

dTECT: Um Modelo para a Avaliação de Unidades Instrucionais para o Ensi- no de Computação na Educação Básica

Autores:

Christiane G. von Wangenheim

Giani Petri

André W. Zibetti

Adriano F. Borgatto

Jean C. R. Hauck

Fernando S. Pacheco

Raul Missfeldt Filho

Versão 1.0

Status: Final

Distribuição: Externa

MAIO - 2017

© 2017 **INCoD** – Instituto Nacional para Convergência Digital

Todos os direitos reservados e protegidos pela Lei 9.610 de 19/02/1998. Nenhuma parte deste documento, sem autorização prévia por escrito do Instituto, poderá ser reproduzida ou transmitida sejam quais forem os meios empregados: eletrônicos, mecânicos, fotográficos, gravação ou quaisquer outros.



Instituto Nacional para Convergência Digital

Universidade Federal de Santa Catarina - UFSC
Campus Universitário João David Ferreira Lima - Trindade
Departamento de Informática e Estatística - Sala 320
Florianópolis-SC - CEP 88040-970

Fone / FAX: +55 48 3721-9516 R.17
www.incod.ufsc.br

ISSN 2236-5281

Relatório Técnico do Instituto Nacional para Convergência Digital/
Departamento de Informática e Estatística, Centro Tecnológico,
Universidade Federal de Santa Catarina. -- v.1, n.1 (2011).--
Florianópolis: INE, UFSC, 2011 -

Semestral
Resumo em português

ISSN 2236-5281

1. Convergência digital. 2. Tecnologia da informação.
3. Informática na saúde. 4. Mídia digital I. Universidade Federal
de Santa Catarina. Departamento de Informática e Estatística.

Resumo

A computação vem sendo inserida na Educação Básica por meio de diversas iniciativas que objetivam popularizar as competências de computação, e ao mesmo tempo, conscientizar e despertar o interesse dos alunos. No entanto, é importante que se tenha evidências dos benefícios esperados destas iniciativas, para então selecioná-las e adotá-las de forma sistemática. Atualmente, a avaliação da qualidade dessas iniciativas é limitada e, por vezes, inexistente. Em muitos casos, a decisão de se utilizar uma destas iniciativas é baseada em suposições de sua eficácia. Deste modo, atualmente, há uma falta de modelos de avaliação e/ou instrumentos de medição sistematicamente desenvolvidos e avaliados que sejam amplamente aceitos para avaliar a qualidade das iniciativas para o ensino de computação na Educação Básica. Assim, o objetivo deste relatório técnico é apresentar o desenvolvimento e avaliação do modelo dTECT (Evaluating TEaching CompuTing), um modelo para a avaliação da qualidade de unidades instrucionais focadas no ensino de computação na Educação Básica com base na percepção dos alunos.

1. Introdução

A inclusão do ensino de computação na Educação Básica é uma tendência mundial que vem se concretizando de diversas maneiras, inclusive informais, como acampamentos de verão, clubes ou em oficinas de pais e filhos (GRESSE VON WANGENHEIM; WANGENHEIM, 2014). Atualmente, existem diversas iniciativas para ensinar computação como, por exemplo, Code.org (<http://www.code.org>), Code.club (<https://www.codeclubworld.org>), Computação na Escola (<http://www.computacaonaescola.ufsc.br/>), entre outras. Com estas iniciativas espera-se a popularização de competências de computação, e ao mesmo tempo, a conscientização e o interesse dos alunos sobre a área de computação (GUZDIAL et al., 2014; GARNELI et al., 2015).

Neste contexto, levando em consideração o número crescente de alternativas de ambientes de programação e unidades instrucionais (UIs), é importante que se tenha evidências dos benefícios esperados destas iniciativas, para então selecioná-las e adotá-las de forma sistemática (DECKER et al., 2016). Segundo Guzdial (2004), as maiores contribuições a serem feitas nesta área não estão na construção de novos ambientes/linguagens de programação, mas em descobrir como estudar os ambientes existentes. Uma compreensão mais precisa a respeito dos resultados do uso dos ambientes de programação permitiria saber se contribuem, de fato, positivamente ao alcance dos seus objetivos e se compensam os custos e esforços envolvidos em adotá-los. No entanto, embora se tenha indícios de que as UIs existentes possam ser ferramentas capazes de aprimorar o processo de ensino e aprendizagem na Educação Básica, e que esse tipo de iniciativa começa a ser usada mais amplamente nas escolas mundialmente, ainda há poucas pesquisas voltadas à análise da contribuição que estas UIs podem trazer para a educação (DECKER et al., 2016).

Atualmente, a avaliação da qualidade dessas UIs é limitada e, por vezes, inexistente (DECKER et al., 2016; GARNELI et al., 2015). Em muitos casos, a decisão de se utilizar uma destas UIs é baseada em suposições de sua eficácia (GROSS; POWERS, 2005; WILSON et al., 2010). Por outro lado, alguns estudos focam exclusivamente na avaliação do grau da aprendizagem alcançado (GROSS; POWERS, 2005; KALELIOĞLU; GÜLBAHAR, 2014). Outros estudos focam em perguntas específicas, como por exemplo, a efetividade de linguagens de programação baseada em blocos (WEINTROP; WILENSKY, 2015; GROVER et al., 2014; PERDIKURI, 2014). No entanto, as percepções e intenções dos alunos também são fatores determinantes para o sucesso da aprendizagem (GIANNAKOS et al., 2013). Porém, poucas avaliações levam também em consideração outros aspectos como a motivação e a experiência do aluno durante a unidade instrucional (CRAIG; HORTON, 2009; GIANNAKOS et al., 2014), ou ainda, as atitudes dos estudantes em relação a aceitação da tecnologia (GIANNAKOS et al., 2013). Neste contexto, há trabalhos que apresentam escalas para mensurar a atitude dos alunos em relação à computação (GARLAND; NOYES, 2008). No entanto, estas escalas são projetadas para a educação superior e parecem ser ultrapassadas no contexto atual de ensino de computação na Educação Básica (GARLAND; NOYES, 2008).

As medições utilizadas para avaliar a qualidade das UIs de ensino de computação variam bastante, desde escalas genéricas para avaliar as atitudes dos alunos em relação a computação até instrumentos de medição desenvolvidos de maneira *ad-hoc*. Muitos são desenvolvidos sem a definição de um modelo para derivar os itens do instrumento de medição com base em constructos teóricos, o que pode deixar a validade dos resultados questionável. Deste modo, atualmente, há uma falta de modelos de avaliação e/ou instrumentos de medi-

ção sistematicamente desenvolvidos e avaliados que sejam amplamente aceitos para avaliar a qualidade de UIs para o ensino de computação na Educação Básica.

Contudo, principalmente quando as UIs são realizadas de maneira mais informal, por exemplo, em oficinas de ensino de programação para pais e filhos fora do ambiente escolar, pode ser inviável a realização de experimentos que necessitem a realização de pré-testes e de inclusão de grupos de controle, pois pode causar uma maior interrupção e pode influenciar o grau de divertimento propiciado pela oficina. Diante disto, no contexto de UIs mais informais uma alternativa viável pode ser a condução de estudos de caso, onde a avaliação da UI é realizada somente ao final da oficina (pós-teste), tipicamente por meio de um questionário para obter a percepção dos alunos (WOHLIN et al., 2012). Uma vantagem deste tipo de estudo é que a avaliação pode ser executada com pouco esforço e de uma forma não intrusiva ao final da unidade instrucional. Estudos baseados na medição de percepções, utilizando instrumentos de medição, tipicamente questionários, são conduzidos em uma variedade de diferentes áreas de pesquisa (DEVELLIS, 2016; TAKATALO et al., 2010, SWEETSER; WYETH, 2005; POELS et al., 2007). Assim, o objetivo deste capítulo é apresentar o desenvolvimento e avaliação do modelo dETECT (*Evaluating TEaching CompuTing*), um modelo para a avaliação da qualidade de unidades instrucionais focadas no ensino de computação na Educação Básica com base na percepção dos alunos.

2. Metodologia

Com o objetivo de desenvolver um modelo de avaliação de unidades instrucionais para o ensino de computação foi realizada uma pesquisa aplicada (MILLER; SALKIND, 2002), dividida em quatro etapas (Figura 1).

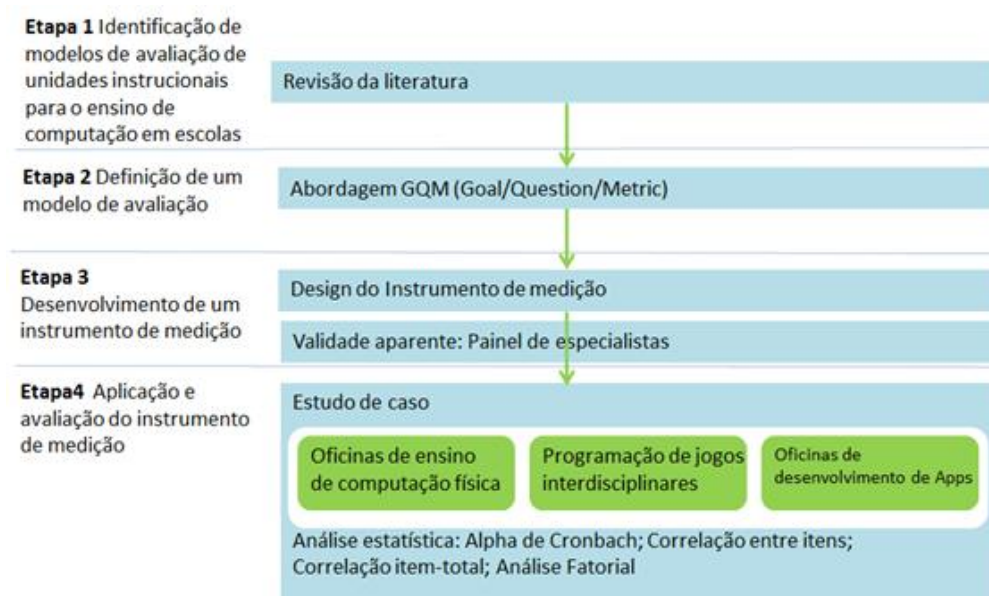


Figura 1. Metodologia

Etapa 1. Levantamento do estado da arte. Foi realizada uma revisão da literatura sobre a bibliografia relacionada a modelos de avaliação de unidades instrucionais para o ensino de computação nas escolas.

Etapa 2. Definição de um modelo de avaliação. O modelo de avaliação dTECT foi desenvolvido para medir a qualidade das unidades instrucionais utilizando a abordagem GQM (Goal/Question/Metric) (BASILI et al., 1994), sistematicamente definindo o objetivo da avaliação e decompondo-o em questões de análise e medidas.

Etapa 3. Definição do instrumento de medição. De modo a operacionalizar a medição, um questionário foi desenvolvido por uma equipe multidisciplinar seguindo (DEVELLIS, 2016; KROSNICK; PRESSER, 2010; MALHOTRA, 2008; KASUNIC, 2005). A definição dos itens do questionário é baseada em itens de estudos semelhantes que foram encontrados na literatura, considerados aderentes ao contexto do presente estudo e ao plano de medição definido.

Etapa 4. Aplicação e avaliação do instrumento de medição. Foi realizado um estudo de caso (YIN, 2009; WOHLIN et al., 2012) para avaliar o instrumento de medição em termos de confiabilidade e validade de constructo. Para a definição da avaliação foi utilizada a abordagem GQM. O objetivo do estudo foi decomposto em aspectos de qualidade e questões de análise para serem analisados na avaliação do instrumento de medição (CARMINES; ZELLER, 1979; DEVELLIS, 2016; TROCHIM; DONNELLY, 2008). Durante o estudo de caso foi aplicado o instrumento de medição do modelo dTECT como parte da avaliação de 16 cursos/oficinas de computação realizados em diferentes instituições de ensino.

A pesquisa foi aprovada pelo CEPESH - Comitê de Ética em Pesquisa com Seres Humanos da UFSC, conforme parecer no. 1.021.541.

3. O Modelo dTECT (*Evaluating TEaching CompuTing*)

O objetivo do modelo de avaliação dTECT é: analisar unidades instrucionais com o propósito de avaliar a percepção da qualidade em termos de qualidade da UI, experiência de computação e percepção da aprendizagem do ponto de vista de alunos no contexto de UIs focadas no ensino de computação na Educação Básica. A partir deste objetivo são derivadas as perguntas de análise e as medidas com base na literatura (Figura 2).

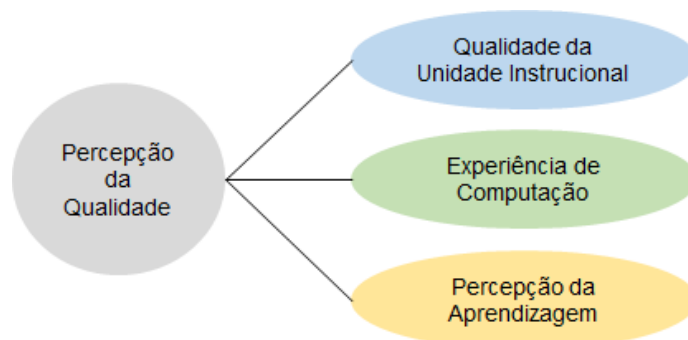


Figura 2. Decomposição dos fatores de qualidade

Entende-se, neste trabalho, que uma unidade instrucional (oficina, curso, etc.) de qualidade é aquela que atinge os seus objetivos de aprendizagem, promove atividades prazerosas que facilitam a aprendizagem e que desperta uma percepção positiva e interesse referente à área de computação.

A medição é operacionalizada pelo desenvolvimento de um instrumento de medição (questionário) a ser respondido pelos alunos após o término da unidade instrucional, de modo a obter a sua percepção sobre qualidade da oficina/curso. Os itens que compõem o questionário (Tabela 1) são derivados de estudos similares encontrados na literatura, considerados aderentes ao contexto desta pesquisa.

Tabela 1. Questionário

No.	Descrição do Item	Formato das respostas
Qualidade da Unidade Instrucional		
1	A oficina foi:	(1) Muito Divertida (2) Divertida (3) Chata (4) Muito chata
2	O tempo das aulas passou:	(1) Muito rápido (2) Rápido (3) Devagar (4) Muito devagar
3	A oficina foi:	(1) Excelente (2) Boa (3) Regular (4) Ruim
Experiência de Computação		
4	Vou mostrar meu programa de computador para outras pessoas:	(1) Sim (2) Não
5	Quero aprender mais sobre como fazer programas de computador:	(1) Sim (2) Não

6	Fazer um programa de computador é:	(1) Muito Divertido (2) Divertido (3) Chato (4) Muito chato
7	Eu gosto de fazer programas de computador:	(1) Sim (2) Não
8	A computação é útil no dia a dia:	(1) Sim (2) Não
9	Você já pensou em trabalhar com computação?	(1) Sim (2) Não
Percepção da Aprendizagem		
10	A oficina foi:	(1) Muito fácil (2) Fácil (3) Difícil (4) Muito Difícil
11	Eu consigo fazer programas de computador:	(1) Sim (2) Não
12	Consigo explicar para um amigo/amiga como fazer um programa de computador:	(1) Sim (2) Não
13	Fazer um programa de computador é:	(1) Muito fácil (2) Fácil (3) Difícil (4) Muito Difícil

O material completo do modelo dETECT está disponível em

http://www.computacaonaescola.ufsc.br/?page_id=45.

4. Definição e Execução da Avaliação do Modelo dTECT

Ao desenvolver modelos de avaliação e questionários, é fundamental analisar se eles estão medindo o que se pretende medir (validade de constructo) e se o mesmo processo de medição produz os mesmos resultados (confiabilidade) (CARMINES; ZELLER, 1979). Por isso, foi avaliado o instrumento de medição do modelo dTECT em termos de confiabilidade e validade de constructo sob o ponto de vista de pesquisadores no contexto de unidades instrucionais para o ensino de computação na Educação Básica. São analisadas as seguintes questões de análise:

Confiabilidade

QA1: Há evidência de consistência interna no instrumento de medição do modelo dTECT?

Validade de Constructo

QA2: Há evidência de validade convergente e discriminante no instrumento de medição do modelo dTECT?

QA3: Quantos fatores subjacentes influenciam as respostas dos itens do instrumento de medição do modelo dTECT?

Para a avaliação do modelo dTECT foram executados 16 estudos de caso aplicando três diferentes unidades instrucionais em 13 diferentes instituições de ensino entre 2015 e 2016, totalizando 477 alunos envolvidos (Tabela 2).

Tabela 2. Visão geral das aplicações e dados coletados

Unidade Instrucional	Local/Data	Número de alunos
Oficina de Computação Física	INE/UFSC - Florianópolis – 08/08/2015	8
	Escola Hamônia - Ibirama/SC – 29/08/2015	13
	INE/UFSC - Florianópolis – 17/10/2015	14
	IFSC - Gaspar/SC – 20 e 22/10/2015	32
	Escola Sabedoria Junior – Florianópolis/SC – 04/11/2015	22
	INE/UFSC – Florianópolis – 07/11/2015	16
	INE/UFSC – Florianópolis – 14/11/2015	15
Unidade Jogos com Scratch	Turmas 5Mat, 5 Vesp, 7A, 7B – Escola Autonomia – Florianópolis/SC	99
	Escola Básica Municipal Prefeito Reinaldo Weingartner –Palhoça/SC	25
	EEB Prof Vitorio Anacleto Cardoso - Gaspar/SC	43
	EEB Zenaide Schmitt Costa - Gaspar/SC	31
	EEB Luiz Franzoi - Gaspar/SC	15
	EEB Ferandino Dagnoni - Gaspar/SC	46
	EEB Prof Dolores dos Santos Krauss - Gaspar/SC	14
	EEB Norma Mônica Sabel - Gaspar/SC	49
Oficina de App Inventor	Turma 1 e 2 – Escola Básica Prof. ^a Herondina Medeiros Zeferino 2016	35
Total		477

O público-alvo são alunos no nível da Educação Básica levando em consideração diversos tipos de unidades instrucionais, tanto dentro de escolas, quanto de forma extracurricular no Brasil. As unidades instrucionais são voltadas ao ensino de computação focando na programação e pensamento computacional (Tabela 3).

Tabela 3. Visão geral das unidades instrucionais aplicadas

Oficina de Computação Física	Unidade Jogos com Scratch	Oficina de App Inventor
		
Integrando Scratch/Snap! com Arduino e peças de hardware numa solução baixo custo, os alunos aprendem dar vida a um robô interativo.	De forma interdisciplinar os alunos aprendem conceitos básicos da computação programando jogos envolvendo diversos conteúdos (p.ex. história, língua portuguesa, geografia, etc.) utilizando Scratch.	Os alunos aprendem programar um app de jogo "Caça Mosquito" utilizando o App Inventor.
		

4.1 Análise dos dados coletados

De modo a obter uma maior precisão e poder estatístico, aumentando o tamanho da amostra, os dados coletados nos 16 estudos de caso foram analisados cumulativamente para responder as questões de análise definidas.

Confiabilidade

QA1: Há evidência de consistência interna no instrumento de medição do modelo dE-TECT?

A consistência interna foi medida por meio do Alfa de Cronbach (1951). O coeficiente alfa indica indiretamente o grau em que um conjunto de itens está medindo um único fator de qualidade (a percepção da qualidade da unidade instrucional). Tipicamente, valores de alfa de Cronbach maiores que 0,7 são considerados aceitáveis (DEVELLIS, 2016), assim, indicando a consistência interna do instrumento de medição. Analisando os 13 itens que compõem o instrumento de medição do modelo dE-TECT (Tabela 1), o valor do coeficiente alfa de Cronbach é aceitável ($\alpha=.787$). Deste modo, pode-se concluir que as respostas entre os

itens são consistentes e precisas, indicando a confiabilidade dos itens do instrumento de medição do modelo dTECT.

Validade de Constructo

QA2: Há evidência de validade convergente e discriminante no instrumento de medição do modelo dTECT?

Validade de constructo de um instrumento de medição refere-se à habilidade de medir o que ele se propõe a medir (CARMINES; ZELLER, 1979; TROCHIM; DONNELLY, 2008). Validade convergente e discriminante são dois subtipos de validade que compõem a validade de constructo. A validade convergente refere-se ao grau que dois itens de um fator de qualidade que teoricamente devem estar relacionados, de fato, estão relacionados. Em contrapartida, validade discriminante testa se conceitos ou medidas que supostamente não devem ter relação, de fato, não estão relacionados (TROCHIM; DONNELLY, 2008). De modo a analisar a validade convergente e discriminante do instrumento de medição do modelo dTECT, as intercorrelações dos itens e correlação item-total são calculadas (DEVELLIS, 2016).

Intercorrelação dos itens. Para analisar a intercorrelação entre os itens, utilizou-se uma matriz de correlação não-paramétrica de Spearman (Tabela 4). A matriz mostra o coeficiente de correlação de Spearman, indicando o grau de correlação entre dois itens (par de itens). Este coeficiente foi utilizado por ser o coeficiente mais apropriado para análises de escalas Likert. Os coeficientes de correlação entre os itens de um mesmo fator de qualidade estão marcados em cores. De acordo com Cohen (1988), uma correlação entre itens é considerável satisfatória, se o coeficiente de correlação é maior que 0,29, indicando que há uma média ou alta correlação entre os itens. Coeficientes de correlação satisfatórios estão destacados em negrito.

Tabela 4. Matriz de correlação de Spearman

Item/ Fator de qualidade	1	2	3	4	5	6	7	8	9	10	11	12	13
	Qualidade da UI				Experiência de Computação					Percepção da Aprendizagem			
1	1,000												
2	,268	1,000											
3	,514	,294	1,000										
4	,243	,090	,230	1,000									
5	,291	,137	,295	,362	1,000								
6	,452	,174	,382	,266	,385	1,000							
7	,211	,160	,181	,325	,378	,395	1,000						
8	,046	,028	,035	,240	,327	,312	,345	1,000					
9	,221	,107	,262	,203	,324	,226	,263	,197	1,000				
10	,174	,203	,153	,067	,133	,144	,022	-,086	,125	1,000			
11	,213	,059	,087	,273	,320	,318	,287	,239	,202	,157	1,000		
12	,162	,094	,135	,312	,246	,194	,266	,192	,247	,143	,450	1,000	
13	,156	,082	,155	,168	,266	,325	,225	,226	,199	,391	,432	,286	1,000

Analisando as intercorrelações entre os itens dos três fatores de qualidade (Tabela 4), pode-se observar que a maioria dos pares de itens possuem média ou alta correlação considerando cada fator de qualidade. Porém, alguns pares de itens apresentaram uma baixa correlação (ex., 1-2, 6-9, 10-11). Mesmo assim, os resultados indicam uma evidência de validade convergente. Por outro lado, alguns pares de itens (ex., 1-6, 3-6, 5-11) apresentaram média ou alta correlação com itens de outro fator de qualidade. Deste modo, não há evidência de validade discriminante. No entanto, a falta de existência de validade discriminante é aceitável, pois, embora o modelo seja dividido em três fatores de qualidade, todos fatores também estão relacionados a um único fator, que é a percepção da qualidade da UI.

Correlação item-total. A correlação item-total é analisada de modo a verificar se qualquer item do instrumento é inconsistente com a média de correlação dos outros itens, e assim, pode ser eliminado (CARMINES; ZELLER, 1979; DEVELLIS, 2016). A Tabela 5 mostra o coeficiente de correlação entre um único item e os outros itens (excluindo o item referido) do instrumento de medição. As correlações são consideradas satisfatórias se o coeficiente de correlação é maior que 0,29 (COHEN, 1988). Itens que apresentam uma baixa correlação são destacados em negrito. A Tabela 5 também apresenta o coeficiente alfa de Cronbach se o item for excluído, neste caso, espera-se que nenhum item ao ser retirado aumente o alfa de Cronbach (DEVELLIS, 2016) do instrumento de medição.

Tabela 5. **Correlação item-total corrigida**

Fator de Qualidade	No. item	Correlação de item total corrigida	Alfa de Cronbach se o item for excluído
Qualidade da UI	1	,511	,764
	2	,269	,794
	3	,459	,769
Experiência de Computação	4	,431	,773
	5	,511	,769
	6	,594	,753
	7	,481	,771
	8	,338	,783
	9	,410	,774
	10	,280	,787
Percepção da Aprendizagem	11	,470	,770
	12	,415	,773
	13	,474	,769

Em geral, as correlações item-total são médias e altas. A maioria dos itens demonstram aceitável correlação item-total e valores satisfatórios de coeficiente alfa de Cronbach se item foi excluído, indicando assim, a validade dos fatores de qualidade. Somente os itens 2 (“O tempo das aulas passou”) e 10 (“A oficina foi:”) apresentaram uma baixa correlação item-total, além disso o item 2 apresenta um pequeno acréscimo no alfa de Cronbach caso seja excluído, deste modo, os resultados indicam que estes itens precisam ser revisados.

QA3: Quantos fatores subjacentes influenciam as respostas dos itens do instrumento de medição do modelo dTECT?

De modo a determinar quantos fatores de qualidade estão subjacentes ao conjunto de itens do instrumento de medição do modelo dTECT foi realizada uma análise fatorial seguindo (BROWN, 2006). Para analisar se os itens do instrumento de medição do modelo dTECT podem ser submetidos a uma análise fatorial, utilizou-se o índice de Kaiser-Meyer-Olkin (KMO). Como resultado, obteve-se um índice KMO de 0,827, indicando que a análise fatorial é apropriada neste caso. Aplicando a análise fatorial, o número de fatores retidos na análise é decidido com base no critério de Kaiser-Guttman, que determina que o número de fatores é igual ao número de fatores com autovalor maior que 1 (CARMINES; ZELLER, 1979). Deste modo, justificando a retenção de três fatores que estão influenciando as respostas das avaliações das UIs. Em relação ao modelo dTECT, isto significa que as respostas do instrumento de medição estão representando três fatores subjacentes, indicando uma decomposição igual a definição original do modelo.

Uma vez identificado o número de fatores subjacentes, outra questão é determinar quais itens contribuem para qual fator. A fim de identificar as cargas fatoriais dos itens, o método de rotação Promax com Normalização de Kaiser é usado, por ser o método de rotação mais amplamente aceito e utilizado (CARMINES; ZELLER, 1979). A Tabela 6 mostra as cargas

fatoriais dos itens associado com os três fatores retidos. A maior carga fatorial de cada item, indicando a qual fator o item está mais relacionado, é marcado em negrito.

Tabela 6. Cargas fatoriais

Fator de Qualidade	No. item	Descrição	Fatores		
			1	2	3
Qualidade da UI	1	A oficina foi:	,146	,763	-,018
	2	O tempo das aulas passou:	-,096	,619	,043
	3	A oficina foi:	,110	,800	-,078
Experiência de Computação	4	Vou mostrar meu programa de computador para outras pessoas:	,591	,101	,008
	5	Quero aprender mais sobre como fazer programas de computador:	,571	,217	,035
	6	Fazer um programa de computador é:	,510	,400	,055
	7	Eu gosto de fazer programas de computador:	,683	,114	-,053
	8	A computação é útil no dia a dia:	,783	-,213	-,090
Percepção da Aprendizagem	9	Você já pensou em trabalhar com computação?	,401	,130	,165
	10	A oficina foi:	-,425	,239	,823
	11	Eu consigo fazer programas de computador:	,415	-,177	,546
	12	Consigo explicar para um amigo/amiga como fazer um programa de computador:	,432	-,139	,410
	13	Fazer um programa de computador é:	,230	-,102	,720

Analisando as cargas fatoriais dos itens, pode-se observar que, o primeiro fator (fator 1) é representado por um conjunto de 7 itens (4, 5, 6, 7, 8, 9 e 12). Assim, este fator está diretamente relacionado ao fator de qualidade sobre a experiência de computação proporcionado pela unidade instrucional (Tabela 2). Pois, com exceção do item 12, todos os itens correspondem ao referido fator de qualidade no modelo dTECT. Embora o item 12 apresenta a maior carga fatorial no fator 1, ele também apresenta uma carga fatorial semelhante (,410) no fator 3, portanto, mostrando que este item contribui para os dois fatores de qualidade (experiência de computação e percepção da aprendizagem). Em relação ao fator 2, é considerado um conjunto de três itens (1, 2 e 3). Esse resultado parece sugerir que esses itens estão relacionados ao fator sobre a qualidade da unidade instrucional do modelo dTECT. Pois, de fato, esses referidos itens correspondem a um mesmo fator de qualidade (qualidade da UI) na definição original do modelo dTECT (Tabela 1). Analisando os resultados do fator 3, ele inclui um conjunto de três itens (10, 11 e 13), indicando que estes itens estão relacionados a um único fator de qualidade.

5. Discussão

Os resultados obtidos apresentam evidências suficientes para considerar o modelo dTECT confiável e válido para a avaliação de unidades instrucionais para o ensino de computação na Educação Básica.

Em termos de confiabilidade, os resultados da análise indicam um coeficiente alfa de Cronbach aceitável (Cronbach's alpha $\alpha=.787$), indicando a consistência interna do conjunto de itens do instrumento de medição do modelo dTECT. Assim, indicando que os itens do instrumento de medição do modelo dTECT são consistentes e precisos para avaliar a qualidade de unidades instrucionais.

Em termos de validade de constructo, em relação a validade convergente, identificou-se que os itens tendem a ter uma média e alta correlação entre os itens, principalmente entre os itens de um mesmo fator de qualidade (ex., qualidade da UI, experiência de computação e percepção da aprendizagem). Deste modo, pode-se concluir que existe evidência de validade convergente considerando os fatores de qualidade. Isso indica que, os itens do instrumento de medição do modelo dTECT parecem estar medindo realmente o que se propõem a medir (ex., qualidade da UI, experiência de computação e percepção da aprendizagem). No entanto, alguns itens apresentam baixa correlação, tanto em um único fator de qualidade quanto em relação aos outros fatores, isto pode indicar que alguns itens precisam ser revisados, principalmente quanto a sua descrição. Com relação a validade discriminante, em geral, a maioria dos itens apresentam uma correlação baixa com itens de outro fator de qualidade. Porém, alguns pares de itens (ex., 1-6, 5-11) apresentaram uma média ou alta correlação com itens de outro fator de qualidade. Deste modo, os resultados não indicam evidências de validade discriminante. No entanto, neste caso, a falta de existência de validade discriminante é aceitável, pois, embora o modelo seja dividido em três fatores de qualidade, todos fatores também estão relacionados a um único fator, que é a percepção da qualidade da unidade instrucional, como proposto na composição original do modelo dTECT (Figura 2).

Analisando a correlação item-total, novamente, a grande maioria dos itens apresentam valores de correlação satisfatórios com os outros itens do instrumento de medição, indicando que o conjunto de itens do instrumento de medição do modelo dTECT estão relacionados para medir o que se propõem a medir (percepção da qualidade de UI).

Com base nos resultados da análise fatorial, identificou-se que os dados coletados nos estudos de caso são explicados por três fatores. Confirmando assim, a estrutura inicial definida para o modelo dTECT, claramente agrupando os itens conforme seu fator de qualidade definido (qualidade da UI, experiência de computação e percepção da aprendizagem).

Ameaças à validade

Devido as características deste tipo de pesquisa, este trabalho está sujeito a ameaças à validade. Assim, foram identificadas possíveis ameaças e aplicadas estratégias de mitigação de modo a minimizar o seu impacto nesta pesquisa. Algumas ameaças estão relacionadas ao tipo de estudo. Deste modo, para minimizar esta ameaça foi definido e documentado uma metodologia sistemática para este trabalho. O modelo dTECT foi sistematicamente definido com base na abordagem GQM, decompondo sistematicamente o objetivo de avaliação em questões de análise e medidas. O instrumento de medição, foi desenvolvido seguindo um método definido na literatura e envolvendo uma equipe multidisciplinar de pes-

quisadores. Além disso, para a avaliação do instrumento de medição do modelo dTECT um estudo de caso também foi sistematicamente definido e documentado. Outro risco refere-se à qualidade dos dados agrupados em uma amostra, em termos de padronização dos dados e adequação ao modelo. Em relação a isso, como este estudo é exclusivo para avaliar UIs que utilizam o modelo dTECT, o risco é minimizado, pois todos os estudos de caso utilizam o mesmo instrumento de medição (com mesmo padrão de resposta) para a coleta de dados. Outra ameaça refere-se ao agrupamento de dados de diferentes contextos. Para minimizar este risco, foram considerados estudos de caso somente de UIs para o ensino de computação exclusivamente em nível de Educação Básica.

Em termos de validade externa, uma ameaça para generalizar os resultados está relacionada ao tamanho da amostra e diversidade dos dados utilizados na avaliação. Em relação a isso, a avaliação do modelo dTECT usou dados coletados em 16 estudos de caso avaliando três diferentes unidades instrucionais em 13 instituições de Educação Básica, totalizando uma amostra de 477 estudantes. Em termos de significância estatística, o tamanho da amostra é considerado satisfatório permitindo a geração de resultados significantes.

Em relação a confiabilidade do estudo, uma ameaça refere-se a quanto os dados e a análise são dependentes dos pesquisadores específicos. Para mitigar esta ameaça, foi documentado uma metodologia sistemática, tanto para o desenvolvimento do modelo dTECT quanto para o estudo de caso de sua avaliação, definindo claramente o objetivo do estudo, o processo de coleta de dados e os métodos estatísticos utilizados para a análise dos dados. Além disso, outra ameaça refere-se a correta escolha dos métodos estatísticos para a análise dos dados. Para minimizar esta ameaça, a análise estatística baseou-se no guia de desenvolvimento de escalas proposto por Devellis (2016), que está alinhado aos procedimentos para a avaliação de consistência interna e validade de constructo de um instrumento de medição (Trochim & Donnelly, 2008).

6. Conclusões

Embora a avaliação de unidades instrucionais para o ensino de computação seja essencial para sua melhoria contínua, efetiva e eficiente aplicação, poucos esforços têm sido feitos para o desenvolvimento de modelos sistemáticos de avaliação. Neste contexto, este trabalho apresenta uma primeira proposta na direção de levar em consideração restrições práticas ao executar essas avaliações em programas de extensão mais informais.

Com base na literatura e em experiências práticas, o modelo de avaliação dTECT, incluindo um instrumento de medição com 13 itens foi desenvolvido sistematicamente e aplicado (pós-unidade) em 16 oficinas/cursos na Educação Básica no Brasil. Os resultados da análise das respostas de 477 alunos, indicam que o instrumento de medição do modelo dTECT é aceitável em termos de confiabilidade e validade de constructo. Em termos de confiabilidade, o Alfa de Cronbach é de $\alpha=0,787$, indicando uma aceitável consistência interna, o que significa que as respostas entre os itens são consistentes e precisas. Os resultados da análise também indicam um aceitável grau de correlação entre os itens em relação aos fatores de qualidade. Assim, sugerindo que o instrumento de medição do modelo dTECT pode ser um instrumento confiável e válido para medir a percepção dos estudantes em relação a qualidade de unidades instrucionais para o ensino de computação. Os resultados da análise fatorial, indicam que três fatores subjacentes influenciam as respostas dos itens do instrumento de medição do modelo dTECT. Este resultado confirma a estrutura original do modelo dTECT (Figura 2), que define três fatores de qualidade (qualidade da UI, experiência de computação e percepção da qualidade) para a avaliação de unidades instrucionais.

Agradecimentos

Este trabalho teve suporte do CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico – www.cnpq.br), uma entidade do governo brasileiro com foco no desenvolvimento científico e tecnológico e o *Google Rise Award*.

Referências

- BASILI, V. R.; CALDIERA, G.; ROMBACH, H. D. **Goal, Question Metric Paradigm**. In: MARCINIAK, J. J. (Ed.). *Encyclopedia of Software Engineering*. John Wiley & Sons, 1994. p. 528–532.
- BROWN, T. A. **Confirmatory factor analysis for applied research**. 1. ed. New York: The Guilford Press, 2006. 475 p.
- CARMINES, E. G.; ZELLER, R. A. **Reliability and validity assessment**. 1. ed. Beverly Hills: Sage Publications Inc., 1979. 75 p.
- CRAIG, M; HORTON, D. Gr8 designs for Gr8 girls: a middle-school program and its evaluation. In: 40TH ACM TECHNICAL SYMPOSIUM ON COMPUTER SCIENCE EDUCATION, 2009. Chattanooga. **Anais...** Chattanooga, TN, USA, 2009, p. 221-225.
- CRONBACH, L. J. Coefficient alpha and the internal structure of tests. **Psychometrika**, vol. 16., no. 3, p. 297–334, 1951.
- DECKER, A.; MCGILL, M. M.; Settle, A. Towards a Common Framework for Evaluating Computing Outreach Activities. In: ACM TECHNICAL SYMPOSIUM ON COMPUTER SCIENCE EDUCATION, 2016. Memphis. **Anais...** Memphis, TN, USA, 2016 p. 627-632.
- DEVELLIS, R. F. **Scale development: theory and applications**. 4. ed. SAGE Publications, 2016. 280 p.
- GARLAND, K. J.; NOYES, J. M. Computer attitude scales: How relevant today? **Computers in Human Behavior**, vol. 24, no. 2, p. 563–575, 2008.
- GARNELI, V.; GIANNAKOS, M. N.; CHORIANOPOULOS, K. Computing Education in K-12 Schools: A Review of the Literature. In: IEEE GLOBAL ENGINEERING EDUCATION CONFERENCE, 2015. Tallin. **Anais...** Tallinn, Estonia, 2015. p. 543-551.
- GIANNAKOS, M. N.; HUBWIESER, P.; CHRISOCHOIDES, N. How Students Estimate the Effects of ICT and Programming Courses. In: ACM TECHNICAL SYMPOSIUM ON COMPUTER SCIENCE EDUCATION, 2013. Denver. **Anais...** Denver, CO, USA, 2013. p. 717-722.
- GIANNAKOS, M. N.; JACCHERI, L.; LEFTHERIOTIS, I. Happy Girls Engaging with Technology: Assessing Emotions and Engagement Related to Programming Activities. In: INTERNATIONAL CONFERENCE ON LEARNING AND COLLABORATION TECHNOLOGIES, 2014. Heraklion. **Anais...** Heraklion, Greece, 2014, p. 398-409.
- GRESSE VON WANGENHEIM, C.; WANGENHEIM, A. Teaching Game Programming in Family Workshops. **IEEE Computer Magazine**, vol. 47, no. 8, p. 84-87, 2014.
- GROSS, P.; POWERS, K. Evaluating Assessments of Novice Programming Environments. In: INTERNATIONAL WORKSHOP ON COMPUTING EDUCATION RESEARCH, 2005. Seattle. **Anais...** Seattle, WA, USA, 2005. p. 99-110.

GROVER, S.; COOPER, S.; PEA, R. Assessing computational learning in K-12. In: CONFERENCE ON INNOVATION & TECHNOLOGY IN COMPUTER SCIENCE EDUCATION, 2014. Uppsala. **Anais...** Uppsala, Sweden, 2014. p. 57–62.

GUZDIAL, M. **Programming environments for novices**. In Computer Science Education Research. S. Fincher and M. Petre (Eds.). Swets and Zeitlinger. Chapter 3, 2004.

GUZDIAL, M.; ERICSON, B.; MCKLIN, T.; ENGELMAN, S. Georgia Computes! An intervention in a US state, with formal and informal education in a policy context. **ACM Transactions on Computing Education**, vol. 14, no. 2, Article 13, 2014.

KALELIOĞLU, F.; GÜLBAHAR, Y. The Effects of Teaching Programming via Scratch on Problem Solving Skills: A Discussion from Learners' Perspective. **Informatics in Education**, vol. 13, no. 1, p. 33–50, 2014.

KASUNIC, M. **Designing an effective survey**. Carnegie-Mellon University Pittsburgh Pa Software Engineering Inst, 2005.

KROSnick, J. A.; PRESSER, S. **Questionnaire design**. In J. D. Wright & P. V. Marsden (Eds.), *Handbook of Survey Research*. 2. ed. West Yorkshire, England: Emerald Group, 2010.

MALHOTRA, N. K.; BIRKS, D. F. **Marketing Research: An Applied Approach**. 3. ed. Trans-Atlantic Publications, 2008. 816p.

MILLER, D. C.; SALKIND, N. J. **Handbook of Research Design and Social Measurement**. 6. ed. SAGE Publications, 2002. 808 p.

PERDIKURI, K. Students' Experiences from the use of MIT App Inventor in classroom. In: PANHELLENIC CONFERENCE ON INFORMATICS, 2014. Athens. **Anais...** Athens, Greece, 2014. p.1-6.

POELS, K.; KORT, Y. D.; IJSSELSTEIJN, W. (2007). **It is always a lot of fun!: exploring imensions of digital game experience using focus group methodology**. In: CONFERENCE ON FUTURE PLAY, 7., 2007. Toronto. **Proceedings....** New York: ACM, 2007. p. 83-89.

SWEETSER, P.; WYETH, P. GameFlow: a model for evaluating player enjoyment in games. **Computers in Entertainment**, vol. 3, no. 3, p. 1-24, 2005.

TAKATALO, J., HÄKKINEN, J., KAISTINEN, J., & NYMAN, G. (2010). **Presence, Involvement, and Flow in Digital Games**. In: Bernhaupt, R. (Ed.). *Evaluating User Experience in Games: Concepts and Methods*, Springer, 2010. p. 23-46.

TROCHIM, W. M.; DONNELLY, J. P. **Research methods knowledge base**. 3. ed. Mason, OH: Atomic Dog Publishing, 2008. 361p.

WEINTROP, D.; WILENSKY, U. To Block or Not to Block, That is the Question: Students' Perceptions of Blocks-based Programming. In: CONFERENCE ON INTERACTION DESIGN AND CHILDREN, 2015. New York. **Anais...** New York, NY, USA, 2015. p. 199-208.

WILSON, A.; MOFFAT, D. C. Evaluating Scratch to introduce younger schoolchildren to Programming. In: ANNUAL WORKSHOP OF THE PSYCHOLOGY OF PROGRAMMING INTEREST GROUP, 2010. Leganés. **Anais...** Leganés, Spain, 2010.

WOHLIN, C.; RUNESON, P.; HOST, M.; OHLSSON, M. C.; REGNELL, B.; WESSLEN, A. **Experimentation in Software Engineering**. Springer, 2012. 236p.

YIN, R.K. **Case study research: design and methods**. 4. ed. Beverly Hills: Sage Publications, 2009. 312p.