos identificados na literatura, a proposta combina avaliação de interface, análise do *feedback* explicativo e comparação entre versões, articulando dados quantitativos e qualitativos para compreender como decisões de design influenciam a percepção de apoio ao estudo de lógica proposicional.

3. Metodologia

Este estudo configura-se como uma pesquisa aplicada, de abordagem mista (quantitativa e qualitativa) e objetivos descritivos e exploratórios [Prodanov and de Freitas 2013]. O foco é comparar as versões original e reformulada do módulo “Aprender” do TRUEBLE, investigando como mudanças na organização do conteúdo, na navegação e no *feedback* são percebidas pelos estudantes em termos de apoio ao estudo de lógica proposicional. A descrição detalhada do aplicativo e da reformulação avaliada é apresentada na Seção 4.

A avaliação foi realizada por meio de um teste A/B com delineamento cruzado, no qual os participantes interagiram com ambas as versões em ordens alternadas [Quin et al. 2024, Pavlovich and Kirynovich 2021]. Participaram oito estudantes dos cursos de Ciência da Computação e Engenharia de Software, matriculados ou egressos da disciplina de Lógica Matemática. Todos foram informados sobre os objetivos da pesquisa, os procedimentos de coleta, a participação voluntária e o anonimato dos dados, formalizando sua concordância por meio do Termo de Consentimento Livre e Esclarecido (TCLE). A amostra foi constituída por conveniência. A Figura 1 apresenta o delineamento metodológico do estudo.

Antes da aplicação definitiva, foi realizado um teste piloto com dois estudantes de Engenharia de Software sem contato prévio com o TRUEBLE, com o objetivo de verificar a clareza das instruções, a adequação das tarefas e o funcionamento geral do protocolo [Krug 2010]. O piloto resultou apenas em ajustes pontuais na redação das instruções. No teste A/B, os participantes foram divididos em dois grupos: o primeiro utilizou inicialmente a versão original e, em seguida, a reformulada; o segundo realizou o percurso inverso. Essa estratégia buscou reduzir possíveis efeitos de ordem, aprendiza-

Figura 1. Delineamento metodológico do estudo.



Fonte: gerada por inteligência artificial.

gem e familiarização, permitindo que cada participante atuasse como seu próprio controle.

As sessões foram presenciais e moderadas, seguindo um roteiro padronizado de tarefas. Os participantes deveriam localizar o tópico de semântica, ler o conteúdo, explicar sua compreensão, retornar ao tópico, resolver o exercício e refazê-lo após uma resposta incorreta. Após o uso de cada versão, os participantes responderam a um questionário pós-tarefa composto por doze itens fechados em escala Likert de cinco pontos e duas questões abertas [South et al. 2022]. As perguntas aplicadas são apresentadas na Tabela 2.

Tabela 2. Perguntas do questionário pós-tarefa.

Item	Pergunta
P01	Com pouco tempo de uso, consegui entender como o aplicativo funciona.
P02	Durante o uso, o funcionamento do aplicativo estava claro.
P03	O design da interface facilitou minha concentração no conteúdo.
P04	Foi fácil localizar e retornar ao conteúdo que eu procurava.
P05	As instruções de cada exercício estavam claras.
P06	Senti-me seguro(a) ao interagir com os exercícios.
P07	Foi fácil refazer o exercício após uma resposta incorreta.
P08	Os rótulos, ícones e botões eram legíveis e autoexplicativos.
P09	O conteúdo foi apresentado de forma clara.
P10	Os exemplos presentes no material me ajudaram a entender os conceitos.
P11	Sinto que o aplicativo pode ser utilizado como suporte para o estudo de conteúdos sobre lógica matemática.
P12	No geral, eu gostei desta versão do módulo.
P13	Você encontrou algo que tenha sido confuso, frustrante ou pouco claro? Se sim, descreva.
P14	Você tem alguma sugestão para melhorar a experiência de aprendizagem?

Após o teste A/B, os participantes participaram de uma entrevista semiestruturada [Rosala and Pernice 2023] para aprofundar percepções sobre organização dos conteúdos, navegação, *feedback* e percepção de progresso. O roteiro utilizado é apresentado na Tabela 3.

Os dados quantitativos foram analisados pelas medianas dos itens P01–P12 e pelo escore global de cada participante, correspondente à média dos itens fechados. As versões

Tabela 3. Perguntas da entrevista com foco em aspectos de aprendizagem.

Item	Pergunta
E01	Como foi sua experiência ao realizar as atividades nas duas versões, como ler o conteúdo, resolver e refazer exercícios?
E02	O que você percebeu de diferente na forma como os materiais foram apresentados entre as versões?
E03	A navegação e a organização dos tópicos influenciaram sua capacidade de relacionar ou conectar os conteúdos?
E04	Quando você errou um exercício, que tipo de retorno o sistema apresentou e como foi para você lidar com isso?
E05	Em qual versão foi mais fácil perceber seu progresso ou saber em que ponto do estudo você estava?
E06	Se você tivesse que continuar estudando usando apenas uma das versões, qual escolheria e por quê?

foram comparadas pelo teste de Wilcoxon para amostras pareadas [Wilcoxon 1945], com tamanho de efeito calculado pela correlação rank-biserial. A consistência interna foi estimada pelo alfa de Cronbach. Devido à amostra reduzida, os resultados foram interpretados exploratoriamente e integrados aos dados qualitativos.

As respostas abertas e as entrevistas foram analisadas por meio de análise temática [Braun and Clarke 2006], envolvendo familiarização com os dados, codificação inicial, agrupamento em temas e revisão dos temas identificados. A codificação foi realizada por uma pesquisadora e revisada pelos demais autores. Essa análise permitiu identificar categorias relacionadas à navegação e organização do conteúdo, *feedback* após erros, percepção de progresso e conforto visual. Por fim, os resultados quantitativos e qualitativos foram integrados para responder aos objetivos do estudo.

4. O Aplicativo TRUEBLE e a Reformulação do Módulo Aprender

Estudos anteriores [de Abreu Dias and Finger 2020] identificaram que aplicativos disponíveis na Google Play Store, embora eficientes na geração automática de tabelas-verdade, apresentavam limitações em relação a abordagens pedagógicas, recursos explicativos e interatividade voltados à aprendizagem. Diante desse cenário, o TRUEBLE foi concebido como uma solução educacional integrada, combinando conteúdo teórico, prática guiada e verificação de resultados, com foco na compreensão conceitual progressiva e no desenvolvimento do raciocínio lógico.

O aplicativo foi estruturado de forma modular, contemplando diferentes etapas do processo de ensino-aprendizagem. O módulo “Aprender” é destinado à apresentação de conteúdos e à resolução de exercícios sobre semântica e tabelas-verdade. O módulo “Construir Tabelas” é voltado à elaboração prática das tabelas-verdade, permitindo a inserção de fórmulas lógicas e o preenchimento manual das colunas, favorecendo a reflexão sobre operadores e proposições. Um terceiro módulo, denominado “Verificar Tabelas”, foi previsto no aplicativo, mas não chegou a ser desenvolvido. A Figura 2 apresenta a interface da versão original do TRUEBLE, com destaque para a tela inicial, o módulo “Aprender” e o módulo “Construir Tabelas”.

As evidências empíricas reportadas em avaliações anteriores realizadas com estudantes, especialmente nos trabalhos de [Moura et al. 2024, Arrojo et al. 2024], fundamentaram a necessidade de reestruturação do TRUEBLE e orientaram as intervenções de design analisadas neste estudo. Também foi realizada uma análise heurística da interface, conduzida por um avaliador com base nas heurísticas de usabilidade propostas por [Nielsen 2005], a fim de aprofundar os problemas observados quanto à organização visual, à clareza de rótulos, ao fluxo de navegação e ao *feedback* oferecido aos estudantes.

A partir desse diagnóstico, procedeu-se à reengenharia dos requisitos do sis-

Figura 2. Interface da versão original do TRUEBLE.**(a) Tela inicial e módulo “Aprender”.****(b) Módulo “Construir Tabelas”.**

Fonte: elaboração dos autores.

tema, priorizando aspectos diretamente relacionados ao apoio ao estudo, como *feedback* formativo, clareza instrucional, organização do conteúdo e consistência visual. Considerando que ambientes educacionais digitais devem articular recursos tecnológicos e estratégias pedagógicas para apoiar o estudo e a progressão do estudante [de Oliveira Martins et al. 2016], a intervenção concentrou-se no módulo “Aprender”, responsável pela mediação inicial e pela apresentação dos conteúdos.

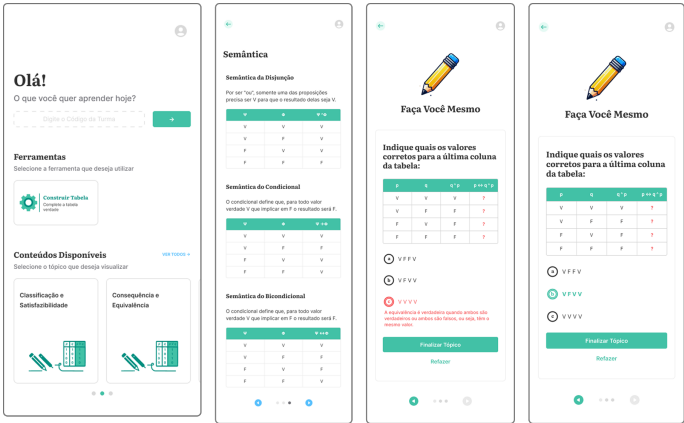
O antigo módulo “Aprender” foi reformulado, dando origem à seção “Conteúdos Disponíveis”. A Figura 3 apresenta a proposta de reformulação e evidencia a nova estrutura instrucional adotada.

Nessa nova organização, os tópicos de lógica proposicional passaram a ser apresentados em *cards* temáticos que articulam explicações teóricas, exemplos e exercícios de fixação. Cada *card* incorpora um indicador de progresso, possibilitando ao estudante acompanhar seu avanço e favorecendo a percepção de progressão no estudo. Além disso, o *feedback* após respostas incorretas passou a apresentar explicações mais detalhadas, buscando apoiar a revisão dos erros e a compreensão dos conceitos envolvidos [Hattie and Timperley 2007, Shute 2008].

5. Resultados e Discussão

Esta seção apresenta os resultados obtidos a partir dos questionários pós-tarefa, das respostas abertas e das entrevistas semiestruturadas. A discussão articula os dados quantitativos e qualitativos para interpretar como os participantes perceberam as diferenças entre a versão original e a versão reformulada do módulo “Aprender”.

Figura 3. Protótipo de alta fidelidade do módulo “Aprender” reformulado.



Fonte: elaboração dos autores.

5.1. Perfil dos participantes

Participaram do estudo oito estudantes de graduação da área de Computação, sendo quatro de Ciência da Computação e quatro de Engenharia de Software. Quanto à faixa etária, três participantes tinham entre 18 e 20 anos, quatro entre 21 e 23 anos e um entre 24 e 26 anos. Em relação ao gênero, seis se identificaram como masculino e duas como feminino. A Tabela 4 sintetiza o perfil dos participantes, identificados de P01 a P08 e agrupados conforme a ordem de utilização das versões no delineamento cruzado.

Tabela 4. Perfil dos participantes.

Part.	Grupo	Curso	Faixa etária	Conhecimento prévio
P01	A→B	Ciência da Computação	21–23 anos	Intermediário
P02	A→B	Engenharia de Software	24–26 anos	Básico
P03	A→B	Ciência da Computação	21–23 anos	Intermediário
P04	A→B	Ciência da Computação	21–23 anos	Intermediário
P05	B→A	Engenharia de Software	18–20 anos	Intermediário
P06	B→A	Engenharia de Software	18–20 anos	Intermediário
P07	B→A	Engenharia de Software	21–23 anos	Intermediário
P08	B→A	Ciência da Computação	18–20 anos	Intermediário

Quanto ao conhecimento prévio em lógica matemática, sete participantes declararam possuir nível intermediário e um nível básico. Esse perfil sugere familiaridade suficiente para avaliar a clareza dos conteúdos e exercícios, ao mesmo tempo em que representa um público potencialmente beneficiado por recursos de apoio ao estudo.

Em relação ao uso de recursos digitais para aprendizagem, quatro participantes afirmaram utilizá-los sempre, três frequentemente e um às vezes. As ferramentas mais mencionadas foram vídeos explicativos no YouTube (oito participantes) e livros ou PDFs digitais (sete participantes). Apesar desse uso recorrente, quatro participantes relataram dificuldade em encontrar materiais digitais de qualidade sobre lógica matemática, enquanto todos avaliaram os recursos encontrados como apenas razoáveis.

O predomínio de participantes com conhecimento intermediário e o uso frequente de recursos digitais sugerem que os estudantes possuíam repertório para identificar problemas de organização, navegação e *feedback*, bem como para comparar a experiência proporcionada pelo TRUEBLE com outras formas de estudo digital.

5.2. Análise quantitativa

Considerando os oito participantes em conjunto, todos apresentaram escore global superior para a versão reformulada em comparação à versão original. O escore global corresponde à média das respostas aos itens P01 a P12 do questionário pós-tarefa. Os escores variaram entre 2,08 e 4,33 na versão original, com mediana de 3,42, e entre 4,17 e 5,00 na versão reformulada, com mediana de 4,71. A Tabela 5 apresenta os escores médios globais por participante e as diferenças pareadas entre as versões.

Tabela 5. Escores médios globais por participante e diferenças pareadas.

Part.	Grupo	Escore orig.	Escore ref.	Dif.	Dir.
P01	A→B	2,58	4,92	+2,33	B
P02	A→B	2,08	4,17	+2,08	B
P03	A→B	3,50	4,83	+1,33	B
P04	A→B	4,33	4,67	+0,33	B
P05	B→A	4,33	4,75	+0,42	B
P06	B→A	3,42	4,25	+0,83	B
P07	B→A	3,42	4,58	+1,17	B
P08	B→A	2,83	5,00	+2,17	B
Mediana	–	3,42	4,71	+1,25	8/8

O teste de Wilcoxon para amostras pareadas, aplicado sobre os escores globais, indicou, em caráter exploratório, diferença favorável à versão reformulada ($W = 0$; $p = 0,008$, bilateral). O tamanho de efeito rank-biserial obtido foi $r = 1,00$, valor que reflete o fato de que todas as oito diferenças individuais foram positivas e ocorreram na mesma direção. Esse resultado decorre da unanimidade direcional das avaliações — e não implica ausência de variação interna entre os participantes, uma vez que as diferenças individuais variaram entre +0,33 e +2,33. Com $n = 8$ participantes, $r = 1,00$ deve ser interpretado como indicador da consistência direcional das respostas na amostra, e não como evidência generalizável de efeito máximo em nível populacional. Considerando o número reduzido de participantes, os dados sugerem uma percepção sistematicamente mais positiva da versão reformulada, interpretação que deve ser confirmada em estudos com amostras maiores.

Para verificar possíveis efeitos de ordem, os escores foram observados separadamente por grupo. O padrão favorável à versão reformulada ocorreu tanto no grupo que utilizou primeiro a versão original quanto no grupo que utilizou primeiro a versão reformulada. No Grupo 1, a mediana da diferença foi de +1,33, enquanto no Grupo 2 foi de +1,17. As Figuras 4a e 4b ilustram as medianas por item em cada grupo. A direção das avaliações não se inverteu em nenhum dos grupos, o que reforça a consistência do padrão observado.

A análise por dimensão permitiu organizar os itens do questionário de forma mais sintética. Os itens P01, P02 e P08 foram agrupados na dimensão clareza e usabilidade geral; os itens P03, P04 e P09 na dimensão navegação e organização do conteúdo; os itens P05, P06 e P07 na dimensão suporte à realização de exercícios; os itens P10 e P11 na dimensão percepção de apoio ao estudo; e o item P12 foi considerado como satisfação geral. A Tabela 6 apresenta as medianas por dimensão.

As dimensões com maior diferença foram navegação e organização do conteúdo, suporte à realização de exercícios e satisfação geral, indicando que a reformulação foi percebida principalmente como melhoria na experiência de interação e no apoio à realização

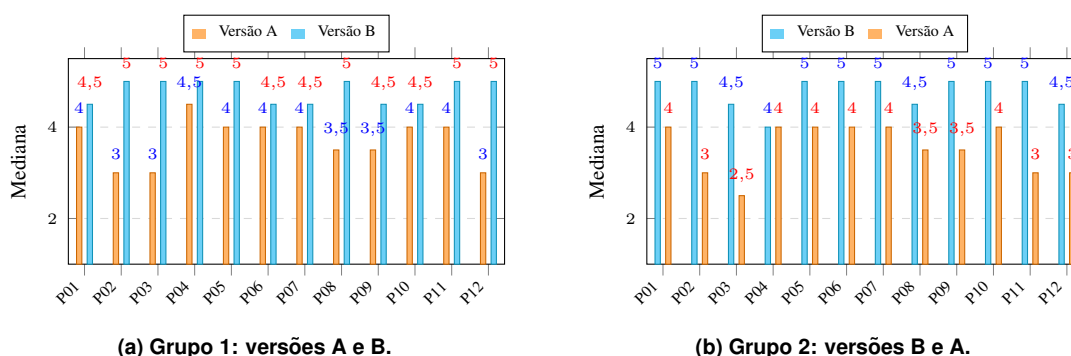


Figura 4. Comparação das medianas por item entre as versões, por grupo.

Tabela 6. Medianas por dimensão no conjunto dos participantes.

Dimensão	Mediana original	Mediana reformulada	Dif.
Clareza e usabilidade geral	3,5	5,0	+1,5
Navegação e organização do conteúdo	3,0	5,0	+2,0
Suporte à realização de exercícios	3,0	5,0	+2,0
Percepção de apoio ao estudo	4,0	5,0	+1,0
Satisfação geral	3,0	5,0	+2,0

das atividades. A dimensão percepção de apoio ao estudo apresentou diferença menor (+1,0), pois a versão original já possuía mediana relativamente alta (4,0), sugerindo que o aplicativo já era percebido como útil para o estudo e que a nova organização ampliou essa percepção de forma mais incremental. Cabe observar que a versão reformulada apresentou mediana igual a 5,0 em todas as dimensões avaliadas. Esse resultado indica uma percepção fortemente favorável, mas também sugere possível efeito teto, reduzindo a sensibilidade do instrumento para discriminar diferenças mais sutis entre as dimensões. Em estudos futuros, instrumentos com maior número de itens por dimensão ou escalas complementares poderão captar nuances da experiência que esta análise não foi capaz de distinguir.

No que diz respeito à consistência interna do questionário pós-tarefa, considerando os itens P01 a P12, o alfa de Cronbach foi calculado separadamente para cada versão. Para a versão original, obteve-se $\alpha = 0,903$, enquanto para a versão reformulada obteve-se $\alpha = 0,775$. Ambos os valores situam-se acima do limiar convencional de 0,70 frequentemente adotado em estudos exploratórios, indicando consistência interna satisfatória do instrumento nas duas condições. Ainda assim, esses resultados devem ser interpretados com cautela em razão do número reduzido de participantes.

5.3. Análise qualitativa

As respostas abertas e as entrevistas reforçam os resultados quantitativos. A partir da análise temática, os relatos foram organizados em quatro categorias: organização e navegação; *feedback* e revisão de erros; percepção de progresso; e limitações visuais e funcionais. Para preservar o anonimato, os participantes são identificados de P01 a P08, conforme a Tabela 4.

Na categoria organização e navegação, os participantes relataram dificuldades na versão original relacionadas à compreensão de símbolos, à organização dos conteúdos e à condução do estudo. P07 descreveu a experiência com a versão original como “*cansativa, uma vez que não era intuitiva, e a interface era confusa*”, enquanto P06

caracterizou o layout dessa versão como frustrante. Em contraste, a versão reformulada foi percebida como mais organizada, com informações hierarquizadas e fluxo mais intuitivo. P03 destacou que o encadeamento dos conteúdos “*faz mais sentido e agrega para a experiência de aprendizado*”. Esses achados dialogam com a literatura de Interação Humano-Computador, que destaca a importância de interfaces claras, consistentes e organizadas para favorecer a compreensão e reduzir barreiras de uso [Barbosa and Silva 2010, Norman 2013, Nielsen 2005].

Na categoria *feedback* e revisão de erros, o retorno após respostas incorretas foi mencionado por todos os participantes como uma das principais diferenças entre as versões. Na versão original, o *feedback* limitava-se à indicação de acertos e erros; P03, por exemplo, relatou que a exibição da quantidade de respostas corretas somada à presença do botão de nova tentativa gerava certa confusão sobre o próprio desempenho. Na versão reformulada, o *feedback* passou a incluir explicações logo abaixo das respostas e a possibilidade de refazer o exercício. P08 descreveu o retorno como “*objetivo, com uma explicação que ajudava a raciocinar a resposta da questão*” e sugeriu, ainda, uma explicação complementar mais aprofundada para os casos de erro. Esse resultado reforça a importância de conceber o *feedback* não apenas como validação da resposta, mas como recurso de mediação pedagógica [Hattie and Timperley 2007, Shute 2008, de Oliveira Martins et al. 2016].

Na categoria percepção de progresso, relatos pontuais indicaram que o indicador de progresso presente nos *cards* temáticos da versão reformulada facilitou a percepção de avanço durante o estudo. Na versão original, participantes como P07 relataram incerteza sobre o andamento das atividades, associada à navegação tela a tela. Por fim, na categoria limitações visuais e funcionais, P07 relatou desconforto visual associado ao uso predominante de áreas claras na interface, especialmente em períodos prolongados de uso. P06 mencionou um problema pontual no *feedback*, em que uma resposta considerada correta foi indicada como incorreta pelo sistema, e P07 também apontou dificuldade inicial para localizar conteúdos e retornar entre telas no primeiro contato com a versão reformulada. Esses achados indicam que a versão reformulada apresenta avanços relevantes, mas ainda demanda ajustes visuais e funcionais antes de avaliações longitudinais em sala de aula.

6. Ameaças à Validade

Este estudo apresenta limitações que devem ser consideradas na interpretação dos resultados. Em relação à validade interna, embora o delineamento cruzado com contrabalanceamento reduza possíveis efeitos de ordem, não elimina completamente efeitos de aprendizagem, familiarização ou contraste entre as versões. A experiência com a primeira versão pode ter influenciado a avaliação da segunda, especialmente porque os participantes executaram tarefas semelhantes nas duas condições. Além disso, as sessões presenciais e moderadas podem ter influenciado as respostas aos instrumentos.

Quanto à validade de constructo, o estudo avalia percepção de apoio ao estudo e experiência de uso, e não aprendizagem efetiva. Os itens do questionário contemplam aspectos como clareza, organização, *feedback* e satisfação, mas não mensuram diretamente aquisição, retenção ou transferência de conhecimento. O escore global foi calculado a partir de itens Likert, exigindo interpretação cautelosa, embora seu uso seja adequado para análises exploratórias. Adicionalmente, a análise heurística foi conduzida por um

único avaliador, e a análise temática não contou com codificação dupla independente nem cálculo formal de concordância entre avaliadores. As avaliações elevadas atribuídas à versão reformulada também sugerem possível efeito teto, o que pode ter reduzido a capacidade do instrumento de distinguir diferenças mais sutis entre as dimensões avaliadas.

Em relação à validade externa, a amostra foi composta por apenas oito estudantes de uma única instituição, recrutados por conveniência, o que limita a generalização dos resultados. O efeito de novidade da versão reformulada também não pode ser descartado. Além disso, a busca por trabalhos relacionados foi conduzida de forma não sistemática, podendo não contemplar estudos relevantes de outros contextos ou idiomas. Avaliações futuras, com amostras maiores e aplicação em contextos reais de ensino, são necessárias para verificar se as percepções observadas se mantêm ao longo do tempo e se se relacionam com ganhos de desempenho.

7. Considerações Finais

A partir dos resultados obtidos, conclui-se que a reformulação do módulo “Aprender” do aplicativo TRUEBLE ampliou seu potencial de apoio pedagógico ao priorizar a organização estruturada dos conteúdos e o *feedback* explicativo após os erros. Os resultados quantitativos e qualitativos apontam um padrão de percepções mais favorável à versão reformulada, especialmente quanto à clareza do fluxo de navegação, à organização do conteúdo e ao apoio à revisão das atividades.

Respondendo às questões de pesquisa, em relação à **QP1**, a versão reformulada foi percebida como mais adequada para apoiar o estudo de lógica proposicional, apresentando mediana global superior à versão original e avaliações favoráveis em todos os participantes. Em relação à **QP2**, os aspectos mais valorizados foram a organização do conteúdo em *cards*, a navegação mais estruturada e o *feedback* explicativo após erros, percebidos como facilitadores do estudo e da compreensão dos exercícios. Por fim, quanto à **QP3**, o padrão favorável à versão reformulada foi observado nos dois grupos do delineamento cruzado, sem evidências de efeito de ordem relevante na amostra estudada.

Nesse sentido, o trabalho contribui para a área de Informática na Educação ao evidenciar que decisões de design relacionadas à organização do conteúdo, à navegação e ao *feedback* podem ampliar a percepção de apoio pedagógico em aplicações voltadas ao ensino de conteúdos abstratos em Computação. Além disso, os relatos dos participantes indicam que aspectos visuais e funcionais, como conforto visual e consistência do *feedback*, também influenciam a experiência de estudo.

Como limitação, destaca-se que o estudo foi realizado com número reduzido de participantes e concentrou-se na percepção dos estudantes sobre o uso das duas versões, não permitindo afirmar ganhos de aprendizagem efetiva. Como trabalhos futuros, pretende-se incorporar correções e aprimoramentos no aplicativo, especialmente quanto ao conforto visual, ao uso de cores, ao problema pontual identificado no *feedback* e à flexibilidade dos modos de visualização. Posteriormente, planeja-se conduzir uma avaliação longitudinal em ambiente de sala de aula, envolvendo turmas de diferentes cursos da área de Computação, com o objetivo de analisar como a ferramenta apoia o estudo em contextos reais de ensino e explorar possíveis relações com o desempenho acadêmico dos estudantes.

Cuidados Éticos

A pesquisa contou com participação voluntária de estudantes, que foram informados sobre os objetivos do estudo, os procedimentos de coleta e assinaram o Termo de Consentimento Livre e Esclarecido (TCLE). Os dados foram tratados de forma agregada e anonimizada, sem identificação individual dos participantes. Por se tratar de uma atividade avaliativa em contexto educacional, sem coleta de dados sensíveis ou intervenções de risco, o estudo foi conduzido em conformidade com os princípios éticos aplicáveis à pesquisa educacional.

Declaração sobre uso de Inteligência Artificial

Neste trabalho, utilizou-se a ferramenta ChatGPT como apoio à organização textual, à geração de código para gráficos e tabelas e à produção da figura que ilustra o delineamento metodológico, a partir de informações fornecidas e conferidas pelos autores. As figuras das interfaces do aplicativo foram elaboradas pelos próprios autores, sem uso de geração automática por inteligência artificial. A coleta, a organização, a análise e a interpretação dos dados, bem como a revisão final do texto, permaneceram sob responsabilidade integral dos autores.

Agradecimentos

Os autores agradecem pelas bolsas de Iniciação Científica concedidas pelo CNPq e pela FAPERGS.

Referências

- Andrade, B., Matos, P., Carvalho, V., Santos, W., and Santos, A. (2014). Avaliação da aceitabilidade do software educacional sr. bohr como ferramenta lúdica para o ensino-aprendizagem. In *Anais do XX Workshop de Informática na Escola*, pages 96–105, Porto Alegre, RS, Brasil. SBC.
- Arrojo, M., Dias, B., Poltronieri, I., and Finger, A. (2024). Trueble: Desenvolvimento e avaliação de um app para o ensino de tabelas verdade. In *XXXV Simpósio Brasileiro de Informática na Educação (SBIE)*. SBC.
- Bandeira, L., Araújo, N., Farias, F., Barros, A., Queiroz, J., Nunes, I., and Guedes, L. (2019). Instrumento de avaliação do software educacional tinkercad: uma visão fundamentada na bncc. In *Anais do XXV Workshop de Informática na Escola*, pages 1324–1328, Porto Alegre, RS, Brasil. SBC.
- Barbosa, S. and Silva, B. (2010). *Interação Humano-Computador*. Elsevier Brasil.
- Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101.
- Costa, A. P., Diniz, E., Sales, E., Coelho, F., Costa, W. C., and Sousa, L. (2024). Usabilidade de um aplicativo de libras para educação e governança em uma cidade amazônica. In *Anais do XXXV Simpósio Brasileiro de Informática na Educação*, pages 1248–1262, Porto Alegre, RS, Brasil. SBC.
- de Abreu Dias, B. and Finger, A. F. (2020). Aplicativos para o ensino-aprendizagem de lógica matemática: qual a melhor escolha? In *Anais do XXVI Workshop de Informática na Escola (WIE 2020)*, pages 111–120, Porto Alegre, RS, Brasil. Sociedade Brasileira de Computação (SBC).

- de Oliveira Martins, D., Tiziotto, S. A., and Cazarini, E. W. (2016). Ambientes virtuais de aprendizagem (avas) como ferramentas de apoio em ambientes complexos de aprendizagem (acas). *Revista Brasileira de Aprendizagem Aberta e a Distância*, 15.
- Frezato, O., Coleti, T., and Valle, P. (2024). Avaliação heurística de um ambiente gamificado para ensino de programação. In *Anais do XXXII Workshop sobre Educação em Computação*, pages 806–817, Porto Alegre, RS, Brasil. SBC.
- Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1):81–112.
- Krug, S. (2010). *Rocket Surgery Made Easy: The Do-It-Yourself Guide to Finding and Fixing Usability Problems*. New Riders, Berkeley, CA.
- Lugli, L. C. (2018). Prototipagem de soluções tecnológicas, alfabetização matemática na educação infantil e deficiência sensorial: parametrização de características assistivas. Technical report, Universidade Estadual Paulista (Unesp).
- Macena, J., Pires, F., Pessoa, M., and Oliveira, E. (2025). Avaliação da interface de um jogo educacional baseado na aprendizagem multimídia. In *Anais do XXIV Simpósio Brasileiro de Jogos e Entretenimento Digital*, pages 477–488, Porto Alegre, RS, Brasil. SBC.
- Moura, G., Poltronieri, I., and Finger, A. (2024). Trueble: Prototipação e avaliação do aplicativo para ensino de tabelas verdade. In *XXXV Simpósio Brasileiro de Informática na Educação (SBIE)*. SBC.
- Nielsen, J. (2005). 10 usability heuristics for user interface design. Acessado em maio de 2025.
- Norman, D. A. (2013). *The Design of Everyday Things*. MIT Press, 15th edition.
- Pavlovich, Y. G. and Kirynovich, I. F. (2021). A/b testing as an effective instrument for adaptation of the user interface at the iterative model of developing applications for mobile devices. *Doklady BGUIR*, 19(1):30–36.
- Prietch, S. S. and Pazeto, T. A. (2010). Estudo sobre a evasão em um curso de licenciatura em informática e considerações para melhorias. *WEIBASE, Maceió/AL*.
- Prodanov, C. C. and de Freitas, E. C. (2013). *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico*. Feevale, Novo Hamburgo, 2 edition. [Recurso eletrônico].
- Quin, F., Weyns, D., Galster, M., and Costa Silva, C. (2024). A/b testing: A systematic literature review. *Journal of Systems and Software*, 211:112011.
- Rosala, M. and Pernice, K. (2023). Entrevistas com usuários: O básico. Artigo em português (Brasil).
- Santana, R., Rizzi, C., and Boscarioli, C. (2024). Avaliação de software educacional para o ensino e aprendizagem de conceitos introdutórios de programação: Uma revisão sistemática da literatura. In *Anais do XXXV Simpósio Brasileiro de Informática na Educação*, pages 2339–2352, Porto Alegre, RS, Brasil. SBC.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1):153–189.

- Sifuentes, E., Dias, E., Tremarin, G., Moura, G., Poltronieri, I., and Finger, A. (2025). Reconstrução de aplicativo de lógica baseada na reengenharia funcional e pedagógica. In *Anais do XXXVI Simpósio Brasileiro de Informática na Educação*, pages 1206–1218, Porto Alegre, RS, Brasil. SBC.
- Silva, F. S. C. d., Finger, M., and Melo, A. C. V. d. (2006). *Lógica para Computação*. Cengage Learning, São Paulo, 2 edition.
- Silva, M., Freire, A., Graciano, M., Greggi, J., and Pereira, M. (2021). Avaliação heurística interdisciplinar e testes de usabilidade de um jogo sério em educação médica. In *Anais do XXXII Simpósio Brasileiro de Informática na Educação*, pages 630–641, Porto Alegre, RS, Brasil. SBC.
- South, L., Saffo, D., Vitek, O., Dunne, C., and Borkin, M. A. (2022). Effective use of likert scales in visualization evaluations: A systematic review. *Computer Graphics Forum*, 41(3). EuroVis 2022.
- Valente, M. T. (2020). Engenharia de software moderna. *Princípios e Práticas para Desenvolvimento de Software com Produtividade*, 1(24).
- Vasconcelos, V. and Andrade, E. (2018). Análise da evasão de alunos na licenciatura em computação. In *Workshop sobre Educação em Computação (WEI)*. SBC.
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.