

# Avaliação Empírica da Utilização de um Jogo para Auxiliar a Aprendizagem de Programação

Elieser A. de Jesus<sup>1</sup>, André L. A. Raabe<sup>1</sup>

<sup>1</sup>UNIVALI – Universidade do Vale do Itajaí

{elieser, raabe}@univali.br

**Abstract.** *This paper describes a study where we evaluated quantitatively the effect of using a computer game as a support tool for programming learning. In this game the students have built programs to win. Experiments were carried out with three classes in a experimental design including pretest-posttest and control group. No significant difference was observed in the progress of students who played in comparison with students who have not played (the control group). Other results and the validity of the findings are discussed in this paper.*

**Resumo.** *Este artigo descreve uma avaliação quantitativa sobre o efeito da utilização de um jogo de computador como ferramenta de apoio à aprendizagem de programação. Neste jogo os alunos precisavam programar para vencer. Foram realizados experimentos com 3 turmas em um delineamento experimental com pré e pós-teste e grupo de controle. Não foi observada nenhuma diferença significativa no progresso dos alunos que jogaram em comparação com os que não jogaram (o grupo de controle). Outros resultados bem como a validade do experimento são discutidos em maiores detalhes no artigo.*

## 1. Introdução

A aprendizagem de programação é entendida como sendo um dos pilares da formação em computação e áreas afins (CRISTÓVÃO, 2008). Sheard *et al* (2009) mencionam que a grande quantidade de literatura relacionada à programação introdutória é reflexo dos muitos desafios relacionados ao tema.

Alguns pesquisadores como Feldgen e Clúa (2004), Cliburn (2006) e Bayllis (2009) vêm chamando a atenção sobre o crescente interesse dos professores pela utilização de jogos na aprendizagem de programação, já que os jogos fazem parte da realidade cultural de muitos dos alunos. Becker e Parker (2005) dizem que na universidade onde atuam de 65% a 75% dos alunos ingressantes procuram o curso de Ciência da Computação porque são usuários de jogos. Estes autores ainda argumentam que os alunos geralmente entendem muito melhor os problemas relacionados aos jogos do que os tradicionais programas de controle de estoque e folha de pagamento.

Parece haver um senso comum estabelecido de que os jogos provocam melhoria na aprendizagem. Possivelmente, muitos assumem que o fato dos jogos fazerem parte do cotidiano dos alunos é suficiente para torná-los benéficos enquanto ferramentas didáticas. Entretanto, algumas evidências (discutidas a seguir) têm mostrado que talvez haja algum entusiasmo exagerado sobre o assunto, pois as melhorias na aprendizagem

que alguns supõem como certa nem sempre são observadas quando experimentos controlados são realizados.

Sendo assim, este artigo relata os resultados de uma pesquisa onde buscou-se responder a seguinte questão: *De que forma a utilização de um jogo de computador focado na resolução de problemas algorítmicos influencia a aprendizagem de programação introdutória?*

Outra questão também abordada neste artigo está relacionada à forma contextualizada como os desafios dos jogos são apresentados. É comum que exercícios de programação resumam-se a “calcule isto” e “resolva aquilo”, onde o resultado do cálculo ou da resolução não é utilizado para algo que faça sentido dentro de um contexto maior. Então, também se procurou responder a uma segunda questão: *um jogo promove melhorias na aprendizagem ou este mesmo resultado pode ser alcançado apenas com a utilização de problemas de programação melhor contextualizados?*

Buscou-se responder as questões mencionadas anteriormente através da verificação das seguintes hipóteses: *H1- Exercícios de programação com enunciados mais contextualizados melhoram a aprendizagem;* e *H2: O uso do jogo promove melhor desempenho do que apenas a utilização de exercícios mais contextualizados.* Estas hipóteses ainda foram pormenorizadas através da categorização dos alunos em três classes: *ótimos, medianos ou com dificuldades*, pois imaginou-se, por exemplo, que alunos *com dificuldades* poderiam se beneficiar mais do jogo e da contextualização de enunciados do que as demais classes de alunos.

Na seqüência deste artigo são apresentadas evidências sobre o efeito dos jogos na aprendizagem de programação. Em seguida são apresentados os detalhes do experimento relatado neste artigo e os resultados obtidos. Por fim, são apresentadas as conclusões e uma discussão sobre a validade do experimento realizado.

## **2. Evidências Sobre o Efeito dos Jogos na Aprendizagem**

Algumas pesquisas vêm sendo realizadas com o objetivo de mensurar a influência dos jogos de computador na aprendizagem de programação introdutória. Cliburn (2006) realizou um estudo onde concluiu que com a presença dos jogos os alunos tornam-se mais motivados e engajados, mas não foram observadas melhorias nas notas. Bayliss (2007) também não observou nenhuma diferença significativa nas notas após a utilização de jogos. Entretanto, a autora sugere que os benefícios podem aparecer mais tarde.

Bierre e Phelps (2004) relatam, assim como Long (2007), o aumento da motivação dos alunos ao utilizarem o Robocode<sup>1</sup>, um ambiente onde tanques de guerra programados pelos alunos batalham entre si. Entretanto, apesar do aumento da motivação nenhum destes autores relata melhoria nas notas. Barnes *et al* (2007) também concluíram que não houve melhora significativa na aprendizagem após a utilização de um jogo sobre os principais tópicos de programação introdutória. As notas dos alunos continuaram, nas palavras dos próprios autores, surpreendentemente baixas.

---

<sup>1</sup> <http://robocode.sourceforge.net/>

Por outro lado, Eagle e Barnes (2009) observaram melhorias nas notas de alunos que utilizaram um jogo para praticar especificamente o conceito de *loops*. Bayllis (2009) também cita resultados significativamente positivos no ensino de programação introdutória baseado na utilização de jogos. Feldgen e Clúa (2004) relatam uma diminuição na taxa de abandono do curso depois da inserção dos jogos no currículo, ainda que os próprios autores reconheçam que tal efeito seja uma consequência da implementação de diversas mudanças, entre elas a inserção dos jogos.

As pesquisas discutidas anteriormente mostram que tanto existem indícios de que os jogos promovem melhorias na aprendizagem de programação quanto existem evidências de que isto não acontece. Esta indefinição sobre o efeito dos jogos na aprendizagem de programação foi o que motivou a realização do experimento relatado neste artigo, cujos detalhes são apresentados a seguir.

### 3. O Experimento Realizado

Para responder as questões mencionadas na introdução deste artigo foi realizado um experimento envolvendo três turmas de alunos ingressantes em um curso de Ciência da Computação. Uma das turmas foi tratada como *grupo de controle* e as outras duas como os *grupos experimentais*. Todas as turmas participaram de um pré-teste, utilizaram um software e responderam um pós-teste, nesta ordem. O experimento aconteceu ao longo de 18 semanas de aula (um semestre letivo na instituição). Os pré-testes foram aplicados por volta da 4ª semana, as intervenções com os softwares aconteceram entre a 10ª e a 17ª semana em paralelo com as aulas, e o pós-teste foi realizado ao final do semestre, por volta da 17ª semana.

#### 3.1 As intervenções com os softwares

Foram construídos três softwares e cada uma das turmas de alunos utilizou um deles. Em todos os softwares os alunos resolveram um mesmo conjunto de nove problemas algorítmicos envolvendo desvios condicionais, *arrays* e *loops*. Apenas estes três tópicos foram tratados por serem mencionados por Roundtree e Roundtree (2003) e Dale (2006) como alguns dos mais problemáticos na perspectiva dos alunos iniciantes.

Para resolver os *desafios* algorítmicos os alunos utilizaram uma linguagem de programação não-textual apelidada de *Escracho*. As partes de um programa nesta linguagem são representadas por elementos gráficos que podem ser arrastados e encaixados entre si, desde que sejam respeitadas algumas restrições sintáticas. Abordagens similares são utilizadas nos softwares Scratch<sup>2</sup> e Alice<sup>3</sup>. Além disso, em todos os três softwares foi utilizado um mecanismo de correção automática de pequenos programas, utilizado pelos alunos para obter *feedback* sobre as suas soluções.

O que diferenciou os 3 softwares mencionados anteriormente entre si foi o nível de contextualização presente nos enunciados dos problemas algorítmicos. O nível mais alto de contextualização foi utilizado dentro do jogo, um dos três softwares utilizados no experimento. Neste software todos os enunciados descreviam objetos e situações do

---

<sup>2</sup> <http://scratch.mit.edu/>

<sup>3</sup> <http://www.alice.org/>

mundo do jogo. No outro extremo, o software com o nível mais baixo de contextualização, os alunos utilizaram enunciados abstratos. Todas as referências ao jogo e aos seus objetos foram retiradas dos enunciados. Por fim, no software com o nível intermediário de contextualização, os enunciados continham apenas imagens estáticas e textos descrevendo as situações do jogo.

O grupo de controle utilizou o software com o nível mais baixo de contextualização, mencionado anteriormente. O primeiro grupo experimental utilizou o jogo (nível mais alto de contextualização) e o segundo utilizou o software com o nível intermediário de contextualização. Os alunos utilizaram os softwares durante aproximadamente 2 horas nos laboratórios da universidade onde foi realizado o experimento, e depois disso puderam utilizá-los por meio da internet até o dia da realização do pós-teste.

### **3.2 Pré e Pós-Testes e a Mensuração do desempenho das turmas**

Para a realização do pré e do pós-teste foi elaborado um instrumento de avaliação com questões de múltipla escolha envolvendo os tópicos de interesse da pesquisa, a saber: os *arrays*, desvios condicionais e *loops*. Este mesmo instrumento de avaliação foi utilizado tanto para o pré quanto para o pós-teste e foi discutido em Jesus e Raabe (2009).

O pré-teste foi aplicado antes que os professores das turmas participantes do experimento iniciassem a apresentação dos tópicos de interesse da pesquisa. A utilização de cada um dos softwares pelos alunos aconteceu logo depois que estes tópicos foram trabalhados em sala de aula. Depois das intervenções com os softwares, próximo ao final do semestre letivo, finalmente foi aplicado o pós-teste.

De maneira geral, o que se buscou foi mensurar o progresso dos alunos sob diferentes aspectos ao longo do semestre letivo. O progresso das turmas foi mensurado pelo que se chamou de *fator de aprendizagem*, a diferença média entre os escores de pós e pré-testes dos alunos. Além disso, as questões do instrumento de avaliação utilizado no pré e pós-teste foram classificadas segundo a taxonomia de Bloom. Isto possibilitou a análise do progresso dos alunos nos níveis da taxonomia, já que estes níveis diferem entre si quanto a complexidade (KRATHWOL, 2002).

Os progressos das turmas de alunos foram comparados entre si para determinar qual delas obteve melhor desempenho. Em algumas das comparações utilizou-se testes de hipótese, todos realizados com 95% de confiança ( $\alpha = 0.05$ ). Antes da realização destes testes foi verificada a suposição de normalidade dos dados amostrais com o teste de Shapiro-Wilk. Quando a normalidade foi observada utilizou-se os testes paramétricos ANOVA e teste-t pareado. Onde não se pôde considerar os dados como normalmente distribuídos utilizou-se o teste não-paramétrico de Kruskal-Wallis.

## **4. Resultados**

Alguns testes estatísticos foram aplicados na pesquisa relatada neste artigo e utilizou-se o *valor p* para determinar a aceitação ou rejeição de hipóteses. O *valor p* é a probabilidade de se rejeitar uma hipótese nula quando ela é verdadeira. A hipótese nula deve ser rejeitada quando o *valor p* for menor ou igual ao nível de significância adotado para o teste (geralmente adota-se o valor 0,05).

#### 4.1 Progresso dos Alunos do Pré Para o Pós-Teste

Todos alunos responderam o mesmo pré e pós-teste, sendo que ambos os testes tinham 10 questões de múltipla escolha classificadas segundo a taxonomia de Bloom. Para cada aluno foi calculado o número de acertos no pré e o número de acertos no pós-teste.

Primeiramente verificou-se se todas as turmas possuíam o mesmo nível de conhecimento no momento do pré-teste. O teste ANOVA não detectou nenhuma diferença significativa nas médias dos pré-testes das 3 turmas ( $p = 0.94$ ). Um teste-t pareado mostrou que os escores médios nos pós-testes foram significativamente maiores que os escores dos pré-testes (valores  $p$  na Tabela 1).

Na Tabela 1 pode-se observar que a turma que utilizou o jogo apresentou um escore médio e uma mediana de pós-teste ligeiramente maior que as demais turmas. O teste de Kruskal-Wallis indicou que a diferença não foi significativa ( $p = 0.4$ ).

**Tabela 1. Escores no pré e pós-teste e fatores de aprendizagem (F.A.)**

	Grupo de Controle			Grupo Exp. 1 (utilizou o jogo)			Grupo Exp. 2		
	Pré	Pós	F.A.	Pré	Pós	F.A.	Pré	Pós	F.A.
Média	4,45	6,55	2,09	4,78	7,78	3,00	4,70	6,50	1,80
Desvio padrão	2,34	1,63	0,94	2,11	0,97	1,50	2,56	2,65	0,70
Mediana	5	7	2	5	8	3	4,5	7,5	2
Valor $p$ do teste-t		0,002			0,0027			0,0028	
N		9 alunos			11 alunos			20 alunos	

O objetivo da pesquisa relatada neste artigo foi comparar o progresso das turmas de alunos entre si. Uma das métricas de progresso utilizadas foi a diferença entre os escores de pós e pré-teste, o que se chamou de *fator de aprendizagem*. Sendo assim, comparou-se os fatores de aprendizagem das turmas entre si com o teste de Kruskal-Wallis. Este teste retornou um valor 0.56, não permitindo a rejeição da hipótese nula que assume a igualdade entre as amostras. Sendo assim, pode-se concluir que todas as intervenções realizadas foram equivalentes entre si quanto ao efeito observado na aprendizagem.

#### 4.2 Progresso das Classes de Alunos

Os alunos participantes do experimento também foram classificados segundo dois escores: o desempenho no pós-teste e a média final (MF) na disciplina de programação introdutória. Então, os alunos foram divididos nas seguintes classes: a) *Alunos ótimos* (escore  $\geq 8$  e  $\leq 10$ ); b) *Alunos medianos* (escore  $\geq 6$  e  $< 8$ ); e c) *Alunos com dificuldades* (escore  $< 6$ ). O escore do pós-teste foi utilizado na classificação pelo fato de que no momento do pré-teste os alunos sequer haviam aprendido os tópicos de programação que interessavam para a pesquisa. Já a classificação pela média final (MF) foi utilizada por se entender que qualquer melhoria na aprendizagem produzida pela utilização do jogo deveria se traduzir em médias finais mais altas na disciplina de programação introdutória, ao invés de apenas escores de pós-teste mais altos.

Na Tabela 2 são apresentados os *fatores de aprendizagem* (F.A.) para as classes de alunos mencionadas anteriormente, bem como as médias no pré e no pós-teste.

**Tabela 2. Fatores de aprendizagem (alunos classificados pelo pós-teste e MF)**

Turma	Classe de alunos	Classificação pelo pós-teste				Classificação pela MF			
		N	$\bar{X}_{Pré}$	$\bar{X}_{Pós}$	F.A.	N	$\bar{X}_{Pré}$	$\bar{X}_{Pós}$	F.A.
Grupo de Controle	Com dificuldades	3	2,67	4,33	1,67	7	5,29	6,86	1,57
	Medianos	5	4,20	6,80	2,60	3	3,67	6,67	3,00
	Ótimos	3	6,67	8,33	1,67	1	1,00	4,00	3,00
Grupo Exp. 1 (usou o jogo)	Com dificuldades	0	-	-	-	3	6,00	7,33	1,33
	Medianos	4	5,25	7,00	1,75	5	4,60	8,00	3,40
	Ótimos	5	4,40	8,40	4,00	1	2,00	8,00	6,00
Grupo Exp. 2	Com dificuldades	7	2,14	3,43	1,29	9	6,33	8,00	1,67
	Medianos	3	7,67	6,33	-1,33	7	3,86	6,29	2,43
	Ótimos	10	5,60	8,70	3,10	4	2,50	3,50	1,00

Os fatores de aprendizagem (F.A.) de cada uma das classes foram comparados entre si com o teste de Kruskal-Wallis. No caso da classificação pelo pós-teste encontrou-se os valores 0.28 para os alunos *ótimos*, um valor de 0.16 para os *medianos* e 0.72 para aqueles *com dificuldades*. Na classificação pela média final (MF) os valores encontrados para as classes foram respectivamente 0.87, 0.83 e 0.28. Como nenhum destes valores é menor do que o nível de significância adotado para os testes (0.05) não se pode rejeitar a hipótese nula sobre a igualdade das amostras. Sendo assim, conclui-se que, independentemente da classificação utilizada (score da MF ou pós-teste) e para qualquer uma das classes, nenhuma das turmas apresentou um progresso diferente das outras turmas. Os progressos foram estatisticamente equivalentes entre si.

Apesar de não se ter encontrado nenhuma diferença significativa entre os progressos dos alunos, a turma que utilizou o jogo (Grupo exp. 1) terminou o experimento sem nenhum aluno classificado como *Com dificuldades*. Os alunos *Ótimos* nos dois grupos experimentais apresentaram um progresso maior do que os *Ótimos* no grupo de controle. Os *Medianos* do *Grupo experimental 2* regrediram, apresentando um fator de aprendizagem negativo. No caso dos *Medianos*, o grupo de controle apresentou um progresso maior do que os dois grupos experimentais.

Além da análise do fator de aprendizagem de cada classe de alunos também foi analisada a transição dos alunos de uma classe para outra. Na Tabela 3 são apresentados os percentuais de alunos em cada uma das classes mencionadas anteriormente. Pode se observar que a turma que utilizou o jogo (Grupo exp. 1) progrediu um pouco mais do que as outras. Nos dois grupos experimentais alguns alunos inicialmente classificados como *com dificuldades* (33,3% e 20%) tornaram-se alunos *Ótimos*. Isto não aconteceu com o grupo de controle, o que sugere que os níveis mais altos de contextualização tenham provocado esta diferença. Além disso, o grupo que jogou também tem o maior percentual de alunos *medianos* que se tornaram *alunos ótimos* (44,4%). Todas as 3 turmas iniciaram com aproximadamente o mesmo percentual de alunos *com dificuldades*, e somente a turma que utilizou o jogo (grupo experimental 1) acabou os experimentos sem nenhum indivíduo classificado desta forma.

**Tabela 3. Progresso das classes de alunos entre o pré e o pós-teste**

Alunos que...	Grupo de Controle (N=11)	Grupo Exp. 1 (utilizou o jogo) (N=9)	Grupo Exp. 2 (N=20)
Progrediram de <i>com dificuldades</i> para <i>ótimo</i>	0,0%	33,3%	20,0%
Progrediram de <i>mediano</i> para <i>ótimo</i>	27,3%	44,4%	25,0%
Progrediram de <i>com dificuldades</i> para <i>mediano</i>	27,3%	22,2%	0,0%
Progrediram	54,5%	77,8%	45,0%
Regrediram	0,0%	0,0%	10,0%
Iniciaram <i>com dificuldades</i>	54,5%	55,5%	60,0%
Continuaram <i>com dificuldades</i>	27,3%	0,0%	35,0%
Continuaram <i>medianos</i>	18,2%	22,2%	5,0%
Continuaram <i>ótimos</i>	0,0%	0,0%	5,0%

### 4.3 Progresso nos níveis da taxonomia de Bloom

As questões do instrumento de avaliação utilizado no pré e no pós-teste foram classificadas segundo a taxonomia revisada de Bloom (KRATHWOHL, 2002). O objetivo desta classificação foi mensurar o progresso dos alunos em alguns dos níveis da taxonomia, já que estes níveis são progressivamente mais complexos.

A Tabela 4 apresenta o percentual de acertos nos níveis da taxonomia de Bloom. Estes percentuais foram obtidos como uma razão entre o *número de acertos dos alunos* e o *total de acertos possíveis*, sendo que este total foi obtido pelo *número de questões em um determinado nível da taxonomia* multiplicado pelo *número de alunos*.

**Tabela 4. Desempenho dos alunos nos níveis da taxonomia de Bloom**

Nível da Taxonomia	Turma	N	Nº de acertos		Percentual de acertos		
			Pré teste	Pós teste	Pré teste	Pós teste	Pós – Pré
Entender (3 questões)	Grupo de controle	11	19/33	29/33	57,5%	87,8%	<b>30,3%</b>
	Grupo exp. 1 (jogo)	9	19/27	24/27	70,3%	88,8%	18,5%
	Grupo exp. 2	20	35/60	45/60	58,3%	75,0%	16,7%
Aplicar (5 questões)	Grupo de controle	11	22/55	32/55	40,0%	58,1%	18,1%
	Grupo exp. 1 (jogo)	9	18/45	37/45	40,0%	82,2%	<b>42,2%</b>
	Grupo exp. 2	20	44/100	60/100	44,0%	60,0%	16,0%
Analisar (2 questões)	Grupo de controle	11	8/22	11/22	36,3%	50,0%	13,7%
	Grupo exp. 1 (jogo)	9	6/18	9/18	33,3%	50,0%	16,7%
	Grupo exp. 2	20	15/40	25/40	37,5%	62,5%	<b>25,0%</b>

Analisando os dados da Tabela 4 pode-se observar que os percentuais de acertos em todos os níveis da taxonomia aumentaram do pré para o pós-teste. No momento dos pré-testes as 3 turmas tinham um desempenho muito próximo nos níveis *Aplicar* (aproximadamente 40% de acertos para todas as turmas) e *Analisar* (aproximadamente 35% de acertos para todas as turmas). No pós-teste o grupo que utilizou o jogo acertou 42,2% a mais de questões classificadas no nível *Aplicar* da taxonomia. Além disso, o

grupo experimental 2 (nível intermediário de contextualização) apresentou um desempenho um pouco melhor que as outras turmas no nível *Analisar*. Conforme a Tabela 4 o grupo de controle saiu-se um pouco melhor que os demais grupos nas questões classificadas no nível *Entender* da taxonomia de Bloom.

## 5. Conclusões

A pesquisa relatada neste artigo foi norteadas pelas questões e hipóteses apresentadas na seção de introdução. Com as análises realizadas não foi possível afirmar que a utilização do jogo melhorou a aprendizagem de desvios condicionais, *loops* ou *arrays*. A análise dos dados também mostrou que o aumento do nível de contextualização dos enunciados não promoveu melhores resultados na aprendizagem como se supôs ao início da pesquisa, pois os grupos experimentais apresentaram um desempenho médio equivalente ao grupo de controle.

Os alunos foram classificados como *com dificuldades*, *medianos* ou *ótimos*. Quando se analisou a transição dos alunos entre estas classes o percentual daqueles que progrediram foi um pouco maior no caso da turma que utilizou o jogo. Além disso, todas as 3 turmas iniciaram com aproximadamente o mesmo percentual de alunos *com dificuldades*, e somente a turma que utilizou o jogo acabou o experimento sem nenhum aluno classificado desta forma. Isto parece sugerir que os alunos *com dificuldades* beneficiaram-se com a utilização do jogo. Cabe ressaltar que não houve comprovação estatística destas hipóteses dado o pequeno tamanho das amostras.

A análise do progresso nos níveis da taxonomia de Bloom também sugere (sem comprovação estatística) que os enunciados com níveis de contextualização mais altos promoveram alguma ascensão nos níveis da taxonomia, e que a utilização do jogo, em particular, favoreceu a *aplicação* do conhecimento. Uma discussão sobre a validade do experimento é apresentada a seguir com base no trabalho de Wainer (2007).

### 5.1 Ameaças a Validade Interna

A seguir são discutidas as ameaças que não puderam ser totalmente mitigadas na realização do experimento.

**Ameaças de Instrumentação** – Nesta pesquisa partiu-se do pressuposto que todas as três turmas lidaram com os mesmos problemas algorítmicos, e que *apenas* o nível de contextualização dos enunciados estava variando em cada um dos três softwares que foram utilizados (um software diferente para cada turma de alunos). É possível que as versões de alguns dos problemas algorítmicos tenham se tornado suficientemente diferentes entre si, a ponto de terem se tornado problemas diferentes ao invés de versões de um mesmo problema. Além disso, percebeu-se que alguns dos enunciados não foram totalmente compreendidos pelos alunos, possivelmente pela linguagem utilizada na descrição dos problemas.

Durante os pré e pós-testes verificou-se que o instrumento de avaliação utilizado foi um tanto longo, causando fadiga nos alunos e aumentando as chances de que as últimas questões fossem respondidas de forma aleatória.

**Ameaça de Seleção** - Este tipo de ameaça é preocupante quando grupos com características desejáveis são deliberadamente escolhidos para participarem de uma

situação que se acredita levar ao resultado desejado. Esta pesquisa foi realizada com 3 turmas de alunos. Para minimizar a ameaça por seleção a turma de alunos que era dirigida pelo orientador do trabalho não utilizou o jogo, mas sim o software onde os enunciados possuíam um nível intermediário de contextualização. A turma escolhida para utilizar o jogo foi dirigida por uma professora sem nenhuma ligação com a pesquisa.

**Ameaça de Contaminação** - Nesta pesquisa, um dos grupos experimentais possuía um aluno repetente que já havia participado do experimento em um semestre anterior fazendo parte do grupo de controle. Este aluno pode ter passado informações para seus colegas e provocado alguma contaminação. Além disso, como os experimentos com os dois grupos experimentais aconteceram em paralelo ao longo do mesmo semestre é possível que os alunos tenham interagido entre si e ensinado uns aos outros.

**Ameaça de Comportamento Competitivo** - Os dois grupos experimentais realizaram os experimentos em paralelo, e é possível que alguns alunos tenham entrado em um “clima” de competição. Porém, o grupo de controle realizou o experimento um semestre antes dos grupos experimentais, o que diminuiu muito as chances de competição entre os grupos experimentais e o grupo de controle.

## 5.2 Ameaças a Validade Externa

A validade externa diz respeito às possibilidades de generalização dos resultados. Não se pode generalizar os resultados obtidos neste trabalho para quaisquer outros jogos, pois os resultados aplicam-se somente ao jogo que foi utilizado, nas condições em que foi utilizado. A utilização do mesmo jogo em situações um pouco diferentes já poderia gerar outros resultados. Por exemplo, é possível que se o mesmo jogo for utilizado durante todo o semestre, como uma ferramenta integrada com um número maior de aulas dos professores, os resultados apontem em outras direções.

Também não se pode generalizar a idéia de que níveis de contextualização mais altos não melhoram a aprendizagem. O que se pode afirmar é que: da forma como foi tratada a contextualização nenhum efeito significativo foi observado na aprendizagem de *loops*, *arrays* e desvios condicionais em comparação com o efeito observado no grupo de controle. Este resultado não deve ser generalizado, pois é possível (e é uma crença dos autores deste artigo) que em outras situações, com outras ferramentas e outras metodologias, a contextualização produza resultados positivos na aprendizagem.

## 5.3 Trabalhos futuros

O principal trabalho futuro refere-se a replicação do experimento com um número maior de alunos, pois nestas condições é possível que os efeitos observados tornem-se estatisticamente significativos. Além disso, entende-se que o acréscimo de uma avaliação qualitativa contribuirá muito para o enriquecimento da pesquisa.

## Referências

Barnes, T. et al. Game2Learn: building CS1 learning games for retention. In: annual SIGCSE Conference on Innovation and Technology in Computer Science Education, 12., 2007, Dundee, Scotland. Proceedings... New York:ACM, 2007. p. 121-125.

- Bayliss, J. D. Using games in introductory courses: tips from the trenches. In: ACM technical symposium on Computer science education, 40., 2009, Chattanooga, USA. Proceedings... New York:ACM, 2009. p. 337-341.
- Becker, K.; Parker, J. R. All I ever needed to know about programming, I learned from re-writing classic arcade games. In: Future Play, The International Conference on the Future of Game Design and Technology, 2005, East Lansing, Michigan, USA. Proceedings... 2005.
- Bierre, K. J.; Phelps, A. M. The use of MUPPETS in an introductory java programming course. In: conference on Information technology education, 5., 2004, Salt Lake City, UT, USA. Proceedings... New York:ACM, 2004. p. 122-127.
- Cliburn, D. C. The effectiveness of games as assignments in an introductory programming course. In: Annual ASEE/IEEE Frontiers in Education Conference, 36. 2006, San Diego, California, USA. Proceedings... USA:IEE, 2006. p. 6–10.
- Cristovão, H. M. Aprendizagem de Algoritmos num Contexto Significativo e Motivador: um relato de experiência. In: Congresso da SBC – WEI – Workshop sobre Educação em Computação, 18., Belém do Pará, Pará. Anais... 2008.
- Dale, N. B. Most difficult topics in CS1: results of an online survey of educators. ACM SIGCSE Bulletin, USA, v. 38, n.2, p. 49-53, jun. 2006.
- Eagle, M.; Barnes, T. Experimental Evaluation of an Educational Game for Improved Learning in Introductory Computing. In: Technical Symposium on Computer Science Education, 40., Chattanooga, USA. Proceedings... USA: ACM, 2009. p. 321-325.
- Feldgen, M.; Clúa, O. Games as a motivation for freshman students learn programming. In: ASEE/IEEE Frontiers in Education, 34., Savannah, GA. Proceedings... 2004. p. S1H11-S1H16.
- Jesus, E. A.; Raabe, A. L. A. Interpretações da Taxonomia de Bloom no Contexto da Programação Introdutória. In: XX Simpósio Brasileiro de Informática na Educação, Florianópolis, 2009.
- Krathwohl, D. R. A revision of bloom's taxonomy: an overview. Theory into Practice, n. 41, v. 4, p. 212-218, 2002.
- Long, J. Just For Fun: using programming games in software programming training and education. Journal of Information Technology Education, USA, v. 6, p. 279-290, 2007.
- ROBINS, A.; ROUNDTREE, Janet; ROUNDTREE, Nathan. Learning and Teaching Programming: a review and discussion. Computer Science Education, USA, v. 13, n. 2, p. 137–172, 2003.
- Sheard, J. et al. Analysis of research into the teaching and learning of programming. In: INTERNATIONAL COMPUTING EDUCATION RESEARCH WORKSHOP, 5., 2009, Berkeley, USA. Proceedings... USA: ACM, 2009.
- Wainer, J. Métodos de pesquisa quantitativa e qualitativa para a Ciência da computação. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 17. Rio de Janeiro, Rio de Janeiro. Anais..., SBC, 2007.