

Avaliação Automatizada da Criatividade de Aplicativos Móveis no Contexto Educacional

Nathalia da Cruz Alves
Universidade Federal de Santa Catarina
Florianópolis/Santa Catarina/Brasil
nathalia.alves@posgrad.ufsc.br

Christiane Gresse von Wangenheim
Universidade Federal de Santa Catarina
Florianópolis/Santa Catarina/Brasil
c.wangenheim@ufsc.br

RESUMO

A criatividade é uma habilidade importante do século 21, que pode ser desenvolvida como parte do ensino de computação. Uma das formas de fomentar a criatividade é por meio do ensino do desenvolvimento de artefatos computacionais, como aplicativos móveis. Embora existam diversos modelos de medição da criatividade, a avaliação da criatividade de aplicativos móveis no ensino de computação permanece relativamente inexplorada, sendo que a maioria dos modelos existentes dependem de uma avaliação manual por humanos. Apesar de a avaliação humana ser importante, ela nem sempre contempla todos os aspectos relevantes e pode ser suscetível a vieses, preferências e conhecimentos pessoais. Assim, este artigo apresenta um modelo analítico e automatizado para avaliar a criatividade de aplicativos móveis. De acordo com a definição da criatividade, o modelo avalia a originalidade, a flexibilidade e a fluência. Resultados de análises estatísticas indicam a confiabilidade e a validade do modelo. Espera-se assim contribuir para o avanço da avaliação da criatividade no ensino de computação por meio de um modelo de avaliação consistente, que pode ser complementado com a avaliação humana, permitindo uma avaliação holística beneficiando tanto educadores quanto estudantes.

CCS CONCEPTS

• **Social and professional topics** → Computing education.

PALAVRAS-CHAVE

Criatividade, Avaliação de aprendizagem, Aplicativo móvel, Ensino de computação.

1 INTRODUÇÃO

Na sociedade 4.0, a criatividade é considerada uma das principais competências do século 21, essencial para o sucesso profissional

e pessoal [10][20]. Consequentemente, desenvolver a criatividade dos estudantes tornou-se uma preocupação dominante [39], especialmente devido às transformações impulsionadas pela Inteligência Artificial, que indicam que profissões que demandam um elevado nível de criatividade têm menor probabilidade de serem automatizadas [35]. Neste contexto, é desejável promover o desenvolvimento da criatividade desde a educação básica, a fim de preparar os estudantes para se tornarem cidadãos capazes de enfrentar os desafios do século 21 [7].

A criatividade é definida como a interação entre a aptidão, o processo e o ambiente pelo qual um indivíduo ou um grupo produz um artefato perceptível que é novo e útil conforme definido dentro de um contexto social [31]. Embora a criatividade seja tradicionalmente associada às artes, música e literatura, ela também pode ser cultivada em outras áreas do conhecimento, como a computação, uma disciplina fundamentalmente criativa [11].

No contexto do ensino de computação, a criatividade é uma das habilidades essenciais para enfrentar os desafios no desenvolvimento de sistemas de software, assim como para preparar os estudantes para aprender a criá-los, testá-los e aprimorá-los [34]. Embora a criatividade possa ser analisada com base em diferentes aspectos [32], no contexto da computação, a análise do artefato de software criado pelo aluno é uma alternativa para se estimular a criatividade. Especialmente, considerando que, independentemente do que os professores desejem que os estudantes aprendam, as percepções dos estudantes sobre o que é significativo, e, portanto, onde e como direcionar seus esforços, são fortemente moldadas pela lente da avaliação [11]. Assim, uma das formas de promover a criatividade dentro do ensino de computação é avaliando-a a partir dos artefatos de software produzidos pelos estudantes.

Apesar da importância de medir e avaliar a criatividade no contexto do ensino de computação, existem poucas abordagens para medir a criatividade de artefatos de software [2]. As abordagens existentes enfocam principalmente em resultados de atividades de programação bem definidas [8][18][23][27]. O uso de atividades de programação bem definidas com vários níveis de dificuldade é uma abordagem educacional comumente empregada para promover a aprendizagem estruturada. Esse método permite que os alunos avancem gradualmente em complexidade, estabelecendo assim uma base sólida de conhecimento. No entanto, uma limitação potencial dessa abordagem é que ela pode impedir os alunos de explorar suas

Fica permitido ao(s) autor(es) ou a terceiros a reprodução ou distribuição, em parte ou no todo, do material extraído dessa obra, de forma verbatim, adaptada ou remixada, bem como a criação ou produção a partir do conteúdo dessa obra, para fins não comerciais, desde que sejam atribuídos os devidos créditos à criação original, sob os termos da licença CC BY-NC 4.0.

EduComp'24, Abril 22-27, 2024, São Paulo, São Paulo, Brasil (On-line)

© 2024 Copyright mantido pelo(s) autor(es). Direitos de publicação licenciados à Sociedade Brasileira de Computação (SBC).

ideias e desenvolver sistemas de software com base em seus interesses e, portanto, limitar sua expressão da criatividade.

Uma alternativa às tarefas bem definidas é adotar uma abordagem de aprendizagem baseada em problemas, na qual os alunos desenvolvem projetos de livre escolha. Nesses cenários, a avaliação, geralmente, segue uma abordagem baseada no desempenho, em que o artefato de software que representa o resultado do processo de aprendizagem dos alunos é utilizado como objeto de avaliação [33]. Apesar do reconhecimento da importância de avaliar a criatividade em relação ao desenvolvimento de artefatos de software a partir de projetos livres, faltam abordagens válidas e confiáveis que forneçam suporte sistemático na definição, execução e análise da criatividade de artefatos de software produzidos por estudantes como parte do ensino de computação [2].

A maioria das abordagens atuais para medir e avaliar a criatividade em projetos de computação de livre escolha são baseadas em critérios subjetivos [2]. Nessas abordagens, os instrutores avaliam artefatos computacionais, usando, por exemplo, uma escala ordinal, que inclui descritores como “não muito novo”, “alguma novidade” ou “muito novo” [5]. Tipicamente, vários especialistas participam desta avaliação, muitas vezes chegando a um consenso (por exemplo, [5], em alinhamento com a Técnica de Avaliação Consensual [3][17]). No entanto, estas técnicas podem consumir muitos recursos, especialmente em cenários que envolvem um grande número de artefatos de software para avaliar. Isso é especialmente um problema em maratonas de programação e concursos de aplicativos, nos quais criadores de todo o mundo podem submeter os seus artefatos de software. Embora a avaliação de especialistas forneça resultados contextualmente relevantes na avaliação da criatividade, fornecer uma avaliação consistente e oportuna pode não ser viável na prática [6]. Consequentemente, esses fatores restringem a eficácia das abordagens manuais como única opção de avaliação. A este respeito, a medição automatizada da criatividade na computação tem valor não apenas como um avanço independente, mas também na promoção de pesquisas mais extensas na intersecção da criatividade e da computação.

Assim, este artigo apresenta resultados de uma avaliação da confiabilidade e validade do modelo analítico Creassessment [1] para avaliação da criatividade de artefatos de software com foco em aplicativos móveis criados como resultados de aprendizagem no contexto do ensino de computação. O modelo utiliza técnicas automatizadas de extração de dados e analisa dimensões da criatividade conforme definidas na literatura da criatividade.

2 TRABALHOS RELACIONADOS

Existem algumas abordagens que exploram a avaliação automatizada de indicadores de criatividade de artefatos de software [8][18][22][23][24][25][26][27][29][37]. Essas abordagens enfatizam principalmente a originalidade dos artefatos de computação criados com diversas plataformas de programação, como AgentSheets [8][23], Kodetu [13][18], Scratch [22][24][25], App Inventor [29][37] e JavaScript [27]. Algumas abordagens são personalizadas para atividades bem

definidas com uma solução correta predefinida [13][18][26][27]. Por outro lado, ao contrário de atividades com soluções pré-determinadas, em que a corretude é fundamental, algumas abordagens visam analisar indicadores de criatividade de artefatos de software resultantes de projetos livres [22][24][25][29][37], reforçando a importância de exibir um alto grau de proficiência na geração de soluções criativas em uma variedade de contextos.

Embora as abordagens existentes enfatizem principalmente a originalidade [22][29][37], é essencial ampliar os critérios de avaliação para abranger o conceito de criatividade de forma mais abrangente. Nesse sentido, a abordagem predominante para avaliar a criatividade no domínio geral, geralmente, envolve o uso de testes de pensamento divergente. Esses testes tipicamente abrangem três dimensões: fluência - a capacidade de gerar uma infinidade de ideias, flexibilidade - a habilidade de fazer a transição entre diversas categorias ou perspectivas, e originalidade - a capacidade de criar ideias distintas e não convencionais [21]. Um exemplo de avaliação da criatividade do artefato de software, inspirado em testes de pensamento divergente, é o modelo de avaliação proposto por Kovalkov et al. [24][25], o qual mede fluência, flexibilidade e originalidade em relação ao código, ao visual e ao áudio de artefatos de software criados com Scratch. Apesar de proverem um modelo de avaliação da criatividade de artefatos de software, este é voltado ao ambiente de programação Scratch, e, assim, não permite sua aplicação em outros ambientes e linguagens de programação. Considerando outros tipos de artefatos de software, especificamente aplicativos móveis, que se tornaram parte integrante da vida moderna, este artigo apresenta uma abordagem para avaliar de forma analítica e automatizada a criatividade de aplicativos móveis buscando fechar essa lacuna do estado da arte.

3 MODELO

O modelo Creassessment é utilizado como instrumento para mensurar a criatividade de aplicativos móveis. Considerando as características idiossincráticas de aplicativos, são definidos quatro aspectos de análise: componentes, conteúdo textual, funcionalidades e programação (Tabela 1).

Tabela 1: Definição de aspectos de aplicativos móveis

Aspecto	Descrição
Componentes	Elemento que atende a uma finalidade de interface do usuário (componente visível, como um botão) ou a uma finalidade funcional (componente invisível, como um sensor).
Conteúdo textual (tags)	Conjunto de 10 palavras-chave extraídas do conteúdo textual do aplicativo.
Funcionalidades	Ações e serviços que o aplicativo oferece aos seus usuários.
Programação	Um elemento visual arrastável que representa um comando de programação (como if-then-else).

Seguindo a definição clássica da literatura para o pensamento divergente [15], tipicamente, utilizado como forma de avaliação da criatividade, o modelo Creassessment é composto

por três dimensões: fluência, flexibilidade e originalidade. Considerando a definição clássica da literatura, a definição de cada dimensão é adaptada para o contexto de avaliação de um artefato de software (aplicativo móvel), considerando suas características idiossincráticas:

A fluência refere-se à capacidade de usar um grande número de componentes (botões, mapas, etc.) e comandos de programação (if, for, while, etc.). É medida pela contagem da quantidade total de componentes e comandos de programação empregados em um aplicativo.

A flexibilidade refere-se à capacidade de adaptar componentes e comandos de programação de diferentes tipos para resolver um problema. A flexibilidade envolve uma categorização prévia de cada componente (interface de usuário, sensores, etc.) e comandos de programação (condicional, laços, etc.). É pontuada pela contagem do número de categorias de componentes, comandos de programação e funcionalidades empregadas no aplicativo.

A originalidade refere-se ao uso de componentes de interface de usuário (botão, caixa de texto, etc.), funcionalidades (fazer login, tocar som, etc.) e conteúdo textual identificado por

meio de palavras-chave ('plantas', 'emprego', etc.) incomuns vistas em um universo de aplicativos feitos por pessoas com experiência e formação semelhantes. É medida pelo inverso da frequência que aparece no universo de aplicativos desenvolvidos em um contexto similar de ensino de computação.

Todos os itens são medidos em uma escala de 0 (insuficiente) a 10 (excelente). O modelo conceitual de avaliação foi implementado e automatizado por meio de um módulo Python [1] e integrado em uma ferramenta web (Figura 1) disponível gratuitamente online [1][38].

4 MÉTODOS

O objetivo deste estudo é analisar de forma exploratória o modelo Creassessment [1], a fim de estimar sua confiabilidade e validade de constructo para a medição da criatividade no contexto do ensino de computação. Adotando a abordagem *Goal Question Metric* [4], são derivadas as seguintes questões de análise:

QA1. Há evidência de consistência interna do modelo?

QA2. Há evidência de validade do modelo?

Coleta de dados: Foi utilizado um conjunto de dados que

Aplicativo Healthy Plants



O aplicativo permite buscar por uma planta e ver informações sobre a planta.

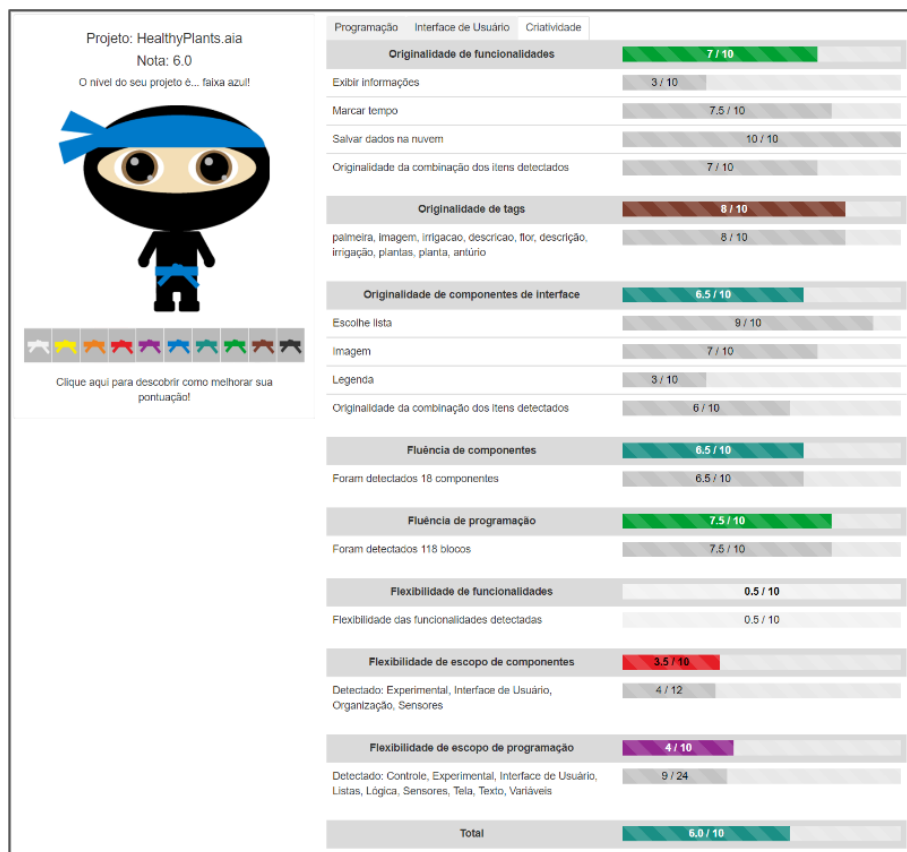


Figura 1: Exemplo de avaliação de um aplicativo [38]

consiste em aplicativos submetidos ao concurso Aplicativo do Mês (*App of the Month*) de 2016 a 2022, disponibilizados pela equipe da Fundação App Inventor/MIT [28]. O Aplicativo do Mês trata-se de um concurso mensal, que ocorreu até 2022, no qual criadores de aplicativos de todas as idades podiam enviar seus aplicativos para serem avaliados por especialistas. Os aplicativos submetidos são feitos no App Inventor - um ambiente de programação baseado em blocos. Durante o período de 6 anos foram feitas 1.923 submissões para o concurso por pessoas de todo o mundo, das quais 1.494 submissões foram únicas, isto é, não duplicadas, e forneceram um *link* para download do aplicativo. Usando um script Python para baixar os aplicativos por meio do compartilhamento de dados pela equipe da Fundação App Inventor/MIT, foram obtidos os códigos-fontes de 1.078 aplicativos. Alguns aplicativos não puderam ser baixados porque o *link* fornecido estava desatualizado e o aplicativo não estava mais disponível. Usando os arquivos dos aplicativos como entrada, foram executadas as avaliações dos 1.078 aplicativos utilizando o modelo analítico e automatizado Creassessment [1]. Foram obtidas as pontuações totais para cada dimensão de avaliação da criatividade e uma pontuação geral da criatividade por meio da média aritmética de todas as pontuações. A Figura 1 apresenta um exemplo de avaliação de um aplicativo com suas respectivas notas usando a ferramenta CodeMaster [38], a qual implementa o modelo Creassessment [1].

Análise dos dados. Os dados contendo as notas obtidas para todas as dimensões (fluência, flexibilidade e originalidade) são analisados em relação a sua confiabilidade e validade. A confiabilidade é analisada em termos de consistência interna por meio do ω (coeficiente ômega) [17]. Diferentemente do coeficiente alfa de Cronbach, comumente utilizado, o coeficiente ω trabalha com as cargas fatoriais, o que torna os cálculos mais estáveis, com um nível de confiabilidade maior e de forma independente do número de itens no instrumento [12]. A validade é analisada via análise fatorial [14] de forma a determinar quantos fatores são subjacentes ao modelo Creassessment. Isso permite identificar em que medida cada item está correlacionado com cada subdimensão por meio de cargas fatoriais.

5 RESULTADOS

5.1 Há evidência de consistência interna do modelo?

A consistência interna foi analisada por meio do ω . Segundo a literatura, um valor entre 0,7 e 0,8 indica confiabilidade aceitável do conjunto de fatores, entre 0,8 e 0,9 confiabilidade boa, e um valor acima de 0,9 indica confiabilidade excelente. Como resultado, obteve-se uma confiabilidade boa de $\omega = 0,86$.

Ao analisar se a confiabilidade aumenta com a eliminação de um item, todos os itens, exceto o item relacionado a conteúdo textual, diminuem o coeficiente ômega se forem eliminados (Tabela 2). Uma razão para o aumento do valor ômega caso a análise da originalidade do conteúdo textual seja eliminada pode ser que a originalidade do conteúdo textual não se correlacione bem com outros itens porque o conjunto de *tags* extraído de um

aplicativo pode ser praticamente qualquer conjunto de palavras. Feita essa ressalva, em geral, esses resultados fornecem uma indicação da consistência interna do modelo.

Tabela 2: Coeficiente ômega ao excluir itens

Dimensão	Aspecto	ω se o item for excluído
Originalidade	Funcionalidades	0,84
	Componentes de interface de usuário	0,84
	Conteúdo textual	0,89
Fluência	Componentes	0,84
	Programação	0,83
Flexibilidade	Componentes	0,82
	Programação	0,81
	Funcionalidades	0,82

5.2 Há evidência de validade do modelo?

Para identificar o número de fatores subjacentes que influenciam as variáveis do modelo Creassessment, foi verificada a possibilidade de realizar uma análise fatorial exploratória usando o índice Kaiser-Meyer-Olkin (KMO). KMO mede a adequação da amostragem com valores entre 0 e 1. Um valor próximo a 1,0 suporta uma análise fatorial e qualquer valor menor que 0,5 provavelmente não é adequado para uma análise fatorial útil [9]. Foi obtido um índice KMO de 0,77, demonstrando que a análise fatorial é adequada neste caso.

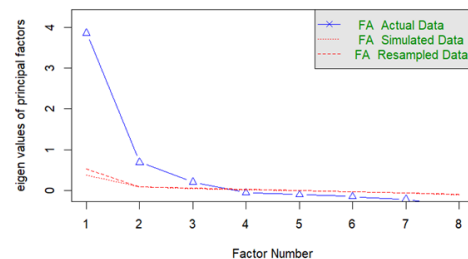


Figura 2: Scree Plots de Análise Paralela

A análise exploratória da matriz de correlação comparada com matrizes aleatórias paralelas indica a existência de 3 fatores na amostra, que são representados pelos triângulos em azul acima da linha pontilhada vermelha (Figura 2). Para identificar se o modelo se ajusta bem aos dados, foi realizada uma análise fatorial exploratória, na qual as estatísticas são pouco sensíveis ao tamanho da amostra. Os valores para avaliar a qualidade do ajuste do modelo são a raiz quadrada média do erro de aproximação (RMSEA), o índice de ajuste comparativo (CFI) e o índice de Tucker-Lewis (TLI). O ajuste é considerado adequado quando $RMSEA < 0,05$, $TLI > 0,90$ e $CFI > 0,90$ [9].

Tabela 3: Resultados para ajuste de análise fatorial exploratória

Número de fatores	RMSEA	TLI	CFI
1	0,2158	0,7847	0,8462
2	0,1453	0,9025	0,9547
3	0,0983	0,9553	0,9888

Os índices de ajuste sugerem que um fator não se ajusta bem aos dados. Um valor RMSEA mais alto indica um ajuste relativamente ruim, pois valores mais próximos de zero são desejáveis. Além disso, tanto o TLI quanto o CFI estão abaixo de 0,9, sugerindo um ajuste abaixo do ideal para um fator. Os índices de ajuste para dois fatores indicam um ajuste melhor em comparação com um fator. O valor RMSEA é menor e os valores de TLI e CFI estão acima de 0,9, sugerindo um ajuste razoavelmente bom para dois fatores. Três fatores mostram o melhor ajuste entre o número de fatores. Cabe ressaltar que o valor RMSEA é o mais baixo, mas ainda acima de 0,05, indicando um ajuste relativamente bom. Já os valores de TLI e CFI estão acima de 0,9, com CFI se aproximando de 1, indicando um ajuste muito bom para três fatores (Tabela 3).

Com base nos resultados fornecidos, foi realizada uma análise fatorial exploratória com três fatores para explorar as dimensões subjacentes representadas pelo conjunto de variáveis observadas (Tabela 4). A carga fatorial reflete a força da relação entre o item e o fator latente que se pretende medir. Para determinar um valor de corte para uma carga fatorial que indica que um item carregou bem em um fator, não existe um padrão universalmente aceito. No entanto, uma regra conservadora comumente usada é que valores maiores que 0,50 são considerados necessários para significância prática [16]. Isso significa que um item com uma carga fatorial de 0,50 ou superior é considerado adequado para medir o fator subjacente.

Tabela 4: Resultados da análise fatorial exploratória com três fatores

Dimensão	Item	F1	F2	F3
Originalidade	Funcionalidades	0,9316	-0,0752	-0,0450
	Componentes de Interface de Usuário	0,5270	-0,1280	0,2539
	Conteúdo textual	-0,1071	-0,1298	0,4185
Fluência	Componentes	0,0825	-0,0756	0,8894
	Programação	-0,0396	0,4706	0,6499
Flexibilidade	Componentes	0,7196	0,2771	-0,0303
	Programação	0,3364	0,7047	0,0973
	Funcionalidades	0,9164	0,0490	0,0725

O primeiro fator (F1) parece mesclar as dimensões de originalidade e flexibilidade. Destaca-se o item originalidade de funcionalidades com alta carga positiva de 0,9316. Isso indica uma forte associação entre este item e a dimensão subjacente representada pelo Fator 1. O item originalidade de componentes de interface do usuário também apresenta uma carga positiva moderada de 0,5270, sugerindo uma relação relativamente mais fraca com o Fator 1. Por outro lado, o item conteúdo textual possui uma carga negativa de -0,1071, indicando que está negativamente associado ao Fator 1. Os itens flexibilidade de componentes e flexibilidade de funcionalidades também apresentam carga positiva alta de 0,7196 e 0,9164 respectivamente. Essas altas cargas fatoriais podem estar relacionadas à estrutura de dados, por exemplo, flexibilidade de componentes e originalidade componentes de interface do usuário estão relacionados ao aspecto de componentes em geral

(veja Tabela 1), assim como a originalidade de funcionalidades e flexibilidade de funcionalidades estão fortemente correlacionadas.

Em relação ao segundo fator (F2), a flexibilidade de programação apresenta uma carga positiva alta de 0,7047. A fluência de programação também apresenta uma carga positiva moderada de 0,4706, implicando uma associação significativa com esse fator. Por outro lado, a originalidade de componentes da interface do usuário e o conteúdo textual têm cargas negativas de -0,1280 e -0,1298, respectivamente, sugerindo uma relação fraca com o Fator 2.

O terceiro fator (F3) parece estar mais relacionado com a fluência, pois ambos os itens da fluência apresentam carga positiva alta, enquanto todos os demais itens têm carga negativa ou positiva relativamente baixa, exceto conteúdo textual, que apresenta carga positiva em F3.

A análise fatorial exploratória revelou a presença de carga fatorial forte de três fatores subjacentes no modelo, indicando um ajuste razoavelmente bom para esses fatores. Os fatores de originalidade e flexibilidade apresentaram altas cargas fatoriais para um mesmo fator, o que pode ser explicado pelo fato de os itens serem similares (funcionalidades e componentes). Cabe ressaltar, no entanto, que os resultados das cargas fatoriais separaram os itens dimensão da fluência em um fator.

5.3 Ameaças à validade

As ameaças à validade da conclusão giram em torno de fatores que impedem a determinação precisa das relações entre a análise dos dados e o resultado. A fim de manter a imparcialidade e evitar a busca ativa de um determinado resultado, foram usados dados de aplicativos reais e independentes submetidos ao concurso App Inventor do Mês, fornecido pela equipe da Fundação App Inventor/MIT. Esse concurso foi escolhido pelo caráter educacional e pelo fato de os critérios incluírem a avaliação da criatividade dos aplicativos submetidos. Todos os aplicativos que tinham um *link* funcional foram baixados. Nenhuma intervenção foi feita para manipular os dados para cumprir os princípios de evitar a busca de um resultado específico. Os aplicativos no conjunto de dados vêm de pessoas de quatro continentes com diversas origens e na comunidade do App Inventor [1]. O conjunto de dados também compreende um período de 6 anos, de 2016 a 2022, permitindo representar uma população diversificada. Para mitigar as ameaças relacionadas ao design, foi utilizada a abordagem sistemática *Goal/Question/Metric* [4] para definir o estudo.

A avaliação dos artefatos foi realizada de forma automatizada seguindo o modelo *Creassessment* com base na definição geral de criatividade. Em termos do construto, foi fornecida uma definição do modelo conceitual baseada na definição clássica da literatura para o pensamento divergente [15] adaptando-a para refletir essas dimensões em aspectos de um único artefato (aplicativo móvel). Particularmente, para flexibilidade e fluência, aproximações foram feitas para refletir a avaliação de um único artefato. Para a flexibilidade, o uso de diferentes aspectos de funcionalidades e blocos de programação pode não necessariamente equivaler à flexibilidade no sentido clássico, mas apenas no contexto de um único artefato. Para a

fluência, o tamanho absoluto do código, isto é, número de comandos, não equivale necessariamente à fluência geral do código, a qual pode estar mais relacionada à sofisticação do uso de construções de código, ou seja, de maneira inteligente. No entanto, estas aproximações fornecem um primeiro passo no fornecimento de um modelo conceitual para avaliar a criatividade de aplicativos móveis.

6 DISCUSSÃO

Os resultados da análise estatística do modelo Creassessment mostram evidências suficientes para considerar a confiabilidade e a validade de construto como um modelo aceitável para medir a criatividade de aplicativos, o qual pode ser complementado por outras análises. A confiabilidade do modelo foi avaliada por meio do coeficiente ômega, que representa as cargas fatoriais e fornece uma medida estável de confiabilidade. O valor ômega obtido de 0,86 indicou boa confiabilidade. A eliminação de itens, exceto conteúdo textual, diminuiu o coeficiente ômega. O índice KMO de 0,77 indicou adequação para análise fatorial, que identificou três fatores. O Fator 1 mesclou originalidade e flexibilidade, o Fator 2 representou a flexibilidade de programação e o Fator 3 demonstrou estar relacionado principalmente à fluência. Em geral, as cargas fatoriais mostraram associações significativas entre os itens e seus fatores subjacentes, com exceção da originalidade do conteúdo textual que não apresentou carga fatorial acima de 0,50 em nenhum fator. É importante destacar que, apesar de esses resultados parecerem intuitivos, não foi encontrada nenhuma análise na literatura evidenciando que a originalidade e flexibilidade de componentes e funcionalidade de artefatos de software podem ter o mesmo fator subjacente. Dessa forma, esses resultados podem indicar, por exemplo, que quanto mais um aplicativo apresenta funcionalidades diferentes, maior é a originalidade de sua interface de usuário. Esses resultados contribuem para o avanço do estudo da criatividade especificamente de artefatos de software com enfoque sobre aplicativos móveis e traz implicações pedagógicas para o fomento da criatividade. Nesse sentido, uma das formas de fomentar a criatividade dos estudantes em termos de desenvolvimento de funcionalidades mais originais pode ser indicando-os a explorarem os componentes disponíveis no ambiente de programação.

Os resultados fornecem uma primeira indicação de que o modelo analítico Creassessment pode ser usado para medir a criatividade de aplicativos móveis criados com o App Inventor considerando originalidade, fluência e flexibilidade. É importante destacar que a originalidade é avaliada em relação a um universo de referência. Usar um universo de referência como base para medir a criatividade em aplicativos pode ser uma abordagem bem-sucedida. Ao estabelecer um universo de referência, o qual representa uma coleção de aplicativos existentes, a originalidade pode ser avaliada com base no desvio dessa norma estabelecida. Além disso, o uso de um universo de referência fornece uma estrutura padronizada e objetiva para medir a criatividade, permitindo comparações consistentes e facilitando a interpretação da medição da criatividade. Pesquisadores podem utilizar o modelo para conduzir estudos e investigações rigorosas

para melhorar a compreensão geral da criatividade no ensino de computação.

6.1 Limitações

Os modelos automatizados podem ter limitações na captura dos aspectos da criatividade e podem ter dificuldades para identificar certos elementos criativos que requerem julgamento humano. Embora o modelo Creassessment tenha sido introduzido como uma ferramenta valiosa para investigar a criatividade, ele não é inerentemente superior em termos de confiabilidade ou validade quando comparado às classificações humanas subjetivas. No entanto, resultados indicam que a abordagem automatizada apresenta uma alternativa viável às avaliações humanas, pois apresenta confiabilidade e validade nesse contexto específico. Outras abordagens automatizadas propostas, como SemDis [6] e AuDrA [30] para outros tipos de artefatos (desenhos e textos) também indicaram que medidas automatizadas, como distância semântica e modelos de aprendizado profundo, são adequadas para mensurar a criatividade.

Cabe ressaltar que o ônus da classificação humana na avaliação da criatividade pode ser substancial, principalmente devido à natureza subjetiva da tarefa e ao potencial de vieses e inconsistências. Os avaliadores humanos geralmente enfrentam desafios para fornecer avaliações consistentes e confiáveis, que podem ser influenciadas por preferências pessoais, experiência, fadiga e restrições de tempo. Além disso, o processo de revisar e classificar manualmente um grande volume de trabalhos criativos pode consumir muito tempo e muitos recursos. Assim, embora reconhecendo as limitações, o uso de modelos automatizados na avaliação da criatividade pode aumentar a eficiência, objetividade e escalabilidade no processo de avaliação, bem como identificar padrões e nuances em trabalhos criativos que podem ser negligenciados ou subestimados por avaliadores humanos.

7 CONSIDERAÇÕES FINAIS

Este trabalho apresentou a avaliação de um modelo automatizado para avaliar a criatividade de aplicativos móveis. Em geral, os resultados da avaliação mostram que o modelo analítico Creassessment representa um instrumento com confiabilidade e validade aceitáveis, que pode ser usado para avaliar a criatividade de aplicativos como parte do ensino de computação. Para isso, foi adotada uma metodologia sistemática, utilizando como dados artefatos de software retirados do concurso App Inventor do Mês com apoio da Fundação App Inventor/MIT para representar uma gama diversificada de domínios e níveis de complexidade. Apoiado por meio da ferramenta online CodeMaster [38], o modelo Creassessment [1] fornece suporte automatizado que ajuda a garantir a consistência e a precisão dos resultados da avaliação, bem como a eliminar o viés. Isto preenche uma lacuna crucial no estado da arte, fornecendo uma abordagem abrangente para avaliar a criatividade de aplicativos, promovendo assim o desenvolvimento de habilidades inovadoras e criativas dos alunos como parte do ensino de computação por meio do

desenvolvimento de aplicativos. Além disso, o modelo pode auxiliar em reduzir a carga de trabalho dos professores e deixá-los livres para dedicar mais tempo para realizar avaliações complementares sobre fatores que não são facilmente automatizados, e fornece insumos para atribuição de nota e feedback.

Considerando os resultados positivos em termos de confiabilidade e validade do modelo, é possível realizar estudos longitudinais sobre o desenvolvimento da criatividade para investigar as trajetórias de desenvolvimento da criatividade no desenvolvimento de artefatos de software. Ao acompanhar o desenvolvimento da criatividade dos indivíduos ao longo do tempo, é possível examinar como diferentes fatores, tais como nível educacional, experiências e intervenções, influenciam o desenvolvimento da criatividade. Considerando também as medidas de criatividade existentes, uma pesquisa futura envolve a comparação dos resultados obtidos por meio do modelo *Creassessment* com outras medidas de criatividade bem estabelecidas, como os Testes de Pensamento Criativo de Torrance [36]. Além disso, os traços de personalidade podem ser estudados por meio do modelo *Big Five*, bem como o potencial de criatividade em domínios por meio da Escala de Domínios de Criatividade de Kaufman [19]. Ao realizar estudos comparativos com diferentes ferramentas, é possível obter informação sobre a compatibilidade do modelo com as ferramentas existentes de avaliação da criatividade, permitindo uma compreensão mais profunda dos pontos fortes e limitações de cada abordagem de medição, destacando as suas contribuições únicas para avaliar a criatividade. Além disso, investigar as relações entre os resultados de diferentes medições pode fornecer uma compreensão mais abrangente e diferenciada das capacidades criativas dos indivíduos e fornecer informações valiosas para educadores, pesquisadores e profissionais na concepção de estratégias eficazes para nutrir e desenvolver a criatividade dentro da educação em computação.

AGRADECIMENTOS

Agradecemos aos pesquisadores associados a Fundação App Inventor/MIT, os quais forneceram apoio para a obtenção de acesso aos aplicativos do concurso Aplicativo do Mês. Os autores expressam, ainda, sua gratidão aos revisores anônimos por seus comentários valiosos e sugestões úteis. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

REFERÊNCIAS

- [1] Nathalia da Cruz Alves. Assessing the Creativity of Mobile Applications in Computing Education. 2023. Tese (Doutorado em Ciência da Computação) – PPGCC/Universidade Federal de Santa Catarina. <https://repositorio.ufsc.br/handle/123456789/251989>.
- [2] Nathalia da Cruz Alves, Christiane. Gresse von Wangenheim, Lúcia Helena Martins-Pacheco (2021). Assessing Product Creativity in Computing Education: A Systematic Mapping Study. *Informatics in Education*, 20(1), 19-45. DOI: <https://doi.org/10.15388/infedu.2021.02>.
- [3] Teresa M. Amabile. 1996. *Creativity in Context*. Westview Press, Boulder, CO.
- [4] Victor R. Basili, Gianluigi Caldiera, H. Dieter Rombach. 1994. The Goal Question Metric Approach. In *Encyclopedia of Software Engineering*. Wiley.
- [5] Satabdi Basu. 2019. Using Rubrics Integrating Design and Coding to Assess Middle School Students' Open-ended Block-based Programming Projects. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 1211–1217. DOI: <https://doi.org/10.1145/3287324.3287412>.
- [6] Roger E. Beaty, Dan R. Johnson. 2021. Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior Research* 53, 2, 757–780. DOI: <https://doi.org/10.3758/s13428-020-01453-w>.
- [7] R. A. Beghetto. 2010. Creativity in the classroom. In J. C. Kaufman & R. J. Sternberg (Eds.), *Cambridge Handbook of Creativity* (pp. 447-463). New York: Cambridge University Press.
- [8] Vicki Bennett, Kyu Han Koh, Alexander Repenning. 2013. Computing creativity: Divergence in computational thinking. In *Proceedings of the 44th ACM Technical Symposium on Computer Science Education*. ACM, 359-364. DOI: <https://doi.org/10.1145/2445196.2445302>.
- [9] Timothy A. Brown. 2006. *Confirmatory factor analysis for applied research*. New York.
- [10] David Cavallo, Helena Singer, Alex S. Gomes, Ig I. Bittencourt, Ismar F. Silveira. 2016. Inovação e Criatividade na Educação Básica: Dos conceitos ao ecossistema. *Revista Brasileira de Informática na Educação*, 24(2).
- [11] Sally A. Fincher, Anthony V. Robins (Eds.). 2019. *The Cambridge Handbook of Computing Education Research*. Cambridge: University Press.
- [12] David B. Flora. 2020. Your Coefficient Alpha Is Probably Wrong, but Which Coefficient Omega Is Right? A Tutorial on Using R to Obtain Better Reliability Estimates. *Advances in Methods and Practices in Psychological Science*, 3(4), 484-501. <https://doi.org/10.1177/2515245920951747>.
- [13] Lilach Gal, Arnon Hershkovitz, Andoni Eguiluz, Mariluz Guenaga, Pablo Garaizar. 2017. Suggesting a Log-Based Creativity Measurement for Online Programming Learning Environment. In *Proceedings of the 4th Conference on Learning at Scale*, 273-277. DOI: <https://doi.org/10.1145/3051457.3054003>.
- [14] Louis W. Gorfelfeld. 1995. An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, 55(3), 377-393.
- [15] Joy Paul Guilford. 1950. Creativity. *American Psychologist*, 5, 444-454. DOI: <https://doi.org/10.1037/h0063487>.
- [16] Joseph F. Hair, et al. 2009. *Multivariate data analysis* (7th ed.). Prentice Hall.
- [17] Andrew F. Hayes, Jacob J. Coutts. 2020. Use Omega Rather than Cronbach's Alpha for Estimating Reliability. *But...*, *Communication Methods and Measures*, 14, 1, 1-24. DOI: <https://doi.org/10.1080/19312458.2020.1718629>.
- [18] Arnon Hershkovitz, Raquel Sitman, Rotem Israel-Fishelson, Andoni Eguiluz, Pablo Garaizar, Mariluz Guenaga. 2019. Creativity in the acquisition of computational thinking. *Interactive Learning Environments* 27, 6 (Sep. 2019), 813-829. DOI: <https://doi.org/10.1080/10494820.2019.1610451>.
- [19] James C. Kaufman. 2012. Counting the muses: Developmental of the Kaufman Domains of Creativity Scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts*, v. 6, n. 4, p. 298–308. DOI: <https://doi.org/10.1037/a0029751>.
- [20] James C. Kaufman, Ronald A. Beghetto. 2009. Beyond Big and Little: The Four C Model of Creativity. *Review of General Psychology* 13, 1 (Mar. 2009), 1–12. DOI: <https://doi.org/10.1037/a0013688>.
- [21] James C. Kaufman, Jonathan A. Plucker, John Baer. 2008. *Essentials of Creativity Assessment*. John Wiley & Sons.
- [22] Prapti Khawas, Peeratham Techapolokul, Eli Tilevich. 2019. Unmixing Remixes: The How and Why of Not Starting Projects from Scratch. In *Proceedings of the IEEE Symposium on Visual Languages and Human-Centric Computing*, 169-173. Memphis, TN, USA. DOI: <https://doi.org/10.1109/VLHCC.2019.8818834>.
- [23] Kyu Han Koh. 2011. Computing indicators of creativity. In *2011 IEEE Symposium on Visual Languages and Human-Centric Computing*. DOI: <https://doi.org/10.1109/vlhc.2011.6070407>.
- [24] Anastasia Kovalkov, Benjamin Paaßen, Avi Segal, Niels Pinkwart, Kobi Gal. 2021. Automatic Creativity Measurement in Scratch Programs Across Modalities. *IEEE Transactions on Learning Technologies* 14, 6 (Dec. 2021), 740-753. DOI: <https://doi.org/10.1109/TLT.2022.3144442>.
- [25] Anastasia Kovalkov, Avi Segal, Kobi Gal. 2020. Inferring Creativity in Visual Programming Environments. In *Proceedings of the Seventh ACM Conference on Learning @ Scale*. Association for Computing Machinery, New York, NY, USA, 269–272. DOI: <https://doi.org/10.1145/3386527.3406725>.
- [26] Jiwen Luo, Feng Lu, Tao Wang. 2020. A Multi-Dimensional Assessment Model and Its Application in E-learning Courses of Computer Science. In *Proceedings of the 21st Annual Conference on Information Technology Education*, 187–193. ACM, New York, NY, USA. DOI: <https://doi.org/10.1145/3368308.3415388>.
- [27] Sven Manske, Ulrich Hoppe. 2014. Automated Indicators to Assess the Creativity of Solutions to Programming Exercises. In *Proceedings of the IEEE 14th International Conference on Advanced Learning Technologies*, 497-501. DOI: <https://doi.org/10.1109/ICALT.2014.147>.
- [28] MIT App Inventor. 2023. App of the Month. <https://appinventor.mit.edu/explore/app-month-gallery>.

- [29] Eni Mustafaraj, Franklyn A. Turbak, Maja Svanberg. 2017. Identifying Original Projects in App Inventor. In Proceedings of the 30th International Florida Artificial Intelligence Research Society Conference, 567-572. Florida: Association for the Advancement of Artificial Intelligence. URL: <https://aaai.org/papers/567-flairs-2017-15549/>.
- [30] John D. Patterson, Baptiste Barbot, James Lloyd-Cox, Roger Beaty. 2022. AuDrA: An automated drawing assessment platform for evaluating creativity. PsyArXiv. DOI: <https://doi.org/10.31234/osf.io/t63dm>.
- [31] Jonathan Plucker, Ronald A. Beghetto, Gayle Dow. 2004. Why isn't creativity more important to educational psychologists? Potential, pitfalls, and future directions in creativity research. *Educational Psychologist*, 39, 83–96. DOI: https://doi.org/10.1207/s15326985Sep3902_1.
- [32] Mel Rhodes. 1961. An Analysis of Creativity. *The Phi Delta Kappan*, 42(7), 305-310.
- [33] Graeme D. Ritchie. 2001. Assessing Creativity. In Proceedings of the AISB symposium on AI and creativity in arts and science (pp. 3-11). York: The Society for the Study of Artificial Intelligence and Simulation of Behaviour.
- [34] Valerie J. Shute, Chen Sun, Jodi Asbell-Clarke. 2017. Demystifying computational thinking. *Educational Research Review*, 22, 142-158. DOI: <https://doi.org/10.1016/j.edurev.2017.09.003>.
- [35] Taguma, M. F. (2018). Future of Education and Skills 2030: Conceptual Learning Framework. OECD. Retrieved August 4, 2021 from <https://www.oecd.org/education/2030-project/>.
- [36] Ellis Paul Torrance. 2008. The Torrance Tests of Creative Thinking Norms—Technical Manual Figural (Streamlined) Forms A & B. Bensenville: IL: Scholastic Testing Service.
- [37] Franklyn A. Turbak, Eni Mustafaraj, Maja Svanberg, Michael Dawson. 2017. Work in Progress: Identifying and Analyzing Original Projects in an Open-Ended Blocks Programming Environment. In Proceedings of the 23rd International Conference on Distributed Multimedia Systems, Visual Languages and Sentient Systems, 115-117. Florida: Association for the Advancement of Artificial Intelligence.
- [38] Universidade Federal de Santa Catarina - Computação na Escola. CodeMaster. <http://apps.computacaonaescola.ufsc.br/codemaster/>
- [39] Joke Voogt, Natalie Pareja Roblin. 2012. A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies* 44, 3 (May 2012), 299–321. DOI: <https://doi.org/10.1080/00220272.2012.668938>.