

LEONARDO FELIPE DE ÁVILA CALBUSCH

**VALIDAÇÃO DE CONTEÚDO DO *CTPUZZLE TEST*:
INSTRUMENTO PARA AVALIAÇÃO DO PENSAMENTO
COMPUTACIONAL DE JOVENS DE 14 A 17 ANOS**

Itajaí (SC), abril de 2021



UNIVALI

**UNIVERSIDADE DO VALE DO ITAJAÍ
CURSO DE MESTRADO ACADÊMICO EM
COMPUTAÇÃO APLICADA**

**VALIDADE DE CONTEÚDO DO *CT PUZZLE TEST*:
INSTRUMENTO PARA AVALIAÇÃO DO PENSAMENTO
COMPUTACIONAL DE JOVENS DE 14 A 17 ANOS**

por

Leonardo Felipe de Ávila Calbusch

Dissertação apresentada como requisito parcial à
obtenção do grau de Mestre em Computação
Aplicada.

Orientador: André Luis Alice Raabe, Doutor

Itajaí (SC), abril de 2021

“VALIDAÇÃO DE CONTEÚDO DO CTPUZZLE TEST: INSTRUMENTO PARA AVALIAÇÃO DO PENSAMENTO COMPUTACIONAL DE JOVENS DE 14 A 17 ANOS”

Leonardo Felipe de Ávila Calbusch

‘Esta Dissertação foi julgada adequada para obtenção do Título de Mestre em Computação Aplicada, Área de Concentração Computação Aplicada e aprovada em sua forma final pelo Programa de Mestrado Acadêmico em Computação Aplicada da Universidade do Vale do Itajaí.’

Prof. André Luís Alice Raabe, Dr.
Coordenador do Programa Mestrado em Computação Aplicada

Apresentado perante a Banca Examinadora composta pelos Professores:

Prof. André Luís Alice Raabe, Dr.
Orientador

Prof. Aluizio Haendchen Filho, Dr. (UNIVALI)
Membro

Prof. Rudimar Luís Scaranto Dazzi, Dr. (UNIVALI)
Membro

Profa. Julia Zanetti Rocca, Dra. (UFMT)
Membro

Prof. Wilkerson de Lucena Andrade, Dr. (UFCG)
Membro

AGRADECIMENTOS

A superação desse grande desafio nunca teria sido possível sem o apoio de várias pessoas queridas. Agradeço a Deus por toda a luz que recai em meu caminho. Agradeço ao Instituto Federal Catarinense, do qual orgulhosamente sou professor e que me concedeu condições para a dedicação necessária a esse trabalho. Agradeço a meus colegas de turma, aos colegas do LITE e do Grupo de Pesquisa em Informática na Educação da Univali por me acolherem e dividirem comigo suas experiências, críticas e sugestões que tanto enriqueceram este trabalho. Agradeço aos docentes, à secretaria e ao Colegiado de Curso do MCA, e ao quadro de técnicos da Univali, por seu apoio, sua dedicação, e pelo compromisso com seu trabalho e com a qualidade da formação de seus acadêmicos. Agradeço aos juízes que aceitaram participar da pesquisa e aos estudantes do IFC que se submeteram ao teste, possibilitando a coleta de dados da pesquisa.

Agradeço à minha família, a cada ente querido que de alguma forma e em algum momento da minha vida investiu parte da sua própria vida para me trazer até aqui. Sou grato a cada um por seus sacrifícios, seus ensinamentos, sua dedicação e seu amor. Sou resultado de suas escolhas, somadas as minhas próprias, em uma equação única que não chegaria a esse resultado se algo tivesse sido diferente.

Agradeço à minha maravilhosa esposa Michelle e à minha doce, meiga e carinhosa filha Milena, que se sacrificaram para me dar tempo e condições de me dedicar a este trabalho, abrindo mão da minha presença muitas vezes, sempre me dando apoio, compreensão e amor, incondicionalmente. Obrigado por sempre acreditarem em mim, vocês são a inspiração, a razão e o meio para o meu sucesso, são os amores da minha vida e o principal motivo para a minha resiliência, mesmo nos momentos mais difíceis. Eu amo vocês acima de qualquer coisa no mundo, e sem vocês nada disso teria feito sentido.

E meu último agradecimento é dedicado a meu orientador, Professor André Raabe, um mestre dedicado, humilde e sábio, que mesmo e principalmente nos momentos mais difíceis sempre me recebeu com um sorriso, com empatia e gentileza, transcendendo seu papel. Conheci poucas pessoas tão dedicadas à Educação na sua forma mais plena, valorizando o esforço de seus estudantes, convertendo erros em oportunidades, se preocupando muito mais com o aprendizado, com a experiência e a evolução profissional e pessoal de seus estudantes do que com meros números que na maioria das vezes não representam o processo pedagógico e a construção do conhecimento. Sempre generoso e compreensivo, o Professor Raabe nunca me deixou desistir, nunca me fez sentir desmotivado, nunca duvidou que eu conseguiria, mesmo quando eu mesmo duvidava. Suas palavras e atitudes deixaram uma marca eterna que levarei comigo por toda a minha vida. Por isso e mais, meus profundos, sinceros e eternos agradecimentos ao professor mais dedicado e humano que já tive o prazer e o privilégio de conhecer. Muito obrigado.

VALIDADE DE CONTEÚDO DO *CT PUZZLE TEST*: INSTRUMENTO PARA AVALIAÇÃO DO PENSAMENTO COMPUTACIONAL DE JOVENS DE 14 A 17 ANOS

Leonardo Felipe de Ávila Calbusch

Abril / 2021

Orientador: Prof. Dr. André Luis Alice Raabe

Área de Concentração: Computação Aplicada

Linha de Pesquisa: Informática na Educação

Palavras-chave: Pensamento Computacional, Avaliação do Pensamento Computacional, Validação de Instrumentos de Avaliação.

Número de páginas: 100

RESUMO

Nos últimos anos, várias iniciativas surgiram para implementar o ensino e a avaliação do Pensamento Computacional (PC) nos currículos escolares de diferentes países. Para implementar o PC com sucesso no currículo escolar, é necessário projetar ferramentas e métodos de avaliação apropriados. A maioria das abordagens de avaliação do PC conhecidas baseia-se em tarefas de programação e na adoção de instrumentos que realizam uma análise estática dos projetos desenvolvidos pelos estudantes. Outros utilizam instrumentos na forma de questionários de múltipla escolha compostos por questões objetivas predefinidas que não consideram o processo cognitivo necessário para resolver cada questão. Nesse contexto, a adoção de estratégias de avaliação mais abrangentes pode trazer benefícios para a avaliação do PC nos alunos. O *Computational Thinking Puzzle Test* implementa uma abordagem baseada em *puzzles* que ajuda a avaliar o pensamento crítico e o raciocínio lógico dos estudantes. O teste independe de domínio específico e contempla experimentação e teste de hipóteses, captando as interações do aluno com o instrumento. Este teste tem demonstrado evidências de potencial em avaliar o PC e tem passado por fases de evolução que requerem um novo processo de validação para se estabelecer como um instrumento eficaz e confiável. Este trabalho apresenta uma etapa do processo de validação e aprimoramento do *CT Puzzle Test*. A arquitetura do *software* do instrumento foi aprimorada, os itens do instrumento foram revisados, os métodos de armazenamento e disponibilidade dos resultados foram ajustados. Foram realizadas a validação de conteúdo por meio da análise de juízes e a análise semântica com uma pequena amostra da população. O resultado deste trabalho demonstra evidências da validade do *CT Puzzle Test* como ferramenta de avaliação PC, que aborda o construto por meio de uma perspectiva pedagógica mais ampla do que outros instrumentos conhecidos. A sistematização da aplicação do teste, com a definição de diretrizes e documentação de apoio, permite que o teste seja disponibilizado a qualquer pesquisador, educador ou avaliador clínico interessado em investigar o desenvolvimento do PC em indivíduos de sua área de atuação.

***CT PUZZLE TEST* CONTENT VALIDITY: COMPUTATIONAL THINKING ASSESSMENT TOOL FOR YOUNG PEOPLE AGED 14 TO 17 YEARS**

Leonardo Felipe de Ávila Calbusch

April / 2021

Advisor: André Luis Alice Raabe, Doctor

Area of Concentration: Applied Computer Science

Research Line: Informatics in Education

Keywords: Computational Thinking, Computational Thinking Assessment, Validation of Assessment Instruments.

Number of pages: 100

ABSTRACT

In recent years, several initiatives have emerged to implement the teaching and assessment of Computational Thinking (CT) in school curricula in different countries. To successfully implement CT in the school curriculum, it is necessary to design appropriate assessment tools and methods. Most of the known CT assessment approaches are based on programming tasks and the adoption of instruments that perform a static analysis of the projects developed by students. Others use instruments in the form of multiple-choice questionnaires composed of predefined objective questions that do not consider the cognitive process necessary to solve each question. In this context, the adoption of more comprehensive assessment strategies can bring benefits for the assessment of CT in students. The Computational Thinking Puzzle Test implements a puzzle-based approach that helps to assess students' critical thinking and logical reasoning. The test does not depend on a specific domain, and includes experimentation and hypothesis testing, capturing the student's interactions with the instrument. This test has shown evidence of potential in evaluating CT and has gone through phases of evolution that require a new validation process to establish itself as an effective and reliable instrument. This work presents a step in the validation and improvement process of the CT Puzzle Test. The software architecture of the instrument was improved, the items instrument revised, and the storage methods and availability of the results adjusted. Content validation was performed through the analysis of judges and semantic analysis with a small sample of the population. The result of this work demonstrates evidence of the validity of the CT Puzzle Test as a CT assessment tool, which addresses the construct through a broader pedagogical perspective than that of other known instruments. The systematization of the test application, with the definition of guidelines and supporting documentation, allows the test to be made available to any researcher, educator or clinical evaluator interested in investigating the development of CT in individuals in their area of expertise.

LISTA DE ILUSTRAÇÕES

Figura 1. <i>CTt</i> , item 11: <i>loops</i> “repetir até + repetir vezes”; “setas visuais”.....	38
Figura 2. <i>CTt</i> , item 18: <i>loops</i> “repetir até” + condicional <i>if/else</i> ; “blocos visuais”.....	38
Figura 3. <i>CTt</i> , item 18: <i>loops</i> “repetir até” + condicional <i>if/else</i> ; “blocos visuais”.....	39
Figura 4. <i>Dr. Scratch</i> : relatório de avaliação e <i>feedback</i> para o projeto “ <i>Alice in Wonderland</i> ”.....	41
Figura 5. <i>Bebras Tasks</i> item 1: abastecimento de água (estrutura lógica/binária).....	42
Figura 6. <i>Bebras Tasks</i> item 2: lavanderia rápida (paralelismo / algoritmos).....	43
Figura 7. <i>Bebras Tasks</i> item 3: ábaco (abstração, decomposição, algoritmos).....	43
Figura 8. Fase 1 do <i>CT Puzzle Test</i>	47
Figura 9. Fase 2 do <i>CT Puzzle Test</i>	48
Figura 10. Fase 3 do <i>CT Puzzle Test</i>	49
Figura 11. Fase 4 do <i>CT Puzzle Test</i>	50
Figura 12. Fase 5 do <i>CT Puzzle Test</i>	51
Figura 13. Fase 6 do <i>CT Puzzle Test</i>	52
Figura 14. Fase 7 do <i>CT Puzzle Test</i>	53
Figura 15. Modelo de arquitetura do <i>CT Puzzle Test</i> versão <i>Flash</i>	64
Figura 16. Modelo de arquitetura do <i>CT Puzzle Test</i> versão <i>HTML5</i>	65
Figura 17. Novo modelo de arquitetura do <i>CT Puzzle Test</i>	66
Figura 18. Dependência entre os módulos do <i>CT Puzzle Test</i>	67
Figura 19. Caso de Uso “Cadastro de Aplicador”.....	68
Figura 20. Caso de Uso “Aplicar Teste”.....	69
Figura 21. Caso de Uso “Cadastro Estudante”.....	69
Figura 22. Caso de Uso “Realizar Teste”.....	70
Figura 23. Modelo conceitual simplificado.....	70
Figura 24. <i>Diagrama Entidade Relacionamento (DER)</i>	71
Figura 25. Material de apoio ao aplicador – instruções de utilização do teste.....	75
Figura 26. Material de apoio ao estudante - passos para iniciar o teste.....	76
Figura 27. Nível de familiaridade dos especialistas com as definições do PC.....	78
Figura 28. Nível de conhecimento dos especialistas sobre ferramentas de avaliação do PC.....	79
Figura 29. Adequação da quantidade de itens de cada fase.....	85
Figura 30. Nível de adequação da quantidade total de fases do <i>CT Puzzle Test</i>	89
Figura 31. Nível de adequação da quantidade total de itens do <i>CT Puzzle Test</i>	90
Figura 32. Nível de adequação do tempo previsto para conclusão do <i>CT Puzzle Test</i>	90

LISTA DE QUADROS

Quadro 1. Redução de escopo das perguntas de pesquisa	23
Quadro 2. Redução de escopo dos objetivos de pesquisa	24
Quadro 1. Publicações que propõem uma definição para o PC	30
Quadro 2. Conjuntos de palavras mais influentes nas definições do PC	30
Quadro 3. Levantamento de Instrumentos de Avaliação do PC	34
Quadro 4. Classificação do <i>CT Puzzle Test</i> por abordagens de avaliação	36
Quadro 5. Perfis profissionais dos especialistas	77
Quadro 6. Conversão da escala aplicada para o IVC	82

LISTA DE TABELAS

Tabela 1. Distribuição dos pesos das fases do <i>CT Puzzle Test</i> em relação aos pilares do PC.....	54
Tabela 2. Dificuldade dos itens do <i>CT Puzzle Test</i>	83
Tabela 3. Relevância dos itens do <i>CT Puzzle Test</i> na avaliação do PC.....	84
Tabela 4. Adequação das instruções (tutoriais) de cada fase.....	85
Tabela 5. Nível de aderência da fase 1 para cada pilar do PC.....	86
Tabela 6. Nível de aderência da fase 2 para cada pilar do PC.....	86
Tabela 7. Nível de aderência da fase 3 para cada pilar do PC.....	87
Tabela 8. Nível de aderência da fase 4 para cada pilar do PC.....	87
Tabela 9. Nível de aderência da fase 5 para cada pilar do PC.....	87
Tabela 10. Nível de aderência da fase 6 para cada pilar do PC.....	88
Tabela 11. Nível de aderência da fase 7 para cada pilar do PC.....	88
Tabela 12. Peso dos pilares sobre os escores <i>versus</i> IVC da aderência aos pilares.....	88
Tabela 13. Capacidade do <i>CT Puzzle Test</i> em avaliar o PC.....	91
Tabela 14. Adequação da abordagem por meio de <i>puzzles</i>	91
Tabela 15. Adequação da interface gráfica e da linguagem adotada no teste.....	91

LISTA DE ABREVIATURAS E SIGLAS

CAU	Colégio de Aplicação da UNIVALI
CFA	Confirmatory factor analysis
CSTA	Computer Science Teachers Association
CTt	Computational Thinking Test
EFA	Exploratory factor analysis
HTML	Hypertext Markup Language
ISTE	International Society for Technology in Education
IVC	Índice de validade de conteúdo
K12	Kindergarten - 12th grade
MIT	Massachussets Institute of Technology
NRC	National Research Council
PC	Pensamento Computacional
UNIVALI	Universidade do Vale do Itajaí

SUMÁRIO

1 INTRODUÇÃO.....	13
1.1 PROBLEMA DE PESQUISA.....	18
1.1.1 Solução Proposta	19
1.1.2 Delimitação de Escopo	20
1.1.3 Justificativa.....	20
1.2 OBJETIVOS	21
1.2.1 Objetivo Geral.....	21
1.2.2 Objetivos Específicos	21
1.3 METODOLOGIA.....	21
1.3.1 Metodologia da Pesquisa	21
1.3.2 Procedimentos Metodológicos.....	22
1.4 ESTRUTURA DA DISSERTAÇÃO.....	24
2 FUNDAMENTAÇÃO TEÓRICA.....	26
2.1 PENSAMENTO COMPUTACIONAL	26
2.2 AVALIAÇÃO DO PENSAMENTO COMPUTACIONAL	32
2.2.1 <i>Computational Thinking Test</i>	37
2.2.2 <i>Dr. Scratch</i>	40
2.2.3 <i>Bebras Tasks</i>	41
2.3 COMPUTATIONAL THINKING PUZZLE TEST	44
2.3.1 Primeira Fase.....	46
2.3.2 Segunda Fase	47
2.3.3 Terceira fase	48
2.3.4 Quarta fase.....	49
2.3.5 Quinta fase	50
2.3.6 Sexta fase.....	51
2.3.7 Sétima fase	52
2.3.8 Pesos dos Pilares do PC	54
2.4 VALIDAÇÃO DE INSTRUMENTOS DE AVALIAÇÃO	54
2.4.1 CONFIABILIDADE	55
2.4.1.1 Estabilidade.....	55
2.4.1.2 Consistência Interna	55
2.4.1.3 Equivalência.....	56
2.4.2 VALIDADE	56
2.4.2.1 Validade de Conteúdo.....	57
2.4.2.2 Validade de Critério.....	58
2.4.2.3 Validade de Construto	59
3 VALIDAÇÃO DO CT PUZZLE TEST	62
3.1 REVISÃO DO INSTRUMENTO.....	62

3.1.1	Revisão Técnica	63
3.1.1.1	Evolução da Arquitetura do Sistema	63
3.1.1.2	Diagramas de Casos de Uso.....	68
3.1.1.3	Modelo conceitual simplificado.....	70
3.1.1.4	Banco de Dados.....	71
3.1.1.5	Correções da última versão	72
3.1.2	Material de apoio.....	73
3.1.2.1	Material de apoio ao aplicador	74
3.1.2.2	Material de apoio ao estudante	75
3.2	VALIDAÇÃO DE CONTEÚDO	76
3.2.1	Avaliação por juízes	77
3.2.1.1	Amostra	77
3.2.1.2	Instrumentos	79
3.2.1.3	Metodologia.....	81
3.2.1.4	Resultados da avaliação das fases.....	81
3.2.1.4.1	Dificuldade dos itens.....	82
3.2.1.4.2	Relevância dos itens para avaliar o PC	84
3.2.1.4.3	Adequação da quantidade de itens de cada fase.....	84
3.2.1.4.4	Adequação das instruções ao estudante	85
3.2.1.4.5	Nível de aderência de cada pilar do PC.....	86
3.2.1.5	Resultados da avaliação geral do instrumento	89
3.2.2	Análise Semântica	91
3.2.2.1	Amostra	92
3.2.2.2	Instrumentos	92
3.2.2.3	Procedimentos.....	93
3.2.2.4	Resultados	93
4	CONCLUSÕES	95
4.1	CONTRIBUIÇÕES DA DISSERTAÇÃO	96
4.1.1	Publicações em Eventos	97
4.2	TRABALHOS FUTUROS	97
	REFERÊNCIAS	98

1 INTRODUÇÃO

O pensamento computacional (PC) vem sendo estudado nos últimos anos sob diferentes perspectivas, e várias iniciativas têm surgido no intuito de implementar o ensino e a avaliação do PC em currículos escolares de diversos países. No entanto, a comunidade científica ainda não adotou uma definição formal congruente para o PC.

Nos últimos anos, o PC vem sendo apresentado como “um conjunto de habilidades para solução de problemas que deve ser desenvolvido pelas novas gerações de estudantes” (BOCCONI et al., 2016), uma percepção já defendida na década de 1970 por Papert (1971). Em 2006, Jannet Wing introduziu o termo e afirmou que o PC “envolve a solução de problemas, o design de sistemas e a compreensão do comportamento humano, baseando-se nos conceitos fundamentais da ciência da computação” (WING, 2006). Desde então ainda não se estabeleceu um consenso entre os pesquisadores da área sobre a definição do PC (KALELIOGLU et al., 2016), tampouco sobre como ele deve ser integrado aos sistemas educacionais, ensinado e avaliado. Ainda assim, o PC está ganhando espaço nas escolas em todo o mundo.

Nos Estados Unidos, a *Computer Science Teachers Association (CSTA)* em parceria com a *International Society for Technology in Education (ISTE)* trabalham de forma colaborativa para desenvolver materiais que possam auxiliar professores do ensino do PC. Apresentaram, em 2011, uma definição operacional do PC envolvendo os seguintes conceitos computacionais: coleta, análise e representação de dados; decomposição de problemas; abstração; algoritmos e procedimentos; automação; simulação e paralelismo (ISTE; CSTA, 2011). Os conceitos apresentados representam características que um indivíduo deve demonstrar no processo de resolução de problemas.

Brackmann (2017) apresenta uma pesquisa que aponta 4 conceitos fundamentais considerados como pilares para o PC: decomposição; reconhecimento de padrão; abstração e algoritmos. Brackmann afirma que desenvolver os pilares do PC durante a educação básica pode auxiliar na atitude dos indivíduos ao se deparar com um problema.

Para que as tentativas de implantação do PC nas escolas obtenham êxito, é de suma importância projetar artefatos e métodos de avaliação adequados para avaliar os estudantes (KONG et al., 2019). A validação de instrumentos de avaliação do PC é extremamente relevante e urgente

porque, sem ferramentas de avaliação confiáveis e válidas, o PC pode perder seu potencial de ingressar nos currículos educacionais (ROMÁN-GONZÁLEZ et al., 2017).

As abordagens de desenvolvimento dos pilares do PC em estudantes de escolas básicas encontradas na literatura adotam o ensino de programação, seja por meio de disciplinas específicas do currículo escolar, seja por atividades extracurriculares, *workshops* ou minicursos. Para avaliar os pilares do PC desenvolvidos por essas abordagens, se utilizam instrumentos que realizam uma avaliação fechada por meio da análise estática de projetos desenvolvidos pelos estudantes, como o *Dr. Scratch* (MORENO-LEÓN et al., 2015), *Progression of Early Computational Thinking (PECT)* (SEITER; FOREMAN, 2013), *CodeMaster* (WANGENHEIM et al., 2018) e *Quizly* (MAIORANA et al., 2015), entre outros. Esse tipo de avaliação apresenta algumas limitações. Para desenvolver seus projetos, os estudantes precisam de conhecimento prévio sobre o ambiente ou linguagem utilizada, com todas as suas particularidades técnicas. Além disso, esses instrumentos geralmente funcionam no código-fonte dos projetos, normalmente após o estudante terminar seu trabalho.

Outras iniciativas utilizam instrumentos na forma de questionários de múltipla escolha, como o *Commutative Assessment Test* (WEINTROP; WILENSKY, 2015), o *Test for Measuring Basic Programming Abilities* (MÜHLING et al., 2015), o *Israel's Nationwide Exam* (ZUR-BARGURY et al., 2013) e o *Bebras Tasks* (DAGIENE; FUTSCHEK, 2008), entre outros. Esse tipo de teste avalia somente as respostas finais dos estudantes a questões previamente definidas, sem valorizar o processo cognitivo necessário para a resolução de cada questão.

Neste contexto, a adoção de estratégias de avaliação mais amplas pode trazer benefícios para a avaliação do PC em estudantes. Estudos envolvendo métodos de ensino que diferem do formato padrão de palestra / memorização / teste mostraram-se promissores em aumentar o nível de desempenho intelectual autêntico demonstrado pelos alunos (BOALER, 1998; TANNER; ALLEN, 2004). Falkner et al. (2010) apresenta efeitos mais amplos da pedagogia nas habilidades de pensamento crítico de um aluno, estabelecendo definições para métodos de ensino baseados em projetos, problemas e *puzzles*. O método menos abstrato, baseado em projetos, inclui trabalhar em equipes e lidar com incertezas e mudanças de condições. Construído sobre este método está o aprendizado baseado em problemas. Este método envolve a aquisição de conhecimento e raciocínio de domínios específicos com métodos específicos de domínio. O método mais abstrato, o aprendizado de *puzzles*, baseia-se em ambos para desenvolver o pensamento crítico e o raciocínio lógico,

independentemente de um domínio específico (FALKNER et al., 2010; MERRICK, 2010; MICHALEWICZ, 2008).

A abordagem de aprendizado baseada em *puzzles*, segundo Falkner et al. (2010), visa incentivar os estudantes a “pensar em como eles estruturam e resolvem problemas não encontrados no final de algum capítulo de livro”. O objetivo é motivar os alunos e, ao mesmo tempo, aumentar sua consciência matemática e habilidades de resolução de problemas, induzindo-o a discutir estratégias de solução. De acordo com Falkner et al. (2010), essa abordagem deve satisfazer alguns critérios como:

- Independência de contexto: *puzzles* educacionais não devem estar especificamente vinculados a um único domínio de solução de problemas;
- Generalidade: *puzzles* educacionais desenvolvem princípios universais de solução de problemas, ou seja, treinar a habilidade de resolução de problemas deve permitir ao estudante resolver problemas ainda desconhecidos no futuro;
- Simplicidade: *puzzles* educacionais devem ser fáceis de declarar e lembrar, pois isso aumenta a chance de os alunos se lembrarem do método de solução para aplicar em outros problemas no futuro;
- Fator “Eureka”: *puzzles* educacionais devem inicialmente frustrar o solucionador de problemas, mas manter a promessa de resolução. Os solucionadores de problemas geralmente usam a intuição para iniciar sua busca pela solução, e eventualmente, alcançam um “momento Eureka” – quando reconhecem o caminho correto para resolver o *puzzle* e experimentam uma sensação de recompensa por sua esperteza.
- Fator de entretenimento: *puzzles* educacionais devem ser divertidos e envolventes. O entretenimento costuma ser resultado da combinação da simplicidade, frustração, fator Eureka e um cenário interessante, com algum contexto lúdico.

A abordagem por *puzzles* desenvolve habilidades de pensamento crítico e resolução de problemas, e estabelece uma conexão entre a capacidade de resolver *puzzles* e a capacidade de resolver problemas do mundo real. Muitos problemas do mundo real podem ser percebidos como *puzzles* em grande escala, pois ao se deparar com um problema no trabalho ou no cotidiano, o

indivíduo deve elaborar hipóteses e testar estratégias para resolvê-lo. Nessa perspectiva, ao analisar um questionário objetivo, os estudantes deveriam poder testar respostas com base em hipóteses experimentadas, ao invés de ter apenas uma chance de acerto que pune uma tentativa sem sucesso. Nesse contexto, torna-se relevante a proposição de um método eficiente de avaliação do PC baseado em *puzzles*.

Um dos projetos pioneiros de implantação do PC no currículo escolar no Brasil foi introduzido pela Universidade do Vale do Itajaí (UNIVALI). A partir de 2017, o Colégio de Aplicação da UNIVALI (CAU) implantou uma disciplina obrigatória de PC a todos os estudantes desde o 6º ano do Ensino Fundamental até o 3º ano do Ensino Médio. O grupo de pesquisadores da Universidade vem pesquisando o tema nos últimos anos, e Gonçalves (2015) desenvolveu o *CT Puzzle Test*, um instrumento de avaliação do PC, resultante de sua dissertação de mestrado. O *CT Puzzle Test* é composto por itens no formato de *puzzles*, que abordam atividades de resolução de problemas, planejadas para que não seja necessário conhecimento prévio. Os itens do *CT Puzzle Test* são fundamentados nos conceitos computacionais apresentados pela *ISTE* e *CSTA* (2011).

Para estimar a validade do instrumento, Gonçalves conduziu aplicações do teste na própria universidade. Os esforços de Gonçalves na investigação da validade e fidedignidade do *CT Puzzle Test* indicaram evidências fortes de validade do instrumento, mas o próprio autor relata que, “porém, para fins estatísticos, a aplicação com mais indivíduos se faz necessária” (Gonçalves, 2015).

Entre 2017 e 2018, durante a implantação da disciplina obrigatória de PC, o *CT Puzzle Test* foi utilizado por Couto (2018), outra pesquisadora da Universidade, para diagnosticar o desenvolvimento dos pilares do PC nos estudantes do CAU. Durante a análise dos resultados da aplicação do teste, Couto observou que alguns ajustes eram necessários para seus objetivos de pesquisa. A pesquisadora revisou e ajustou os itens do instrumento, e realizou alterações importantes visando diminuir o tempo de resolução necessário e viabilizar sua aplicação em um período regular de aula, como por exemplo a remoção de algumas fases do teste. Couto ainda modificou as fórmulas de cálculo dos escores das fases.

Com alterações tão substanciais, Couto decidiu realizar a validação convergente do instrumento modificado, adotando como critério o *CTt* (ROMÁN-GONZÁLEZ, 2015), um instrumento que já passou por testes para confirmar sua validade e fidedignidade (ROMÁN-GONZÁLEZ, 2015; ROMÁN-GONZÁLEZ et al., 2017b). Para tal, Couto adotou a definição do PC

fundamentada nos quatro pilares apontada por Brackmann (2017), e os dados coletados pelos dois testes foram analisados sob essa mesma perspectiva. Os resultados da investigação sobre a convergência entre os testes *CT Puzzle Test* e *CTt* demonstraram que são fortes as evidências da eficácia do *CT Puzzle Test* em avaliar o PC e corroboraram que a abordagem por *puzzles* e a contemplação da experimentação e teste de hipóteses por meio da captura das interações do estudante com o instrumento é uma abordagem válida para a avaliação do PC.

Para continuar contribuindo com a evolução do instrumento, novas pesquisas são necessárias a partir das modificações que o instrumento sofreu em suas últimas etapas de aprimoramento, e novas etapas de validação apoiadas em novas amostras são a principal contribuição ao estado atual de evolução do instrumento.

Nesse contexto, este trabalho apresenta um processo de aprimoramento e validação do *CT Puzzle Test* baseado nas pesquisas de Gonçalves (2015) e Couto (2018). A validade é considerada “o aspecto mais importante para a avaliação da efetividade de um instrumento de avaliação, uma vez que trata a respeito da qualidade dos escores, mensurando o que de fato se propõe a mensurar” (HOGAN, 2006). Portanto, é fundamental para a continuidade do processo de evolução do *CT Puzzle Test* que este seja submetido a técnicas de investigação de validade de conteúdo, de critério e de construto. Neste trabalho, o teste é submetido à etapa de validação de conteúdo por meio de uma análise por juízes e de uma análise semântica a partir da aplicação com estudantes.

A análise por juízes avalia a pertinência dos itens do *CT Puzzle Test* em mensurar o PC nos estudantes e a validade da abordagem por *puzzles* em relação às abordagens dos demais instrumentos conhecidos. A análise semântica avalia se o teste apresenta boa acessibilidade, legibilidade e navegabilidade por parte dos estudantes. Uma definição do PC é proposta para a o processo de aprimoramento do teste, baseada nos quatro pilares do PC.

A arquitetura do sistema recebe aprimoramentos. O código-fonte do instrumento é revisado, algumas fases são replanejadas e um *website* é desenvolvido para a gestão das aplicações do teste e o armazenamento dos resultados em banco de dados.

Uma documentação de apoio a aplicadores e estudantes é apresentada. A documentação de apoio a aplicadores contempla instruções de acesso e utilização do teste, condições de aplicação, gerência dos resultados, entre outros aspectos relevantes à utilização prática do teste. A documentação

de apoio a estudantes oferece conteúdo instrucional introdutório sobre o teste para orientar o estudante antes da aplicação. Essa documentação permite a disponibilização do teste para uso público e viabiliza a adoção do teste a pesquisadores, educadores, profissionais da saúde e demais interessados em investigar o desenvolvimento do PC em indivíduos na sua área de atuação.

1.1 PROBLEMA DE PESQUISA

O *CT Puzzle Test* foi elaborado por Gonçalves (2015) e aprimorado por Couto (2018) com o objetivo de contribuir para a evolução da rede nomológica¹ de desenvolvimento do PC como um instrumento de avaliação eficiente que adota a abordagem por *puzzles*, uma abordagem mais ampla sobre o aprendizado dos estudantes do que os instrumentos de avaliação do PC conhecidos. No entanto, o instrumento passou por modificações substanciais durante seu processo de evolução, e tais modificações demandam uma revisão do seu processo de validação para que o instrumento se confirme como confiável e válido para a avaliação do PC.

A confiabilidade refere-se, principalmente, à estabilidade, consistência interna e equivalência de uma medida (MARTINS, 2006). Estimativas de confiabilidade são afetadas por diversos aspectos do ambiente de avaliação (avaliadores, características da amostra, tipo de instrumento, método de administração) e pelo método estatístico utilizado. Portanto, os resultados de uma pesquisa utilizando instrumentos de medida só podem ser interpretados quando as condições de avaliação e a abordagem estatística são apresentadas de maneira clara (KOTTNER et al., 2011).

A validade refere-se ao fato de um instrumento medir exatamente o que se propõe a medir (ROBERTS et al., 2006; PASQUALI, 2017). Hogan (2006) comenta que foram distinguidos três aspectos de validade: validade de conteúdo, validade de critério e a validade de construto. A validade de conteúdo refere-se ao grau em que o conteúdo de um instrumento reflete adequadamente o construto que está sendo medido, e para ser avaliada, “é necessário ter um domínio de conteúdo bem definido, sendo assim feita por especialistas que julgam a relevância e a representatividade do conteúdo do teste” (HOGAN, 2006). A validade de critério trata do “grau de eficácia em que um teste prediz um desempenho específico de uma pessoa” (PASQUALI, 2009). A mensuração do coeficiente de validade é feita por meio do cálculo da correlação entre os escores do teste “e uma outra medida

¹ Uma rede nomológica pode ser entendida como um sistema de leis científicas que explicam até que ponto e por que os construtos teóricos estão relacionados entre si e com as medidas correspondentes. A rede também informa sobre as circunstâncias (ou seja, variáveis moderadoras) nas quais essas relações podem ou não ser observadas, e depende tanto de definições claras de construto quanto do desenvolvimento de excelentes medidas (PRECKEL; BRUNNER, 2017).

presente ou futura de desempenho das pessoas, que se chama critério” (HOGAN, 2006). A validade de construto é a “extensão em que um conjunto de variáveis realmente representa o construto a ser medido” (MARTINS, 2006; HAIR JUNIOR et al., 2009). A fim de estabelecer a validade de construto, geram-se “previsões com base na construção de hipóteses, e essas previsões são testadas para dar apoio à validade do instrumento” (HAIR JUNIOR et al., 2009).

Este trabalho busca responder duas questões:

QP 1: “A abordagem de avaliação do PC por meio de *puzzles* na construção dos itens de um instrumento de avaliação é válida e pertinente?”

QP 2: “O *CT Puzzle Test* possui validade de conteúdo, sendo seus itens relevantes e representativos para avaliar o PC?”

1.1.1 Solução Proposta

Para responder às questões de pesquisa, o *CT Puzzle Test* é submetido a uma etapa de validação de conteúdo, por meio da análise de juízes e de uma análise semântica.

Com a finalidade de responder a QP1 foram formuladas as seguintes hipóteses aos questionamentos de pesquisa:

H1: Juízes concordam que a abordagem por *puzzles* é válida e pertinente na avaliação do PC.

H2: Os estudantes testados preferem a abordagem por *puzzles* às abordagens tradicionais de teste de perguntas e respostas.

Com a finalidade de responder a QP2 foram formuladas as seguintes hipóteses aos questionamentos de pesquisa:

H3: Juízes concordam que é alto o nível da capacidade do *CT Puzzle Test* em avaliar o PC em jovens de 14 a 17 anos.

H4: Juízes concordam que é alto o nível de aderência dos itens do *CT Puzzle Test* aos quatro pilares do PC.

H5: Os estudantes conseguem acessar e navegar pelo *CT Puzzle Test*, ler e compreender as instruções e exemplos e registrar as respostas correspondentes.

Para validar as hipóteses foram convidados especialistas para julgar o teste, e foi realizada uma aplicação do teste com estudantes para coleta de dados para permitir a análise semântica.

1.1.2 Delimitação de Escopo

A pesquisa é limitada a coleta de dados de alunos 1º, 2º e 3º ano do Ensino Médio do Campus Brusque do Instituto Federal Catarinense. A faixa etária dos estudantes varia entre 14 e 17 anos de idade. O teste é destinado para esse público-alvo. “A idade é utilizada como critério para a validação de construto de um teste quando este mede traços que são intrinsecamente dependentes de mudanças no desenvolvimento cognitivo/afetivo dos indivíduos” (PASQUALI, 2017). A escolha por esta faixa se baseia no desenvolvimento cognitivo mais elevado dos estudantes, alvo dos objetivos de avaliação do teste.

1.1.3 Justificativa

De acordo com Román-González et al. (2019), diferentes métodos de avaliação e medição implicam em definir operacionalmente um construto, a fim de conceber uma ferramenta de avaliação consequentemente validada. Portanto, “avanços na avaliação podem contribuir decisivamente para consolidar o PC como um conceito sólido, digno de ser estudado e desenvolvido” x.

No entanto, para representar uma contribuição científica, um instrumento de avaliação psicométrica deve ser validado por meio de sua comparação com diferentes métricas a fim de compor uma matriz correlacionada para validar os métodos adotados. A validade ocupa uma posição central na teoria da medida, constituindo-se um parâmetro fundamental e indispensável. Atualmente, é definida como “a medida em que as evidências empíricas embasam as interpretações e os usos propostos para o teste” (PASQUALI, 2017). Apesar de demonstrar potencial para a avaliação do PC, o *CT Puzzle Test* deve ser submetido a um rígido processo de validação para ser considerado confiável e válido.

A partir de uma abordagem psicométrica, o PC ainda é um construto psicológico mal definido, pois “sua rede nomológica não foi completamente estabelecida; isto é, as correlações entre o PC e outras construções psicológicas ainda não foram completamente relatadas pela comunidade científica” (ROMÁN-GONZÁLEZ et al., 2016). Além disso, segundo Mühling et al., (2015), ainda existe uma grande lacuna de testes relacionados ao PC submetidos a um processo abrangente de validação psicométrica. Para que seja possível julgar a eficácia de qualquer currículo que incorpore

o PC, “medidas que permitam aos educadores avaliar o que a criança aprendeu precisam ser validadas. Desenvolver avaliações (padronizadas) da aprendizagem dos alunos é uma área urgente de necessidade para a comunidade relativamente jovem de educação em ciência da computação” (BUFFUM et al., 2015).

1.2 OBJETIVOS

A seguir são apresentados os objetivos gerais e específicos como solução à problemática da pesquisa.

1.2.1 Objetivo Geral

Evidenciar a validade de conteúdo do *CT Puzzle Test* para a avaliação do Pensamento Computacional em estudantes de 14 a 17 anos.

1.2.2 Objetivos Específicos

1. Aprimorar os aspectos técnicos do *CT Puzzle Test* para sua disponibilização via Internet;
2. Disponibilizar orientações de apoio ao aplicador do teste;
3. Apresentar conteúdo introdutório sobre o teste para o estudante.
4. Submeter o *CT Puzzle Test* a etapas de validação de conteúdo.

1.3 METODOLOGIA

Nesta seção é apresentada a metodologia utilizada na pesquisa e seus procedimentos metodológicos, os quais são utilizados para alcançar a solução proposta.

1.3.1 Metodologia da Pesquisa

Conforme Silva e Menezes (2001), a natureza da pesquisa é aplicada, pois se propôs a gerar conhecimentos que possam ser aplicados na prática, tencionando a solução de problemas específicos na temática envolvida. A forma de abordagem do problema é quantitativa, pois se propôs a traduzir em números opiniões e informações com a finalidade de classificá-las e analisá-las utilizando-se métodos estatísticos. Do ponto de vista de seus objetivos, a pesquisa é descritiva, pois teve como objetivo descrever as características de determinada amostra, ou fenômeno, ou o estabelecimento de relações entre variáveis previamente selecionadas. Foram usados questionários para a coleta dos

dados. O questionário “é uma técnica estruturada para coleta de dados, que consiste em uma série de perguntas que o entrevistado deve responder” (MALHOTRA, 2001).

1.3.2 Procedimentos Metodológicos

Durante a fase de fundamentação teórica, foi realizada uma pesquisa bibliográfica, em livros, artigos, teses e dissertações acadêmicas, com o objetivo de identificar as pesquisas até então produzidas e relacionadas ao tema abordado, além de identificar os instrumentos mais utilizados para mensurar o construto em questão.

Na fase seguinte, foram adotadas metodologias de validação do instrumento avaliado por meio de análises estatísticas para medir correlações entre avaliações de especialistas convidados a avaliar o instrumento.

Em seguida, iniciou-se a fase de coleta de dados cujo método para a realização da pesquisa foi o levantamento. A seleção desse procedimento técnico é distintiva “quando a pesquisa envolve interrogação direta com as pessoas cujo comportamento se deseja conhecer, e o pesquisador não interfere no ambiente, que está estudando” (SILVA; MENEZES, 2001).

Por fim, para analisar os dados obtidos, foram utilizadas técnicas multivariadas, que, segundo Hair Junior et al. (2009), “consistem no uso de duas ou mais variáveis que servem de indicadores de uma única medida composta”. A utilização da técnica teve como objetivo possibilitar um estudo mais abrangente dos resultados obtidos. Sua escolha remete ao fato de o trabalho ter uma natureza mais complexa de análise quantitativa. Segundo Malhotra (2001), as técnicas multivariadas são apropriadas para se analisarem dados quando há duas ou mais medidas de cada elemento e as variáveis são analisadas simultaneamente em dois ou mais fenômenos.

1.4 LIMITAÇÕES DA PESQUISA

As principais limitações desta pesquisa decorrem da pandemia da doença COVID-19, uma doença infecciosa causada por um novo tipo de coronavírus identificado pela primeira vez em dezembro de 2019, em Wuhan, na China. Durante os meses subsequentes, a doença se espalhou pelo mundo. O surto inicial deu origem a uma pandemia global que à data de conclusão desta pesquisa já tinha resultado em mais de três milhões de óbitos em todo o mundo. Governos ao redor do globo impuseram diversas restrições de mobilidade às suas populações, associadas a medidas de prevenção da transmissão do vírus que incluíram o fechamento de escolas, comércios e serviços públicos.

O período de levantamento de dados da atual pesquisa coincidiu, infelizmente, com os períodos de restrições para o combate à pandemia. Mesmo com as prorrogações previstas e permitidas pelos regulamentos pertinentes, não foi possível encontrar condições para as aplicações do teste no público-alvo necessárias para o cumprimento do planejamento inicial de executar-se todas as etapas de validação do instrumento. Sem vacina para a doença, governos e estados decretaram *lockdowns* e proibição de aglomerações, implantação de trabalho e estudo remotos e distanciamento social. Dessa forma, não foi possível reunir estudantes para uma aplicação sistemática do instrumento, inviabilizando assim a execução de diversas etapas de validação.

Diante deste contexto, as questões de pesquisa e suas respectivas hipóteses tiveram seu escopo significativamente reduzidos. O quadro 1 apresenta a comparação entre o planejamento inicial e da execução final da pesquisa nesse aspecto:

Quadro 1. Redução de escopo das perguntas de pesquisa

Planejamento inicial	Execução final
<p>QP1: O <i>CT Puzzle Test</i> é um teste confiável para a avaliação do PC, capaz de reproduzir um resultado de forma consistente?</p> <p>H1: A aplicação do <i>CT Puzzle Test</i> em dois momentos às mesmas amostras obtém resultados similares que comprovam a estabilidade do teste.</p> <p>H2: As correlações entre os itens que fazem parte do <i>CT Puzzle Test</i> comprovam a consistência interna de cada item do instrumento em medir o PC.</p> <p>H3: Ao submeter respondentes treinados ao <i>CT Puzzle Test</i>, há equivalência nos resultados obtidos com alto grau de similaridade quanto aos escores.</p>	<p>QP1: A abordagem de avaliação do PC por meio de <i>puzzles</i> na construção dos itens de um instrumento de avaliação é válida e pertinente?</p> <p>H1: Juízes concordam que a abordagem por <i>puzzles</i> é válida e pertinente na avaliação do PC.</p> <p>H2: Os estudantes testados preferem a abordagem por <i>puzzles</i> às abordagens tradicionais de teste de perguntas e respostas.</p>
<p>QP2: O <i>CT Puzzle Test</i> é um teste válido, capaz de medir exatamente aquilo que se propõe a medir de acordo a validade de conteúdo, de critério e de construto?</p> <p>H4: A avaliação do <i>CT Puzzle Test</i> por um comitê de especialistas resulta em um alto índice de validade de conteúdo a partir da concordância sobre os aspectos do instrumento e seus itens (validade de conteúdo).</p> <p>H5: Há correspondência entre a pontuação obtida por uma amostra submetida ao <i>CT Puzzle Test</i> e a pontuação obtida pela mesma amostra submetida ao <i>CTt</i> (validade de critério).</p> <p>H6: Existe correlação forte entre os grupos de itens do <i>CT Puzzle Test</i> e os grupos de itens do <i>CTt</i> que se propõem a avaliar os mesmos pilares do PC e, ao mesmo tempo, correlações fracas entre os grupos de itens dos dois testes que se propõem a avaliar pilares do PC diferentes (validade de construto).</p> <p>H7: Ao aplicar o <i>CT Puzzle Test</i> em amostras com características diferentes, os resultados entre grupo semelhantes são similares, e os resultados entre grupos diferentes são divergentes, comprovando que o instrumento é sensível a ponto de detectar essas diferenças.</p>	<p>QP2: O <i>CT Puzzle Test</i> possui validade de conteúdo, sendo seus itens relevantes e representativos para avaliar o PC?</p> <p>H3: Juízes concordam que é alto o nível da capacidade do <i>CT Puzzle Test</i> em avaliar o PC em jovens de 14 a 17 anos.</p> <p>H4: Juízes concordam que é alto o nível de aderência dos itens do <i>CT Puzzle Test</i> aos quatro pilares do PC.</p> <p>H5: Os estudantes conseguem acessar e navegar pelo <i>CT Puzzle Test</i>, ler e compreender as instruções e exemplos e registrar as respostas correspondentes.</p>

Fonte: Autor.

Da mesma forma, os objetivos da pesquisa também precisaram ser ajustados à redução do escopo da pesquisa. O quadro 2 apresenta estas diferenças:

Quadro 2. Redução de escopo dos objetivos de pesquisa

Planejamento inicial	Execução final
Objetivo geral: Validar o <i>CT Puzzle Test</i> como um instrumento confiável para o diagnóstico do Pensamento Computacional em estudantes de 14 a 17 anos.	Objetivo geral: Evidenciar a validade de conteúdo do <i>CT Puzzle Test</i> para a avaliação do Pensamento Computacional em estudantes de 14 a 17 anos.
Objetivos específicos:	Objetivos específicos:
1. Aprimorar o método de armazenamento e disponibilidade dos resultados do <i>CT Puzzle Test</i> ;	1. Aprimorar os aspectos técnicos do <i>CT Puzzle Test</i> para sua disponibilização via Internet;
2. Aperfeiçoar a sistematização de aplicação do teste (identificação das amostras, parâmetros de aplicação do teste, critérios de espaço e tempo entre outros);	2. Disponibilizar orientações de apoio ao aplicador do teste;
3. Avaliar confiabilidade e validade do <i>CT Puzzle Test</i> ;	3. Apresentar conteúdo introdutório sobre o teste para o estudante.
4. Comparar os resultados das aplicações do <i>CT Puzzle Test</i> com outros instrumentos existentes	4. Submeter o <i>CT Puzzle Test</i> a etapas de validação de conteúdo.

Fonte: Autor.

Evidencia-se, portanto, a redução de escopo imposta pela excepcional situação enfrentada durante o período de restrições em função da pandemia, que limitou as possibilidades de coleta de dados, fundamentais para as etapas necessárias ao cumprimento do planejamento inicial.

1.5 ESTRUTURA DA DISSERTAÇÃO

O trabalho está organizado em 4 capítulos correlacionados. O Capítulo 1 apresentou a introdução, com a contextualização do tema PC abordado em escolas pelo mundo e no Brasil. Em seguida, foram apresentados os problemas de pesquisa, a definição dos objetivos do trabalho, e suas limitações para esclarecer o escopo proposto.

O Capítulo 2 apresenta a fundamentação teórica, introduzindo a conceituação de termos importantes para o melhor entendimento do assunto abordado. São apresentados conceitos anteriores à publicação do artigo de Wing apresentando o termo PC. É apresentada a pesquisa bibliográfica realizada com o objetivo de investigar trabalhos similares, categorização de ferramentas utilizadas para avaliação, habilidades e percepções mensuradas por estas ferramentas. O Capítulo 2 também tem como objetivo evidenciar a contribuição do presente trabalho em relação aos trabalhos já existentes na literatura, e ainda descreve os tipos de validação de instrumentos de avaliação que embasam esta pesquisa.

O capítulo 3 apresenta o desenvolvimento da parte empírica da pesquisa, relata os procedimentos e critérios adotados, descreve as coletas dos dados por meio da aplicação do *CT Puzzle Test*, e finalmente apresenta a análise dos resultados.

No Capítulo 4, são apresentadas as conclusões do trabalho: a relação entre os objetivos específicos e os resultados alcançados. São também propostas diferentes trilhas que podem ser seguidas como continuação deste trabalho possibilitando trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são apresentados os fundamentos dos principais conceitos utilizados nesta pesquisa. São apresentados diferentes conceitos de PC abordados por diversos autores, ferramentas que podem ser utilizadas para desenvolver o PC em indivíduos e métodos de avaliação que podem ser utilizados para mensurar o quanto o PC está desenvolvido em um indivíduo. Uma definição própria do PC é sugerida. Por fim, são apresentadas técnicas de investigação da validade de conteúdo de instrumentos de avaliação.

Antes do termo PC surgir na literatura, outros autores abordaram o construto com diferentes nomenclaturas. Embora até hoje o PC ainda não tenha sido formalmente definido, ele já vem sendo evidenciado por alguns autores há décadas.

2.1 PENSAMENTO COMPUTACIONAL

A implantação de conceitos computacionais na educação já era defendida por Seymour Papert anos antes da adoção do termo “Pensamento Computacional” pela comunidade científica. No final da década de 60, Papert já advogava sobre a ideia de desenvolver tecnologias integráveis ao processo de ensino-aprendizagem que permitissem a exploração da criatividade e capacidade de resolução de problemas dos estudantes. De acordo com o autor, ao invés de meramente utilizar a tecnologia, os estudantes deveriam desenvolver a habilidade de criá-la.

Em suas primeiras iniciativas, Papert, Cynthia Solomon e Wallace Feurzeig criaram, em 1967, a linguagem de computação “Logo” com o objetivo de ensinar alunos do ensino básico a programar. Em 1969, o projeto Logo foi incorporado ao Laboratório de Inteligência Artificial do *MIT* (*Massachusetts Institute of Technology*). Papert, que havia trabalhado na Universidade de Genebra com o psicólogo suíço Jean Piaget, trouxe para o projeto a teoria construtivista, que enfatiza que as crianças constroem significado por meio da interação entre experiência e ideias. Dessa associação de concepções, surgiu seu conceito de construcionismo, que sustenta que o aprendizado acontece quando o aluno está “conscientemente engajado na construção de uma entidade pública” (PAPERT; HAREL, 1991). Papert advogou que a computação poderia ser uma poderosa ferramenta intelectual para todas as crianças, e que a tecnologia poderia se tornar “algo que as próprias crianças aprenderão a manipular, estender, aplicar a projetos, ganhando assim um domínio maior e mais articulado do

mundo, uma sensação do poder do conhecimento aplicado e uma imagem realista e autoconfiante de si mesmos como agentes intelectuais” (PAPERT, 1971).

A expressão completa dessa ideia e sua caracterização como “pensamento computacional” apareceram pela primeira vez no livro de Papert, *Mindstorms* (PAPERT, 1980), embora também se referisse à mesma ideia como “pensamento procedural”. O autor explica que o pensamento procedural “é uma poderosa ferramenta intelectual” e que “a assimilação cultural da presença do computador dará origem à alfabetização em informática”. No entanto, desde 1980, Papert já tentava esclarecer que existia diferença entre o que chamou de “pensamento procedural” e a habilidade de “saber programar”. Segundo o autor, não era adequado interpretar o pensamento procedural como “significando saber programar ou saber sobre os diversos usos feitos do computador. Mas o verdadeiro conhecimento em informática não é apenas saber como fazer uso de computadores e ideias computacionais. É saber quando é apropriado fazê-lo” (PAPERT, 1980). A preocupação com esse esclarecimento persistiu com o crescimento do movimento do PC e está presente até hoje.

Ainda que trabalhos como o de Papert inspirassem perspectivas promissoras na área educacional, houve pouca aceitação na educação básica à época. No entanto, a adoção de computadores no ensino básico começou a ser expandida com o aumento do impacto da tecnologia da informação na sociedade e da presença da informatização na vida cotidiana.

Em 2006, Janet Wing reintroduziu o termo “pensamento computacional”, resgatando a abordagem de Papert de que o PC não era apenas programação e que deveria ser uma habilidade fundamental para todos. Ela afirma que o PC é uma habilidade que todos os indivíduos devem ter, que utiliza diversas ferramentas mentais emprestadas do campo de Ciência da Computação, e que deve ser desenvolvida em todos os estágios da vida dos indivíduos, incluindo a educação básica. Sendo fundamental, ela pode ser aplicada a diversas outras disciplinas, ou seja, não apenas para os indivíduos interessados em seguir uma vida profissional na área de Computação (WING, 2006).

Com o crescimento do interesse pelo desenvolvimento do PC na escola, governos de diversos países começaram a modificar seu currículo nacional nos níveis de ensino fundamental e médio. Com o termo em evidência, algumas entidades começaram a trabalhar para tornar viável a inserção de disciplinas abordando o conteúdo no currículo de escolas básicas.

A *ISTE*, em colaboração com a *CSTA*, trabalharam com líderes da educação superior, indústrias e educação básica (*K-12*) para desenvolver uma definição operacional para o PC com a finalidade de incorporá-lo em currículos escolares: “Pensamento Computacional é um processo de resolução de problemas” – o que é semelhante ao conceito apresentado por Wing. Também são apresentadas características que um indivíduo deve ter (*ISTE*; *CSTA*, 2011):

- Formulação de problemas de forma que computadores e outras ferramentas possam ajudar a resolvê-los;
- Organização lógica e análise de dados;
- Representação de dados por meio de abstrações como modelos e simulações;
- Automatização de soluções utilizando o pensamento algorítmico;
- Identificação, análise e implementação de soluções visando a combinação mais eficiente e eficaz de etapas e recursos;
- Generalização e transferência de soluções para uma ampla gama de problemas.

A *ISTE* e a *CSTA* também descrevem disposições e atitudes que apoiam as características que o indivíduo deve apresentar para desenvolver o PC: (i) Confiança em lidar com complexidade; (ii) Persistência em trabalhar com problemas difíceis; (iii) Tolerância para ambiguidades; (iv) A habilidade de lidar com problemas abertos; (v) A habilidade de comunicar e trabalhar com outros para atingir um objetivo ou uma solução em comum. Além das disposições e atitudes, são mapeados conceitos computacionais: coleta, análise e representação de dados; decomposição de problemas; abstração; algoritmos e procedimentos; automação; simulação e paralelismo. Os conceitos apresentados representam características que um indivíduo deve demonstrar no processo de resolução de problemas. O guia “*Computational Thinking in K-12 teachers resources*” foi montado para ilustrar como os 9 conceitos apresentados podem ser aplicados de maneira progressiva em uma escola básica com as idades referentes ao *K-12* (educação primária e secundária).

No entanto, mesmo com o crescimento do interesse pelo PC, ainda falta uma definição formal sobre o tema (*GROVER*, 2015; *KALELIOGLU et al.*, 2016; *ROMÁN-GONZÁLEZ et al.*, 2017). O *National Research Council (NRC)*, um conselho associado às Academias Nacionais de Ciências,

Engenharia e Medicina dos Estados Unidos, divulgou vários relatórios ao longo dos últimos anos que apresentam esforços para explorar a natureza do PC e suas implicações cognitivas e educacionais, nos quais se evidencia necessária uma definição comum ao termo. O grande desafio é encontrar uma definição para o PC e mantê-la. Na visão de cientistas da computação consultados pelo NRC, o PC é uma habilidade analítica fundamental que todos podem utilizar para resolver problemas, modelar sistemas, entender o comportamento humano, e utilizar computadores para ajudar na automação de uma grande parte do processo intelectual, apoiando a sugestão de que o mesmo deve ser ensinado às crianças (NRC, 2011).

Segundo o NRC, uma abordagem comum para incorporar a computação ao currículo do ensino básico é enfatizar a alfabetização em computadores, que geralmente envolve o uso de ferramentas para criar documentos, páginas da Web ou apresentações multimídia. Uma segunda abordagem comum é enfatizar a programação de computadores, ensinando os alunos a programar em linguagens de programação específicas, como *Java* ou *C++*. Uma terceira abordagem concentra-se em ensino de programação aplicada, como jogos, robôs e simulações. Mas essas três abordagens principais - embora úteis e sem dúvida importantes – “não devem ser confundidas com aprender a pensar computacionalmente” (NRC, 2010).

A diversidade de abordagens teóricas para o PC e a resultante falta de padronização “são problemáticas do ponto de vista educacional” (MORENO-LEÓN et al., 2019). Isso é evidenciado pelo fato de que em muitos contextos educacionais “o PC e a programação (ou codificação) são usadas quase como sinônimos” (BALANSKAT e ENGELHARDT, 2014). De fato, a maioria das abordagens sobre ensino e avaliação do PC conduzidas pelo mundo se baseiam na avaliação de aspectos do pensamento computacional após os estudantes passarem por um treinamento que envolve tarefas de programação, como será detalhado na seção 2.2.

Na tentativa de encontrar aspectos congruentes nas definições do PC publicadas nos últimos anos e assim chegar a uma definição que represente sua essência, Moreno-León et al. (2019) realizaram um estudo que se baseou em uma revisão da literatura de trabalhos e propuseram uma definição própria do Pensamento Computacional. O quadro 1 apresenta os trabalhos investigados, iniciando com Papert em 1980, passando por Wing em 2006, *ISTE* e *CSTA* em 2011 e chegando a trabalhos mais recentes. O quadro 01 apresenta os trabalhos ordenados por data de publicação, e indica a forma de publicação (livro, periódico, revista, conferência ou relatório, entre outros).

Quadro 3. Publicações que propõem uma definição para o PC

Autor	Publicação
(Papert, 1980)	Livro – “ <i>Mindstorms</i> ”
(Wing, 2006)	Revista – “ <i>Communications of the ACM</i> ”
(Lu; Fletcher, 2009)	Boletim informativo – <i>SIGCSE Bulletin</i>
(Denning, 2009)	Revista – “ <i>Communications of the ACM</i> ”
(Wing, 2011)	Revista – “ <i>The Link</i> ”
(Aho, 2011)	Simpósio – “ <i>Ubiquity</i> ”
(Barr; Stephenson, 2011)	Periódico – “ <i>Inroads</i> ”
(ISTE; CSTA, 2011)	Relatório – “ <i>ISTE e CSTA</i> ”
(Furber, 2012)	Relatório – “ <i>Royal Society</i> ”
(Sengupta et al., 2013)	Periódico – “ <i>Education and information technologies</i> ”
(Grover; Pea, 2013)	Periódico – “ <i>Educational researcher</i> ”
(Mannila et al., 2014)	Relatório – “ <i>Working group ITiCSE</i> ”
(Csizmadia et al., 2015)	Relatório – “ <i>CAS</i> ”
(CSTA, 2016)	Relatório – “ <i>CSTA K-12 CS Standards</i> ”
(Tedre; Denning, 2016)	Anais de Evento – “ <i>Koli calling</i> ”
(Wolfram, 2016)	Coluna de opinião

Fonte: Adaptado de Moreno-León et al. (2019)

Para detectar os conceitos congruentes do PC nas definições revisadas, uma análise do tipo “nuvem de palavras” ou rede textual foi realizada em um documento contendo todas essas definições. O quadro 02 apresenta os resultados dessa análise.

Quadro 4. Conjuntos de palavras mais influentes nas definições do PC

Grupo	Palavras no contexto
1	Computador, ciência, ferramenta
2	Problema, resolução, solução
3	Abstração, simulação, decomposição
4	Sistema, informação, algoritmo
5	Lógica, depuração, performance

Fonte: Adaptado de Moreno-León et al. (2019)

Essas palavras são os nós que têm mais conexões dentro de cada grupo, sendo, conseqüentemente, as palavras mais influentes dos temas. Segundo análise dos autores, os resultados mostram que nem a programação nem a codificação emergem entre as palavras mais influentes das principais definições de PC. No entanto, embora “a programação torne os conceitos de PC concretos

e hoje em dia seja um método de fato para o aprendizado e o ensino dessas habilidades” (BOCCONI et al., 2016), essa situação pode mudar no futuro próximo devido a vários fatores. Os autores preveem que por um lado, educadores e pesquisadores podem encontrar outras estratégias para desenvolver habilidades de PC, como já é o caso com o uso de atividades desconectadas (BRACKMANN et al., 2017). Por outro lado, o intenso desenvolvimento de soluções de inteligência artificial, especialmente aquelas baseadas no aprendizado de máquina, pode “alterar drasticamente a maneira como a programação de computadores é executada” (GARCIA et al., 2019).

Os autores sugerem uma distinção entre PC - habilidade cognitiva humana - e habilidades de programação - competência instrumental. Considerando os elementos mais influentes e as comunidades de palavras destacadas pela análise da nuvem de palavras, os autores chegam a conclusão que indica uma definição congruente entre as diversas definições pesquisadas, sugerindo que o PC poderia ser compreendido como “a capacidade de formular e representar problemas para resolvê-los, utilizando ferramentas, conceitos e práticas da ciência da computação, como abstração, decomposição ou uso de simulações” (MORENO-LEÓN et al., 2019). Esta definição esclarece a relação entre os conceitos de PC e de programação, sendo o primeiro uma capacidade cognitiva do sujeito, e o segundo apenas um dos meios para desenvolvê-la.

Brackmann (2017) apresenta uma pesquisa que aponta a definição de 4 conceitos fundamentais considerados como pilares para o PC: (i) decomposição - quebrar um problema ou sistema complexo em partes; (ii) reconhecimento de padrão - procurar similaridades entre problemas; (iii) abstração - foco nas informações importantes, ignorando detalhes irrelevantes e (iv) algoritmos - desenvolver passos ou regras para solucionar um problema.

Brennan e Resnick (2012) avaliam que o PC se manifesta em dimensões que incluem a transferência de conceitos computacionais para outros contextos, as práticas do processo de aprender e ensinar (pensamento incremental e iterativo, testagem e depuração, reutilização e combinação de conhecimentos) e as perspectivas dos indivíduos sobre o mundo à sua volta e sobre si mesmos (autoexpressão, interatividade, colaboração e interrogação).

Blikstein (2008) argumenta que o PC está relacionado a “saber usar o computador como instrumento de aumento do poder cognitivo e operacional humano [...] para aumentar nossa produtividade, inventividade, e criatividade”. Estabelecer congruência entre essas abordagens é fundamental para a compreensão do PC e o planejamento de metodologias avaliativas eficientes.

Fundamentado na análise de todas essas discussões, este trabalho propõe a associação da abordagem de Brackmann com a visão de Blikstein, e sugere que “o Pensamento Computacional é uma extensão da capacidade humana de resolução de problemas, que se fundamenta nos conceitos e práticas relacionados à construção, uso e avaliação de algoritmos, no reconhecimento de padrões, na decomposição de problemas e na abstração”. Esta é a definição adotada para a formulação das perguntas aos juízes e para a revisão das mecânicas de jogo do *CT Puzzle Test*.

2.2 AVALIAÇÃO DO PENSAMENTO COMPUTACIONAL

O ensino do pensamento computacional na educação básica tem recebido atenção em vários países, e a disponibilidade de instrumentos de avaliação validados permite que o PC seja investigado em contextos educativos variados. Nas etapas escolares da infância, investigar o PC pode auxiliar no desenvolvimento cognitivo, raciocínio lógico e capacidade de resolução de problemas das crianças. Na educação de jovens, pode-se desenvolver sua capacidade de transferir habilidades de PC para diferentes tipos de problemas, contextos e situações da vida real. No meio acadêmico, investigar o PC pode contribuir com a definição de currículos, projetos pedagógicos e políticas educacionais. Ainda é possível investigar o desenvolvimento do PC em indivíduos em contextos clínicos, como por exemplo os relacionados a públicos com necessidades especiais.

Um levantamento de Wangenheim et al. (2018) aponta que a avaliação do PC em estudantes, na maioria dos casos, se dá por meio de tarefas de programação que avaliam assertividade e eficiência de código. Outras iniciativas utilizam instrumentos na forma de questionários de múltipla escolha com uma única resposta correta previamente definida, como o *Commutative Assessment Test* (WEINTROP; WILENSKY, 2015), o *Test for Measuring Basic Programming Abilities* (MÜHLING et al., 2015) e o *Bebras Tasks* (DAGIENE; FUTSCHEK, 2008).

Alguns instrumentos de avaliação são apresentados na forma de extensões das linguagens de programação, por meio da análise do código escrito pelo estudante. Alguns exemplos para a linguagem *Scratch* são o *Dr. Scratch* (MORENO-LEÓN et al., 2015) ou o *Ninja Code Village* (OTA et al., 2016). Essas ferramentas adotam análise de código estático e avaliam tipo e número de blocos usados no programa, quantificando conceitos e práticas de PC, como abstração, lógica, fluxo de controle, etc. Outras ferramentas avaliam projetos de programação em outras linguagens baseadas em blocos, como o *Snap! Autograder* (BALL; GARCIA, 2016) ou *App Inventor Quizly* (MAIORANA et al., 2015).

O foco da maioria dessas abordagens está na avaliação de problemas fechados, com uma solução correta previamente conhecida. Esse tipo de avaliação apresenta algumas limitações. Primeiro, porque as ferramentas geralmente funcionam no código-fonte, normalmente depois que o aluno termina seu trabalho. Esse foco no código-fonte também “limita a avaliação, não cobrindo práticas essenciais do PC, como criatividade e colaboração, e algumas vezes não fornece suporte imediato para o aluno acompanhar seu desempenho durante a resolução das atividades” (ROBLES et al., 2018). Segundo, porque o estudante precisa de treinamento no ambiente de programação adotado, fundamentação técnica e familiaridade com a linguagem utilizada, conhecimentos prévios que são necessários para o acesso à avaliação do PC que influenciam no desempenho do estudante que realiza os testes.

Román-González et al. (2019) fazem uma análise dos instrumentos de avaliação do PC conforme sua abordagem avaliativa, e os classificam em 7 modalidades:

- Instrumentos de diagnóstico: visam medir o nível de aptidão do PC do sujeito. Podem ser administrados em condição de pré-teste em indivíduos sem nenhuma experiência prévia em programação, e podem ser aplicadas também na condição pós-teste, após uma intervenção educacional, para verificar se as competências de PC foram desenvolvidas.
- Instrumentos sumativos: aplicados após o indivíduo ser exposto a conteúdo e treinamento para os pilares do PC. Seu principal uso é na condição pós-teste, e está vinculado ao ambiente de aprendizado utilizado (linguagem de programação ou conteúdo administrado).
- Instrumentos formativo-iterativos: visam fornecer *feedback* imediato ao aluno, geralmente de maneira automática, a fim de desenvolver e melhorar suas habilidades em PC. Essas ferramentas não avaliam os indivíduos, mas seus produtos de aprendizado, geralmente projetos de programação. Portanto, essas ferramentas são usadas principalmente durante o processo de aprendizado e são projetadas para um ambiente de programação específico.
- Instrumentos de mineração de dados: focados no processo de aprendizado, recuperam e registram a atividade do aluno durante a execução de uma tarefa. Fornecem dados e

análises de aprendizado a partir das quais os processos cognitivos do sujeito podem ser inferidos.

- Instrumentos de transferência de habilidades: Seu objetivo é avaliar até que ponto os alunos são capazes de transferir suas habilidades de PC para diferentes tipos de problemas, contextos e situações. Normalmente são avaliações textuais compostas de perguntas relacionadas a contextos do mundo real, e o sujeito deve utilizar pilares do PC afim de respondê-las corretamente.
- Escalas de percepções e atitudes do PC: visam avaliar as percepções (por exemplo, percepções de autoeficácia) e atitudes dos sujeitos, não apenas sobre o PC, mas também sobre questões relacionadas, como computadores, ciência da computação, programação de computadores ou mesmo alfabetização digital.
- Avaliação de vocabulário de PC: essas ferramentas pretendem medir vários elementos e dimensões do PC, quando expressas verbalmente pelos sujeitos. Essas expressões verbais foram denominadas como “linguagem do pensamento computacional” (GROVER, 2011).

O quadro 3 apresenta um levantamento de instrumentos de avaliação do PC, baseado nas pesquisas de Román-González et al. (2019), Wangenheim et al. (2018) e Couto (2018).

Quadro 5. Levantamento de Instrumentos de Avaliação do PC

Instrumento	Classificação	Modelo de Avaliação	Conhecimento prévio
<i>Computational Thinking Test</i> (Román-González, 2015; Román-González et al., 2017)	Diagnóstico	Questionário de Múltipla Escolha/ Perguntas compostas por desafios lógicos	Independente
<i>CT Puzzle Test</i> (Gonçalves, 2015; Couto, 2018)	Mista	Jogo composto por <i>puzzles</i> apoiados em pilares do PC, no estilo arrastar e soltar, graduação com base no registro de atividades	Independente
<i>Scrape</i> (Wolz et al., 2011)	Sumativo	Graduação de código fonte de projeto concluído	<i>Scratch</i>
<i>Fairy Assessment</i> (Werner et al., 2012)	Sumativo	Graduação da performance em tarefas desenvolvidas com a ferramenta Alice após treinamento, questionário pré e pós-teste	<i>Alice</i>
<i>Combined taxonomy tests</i> (Meerbaum-Salant et al., 2013)	Sumativo	Aplicação de questionário em pré, interim e pós-teste, após treinamento na ferramenta <i>Scratch</i>	<i>Scratch</i>
<i>Israel's Nationwide Exam</i> (Zur-Bargury et al., 2013)	Sumativo	Questionário de Múltipla Escolha/ Perguntas compostas por trechos de código fonte	<i>Scratch</i>
<i>Hairball</i> (Franklin et al., 2013) (Boe et al., 2013)	Sumativo	Graduação de código fonte de projeto concluído	<i>Scratch</i>
<i>Progression of Early Computational Thinking</i> (Seiter; Foreman, 2013)	Sumativo	Graduação de código fonte de projeto concluído	<i>Scratch</i>
<i>Commutative Assessment Test</i> (Weintrop; Wilensky, 2015)	Sumativo	Questionário de Múltipla Escolha/ Perguntas compostas por trechos de código fonte	<i>Snap! / Java</i>

<i>Test for Measuring Basic Programming Abilities</i> (Mühling et al., 2015)	Sumativo	Questionário de Múltipla Escolha/ Perguntas compostas por trechos de código fonte	Linguagem <i>Karel</i>
<i>Quizly</i> (Maiorana et al., 2015)	Sumativo	Plataforma de criação de perguntas ou tarefas no ambiente App Inventor com <i>feedback</i> para o estudante em tempo real e registro de atividades	<i>App Inventor</i>
<i>Autograder</i> (Ball; Garcia, 2016) (Ball, 2017)	Sumativo	Gradação de código fonte de projeto concluído	<i>Snap!</i>
<i>ITCH</i> (Johnson, 2016)	Sumativo	Gradação de código fonte de projeto concluído	<i>Scratch</i>
<i>Computational Thinking Patterns CTP-Graph</i> (Koh et al, 2010)	Formativo-iterativo	<i>Feedback</i> para o estudante por avaliação de código fonte concluído	<i>AgentSheets</i>
<i>REACT</i> (Koh et al., 2014)	Formativo-iterativo	<i>Feedback</i> para o estudante por avaliação de código fonte em tempo real	<i>AgentSheets</i>
<i>Dr. Scratch</i> (Moreno-León et al., 2015)	Formativo-iterativo	<i>Feedback</i> para o estudante por avaliação de registro de atividade	<i>Scratch</i>
<i>Ninja Code Village</i> (Ota et al., 2016)	Formativo-iterativo	<i>Feedback</i> para o estudante por avaliação de código fonte concluído	<i>Scratch</i>
<i>Code Master</i> (Wangenheim et al. 2018)	Formativo-iterativo	<i>Feedback</i> para o estudante por avaliação de código fonte concluído	<i>App Inventor / Snap!</i>
<i>Shuchi Grover</i> (Grover et al., 2017; Grover et al., 2016)	Mineração de dados	Estatística por avaliação de registro de atividade	<i>Fairy Assessment / Alice</i>
<i>Kodetu</i> (Eguiluz et al., 2017)	Mineração de dados	Estatística por avaliação de registro de atividade	<i>Kodetu / Blockly</i>
<i>Bebras Tasks</i> (Dagiene; Futschek, 2008)	Transferência de habilidades	Questionário de Múltipla Escolha/ Perguntas compostas por desafios lógicos	Independente
CTP-Quiz (Basawapatna et al., 2011)	Transferência de habilidades	Questionário com perguntas sobre a relação entre eventos reais e eventos programados em jogos pré-definido desenvolvidos pelos estudantes	<i>AgentSheets</i>
<i>Dancing Alice</i> (Daily et al., 2014)	Transferência de habilidades	Modelagem de movimentos de dança em personagem no ambiente <i>Alice</i>	<i>Alice</i>
<i>Computational Thinking Scales (CTS)</i> (Korkmaz et al., 2017)	Escalas de percepções e atitudes do PC	Escala de avaliação sobre a autopercepção do sujeito com relação a conceitos do PC	Independente
<i>Computer Programming Self-Efficacy Scale (CPSES)</i> (Kukul et al, 2017)	Escalas de percepções e atitudes do PC	Escala de avaliação sobre a autopercepção do sujeito com relação a conceitos de programação	<i>Scratch / Smallbasic</i>
<i>Robotics and Engineering Workshop</i> (Grover, 2011)	Avaliação de vocabulário	Entrevistas em pré e pós-teste após treinamento	Linguagem <i>Logo</i>
<i>CT Framework</i> (Brennan; Resnick, 2012)	Mista	Gradação de avaliação de código fonte de projeto concluído complementada com entrevistas e tarefas de interpretação de código fonte.	<i>Scratch</i>

Fonte: Adaptado de vários autores, Román-González et al. (2019), Wangenheim et al. (2018), Couto (2018).

Cada instrumento aborda a avaliação do PC sob um aspecto diferente. Brennan e Resnick (2012) afirmaram que avaliar as competências computacionais apenas olhando para os programas criados pelos alunos poderia ser claramente insuficiente, e por isso enfatizaram a necessidade de múltiplos meios de avaliação. Grover (2015) afirma que diferentes tipos de ferramentas complementares de avaliação devem ser sistematicamente combinados para alcançar uma compreensão total e abrangente do PC dos estudantes.

Percebe-se, portanto, uma carência de estratégias de avaliação mais amplas do PC. As abordagens baseadas em questionários e análise de código não capturam informações relevantes sobre como se dá o processo de resolução de problemas pelos estudantes tais como o número de tentativas, o tempo entre tentativas, o número de ações realizadas, número de erros cometidos e outras

informações que são particularmente importantes para avaliação do PC tendo em vista que existem práticas que são tipicamente associadas ao seu desenvolvimento.

Acredita-se que a abordagem de avaliação do PC por meio da resolução de *puzzles* possibilita considerar ao mesmo tempo os conceitos e as práticas relacionadas ao PC além de permitir a coleta de informações relevantes sobre o processo de resolução dos problemas pelos estudantes. Falkner et al. (2010), menciona que o uso de *puzzles* pode motivar e, ao mesmo tempo, aumentar a consciência matemática e habilidades de resolução de problemas dos alunos.

Neste contexto, o *CT Puzzle Test* se destaca pois combina algumas das abordagens de avaliação indicadas por Román-González et al. (2019) (diagnóstico, formativo-iterativo, sumativo, mineração de dados) e é independente de linguagem/ferramenta. O quadro 4 apresenta como cada abordagem se caracteriza no teste.

Quadro 6. Classificação do *CT Puzzle Test* por abordagens de avaliação

Classificação	Caracterização
Diagnóstico	Mede o nível de aptidão do PC por meio de escores e pode ser aplicado em formato pré-teste pois independe de conhecimentos prévios.
Formativo-iterativo	Fornecer feedback imediato e automático sobre acertos e erros; fornece dicas de resolução dos problemas.
Sumativo	Pode ser aplicado após o indivíduo passar por instrução (uma disciplina de PC, por exemplo) para medir seu desenvolvimento
Mineração de dados	Registra as interações do indivíduo com o instrumento e permite a análise do pesquisador sobre essas interações (como tempo, tentativas, utilização de dicas etc.), e não somente os acertos e erros.

Fonte: Autor.

Além disso, o instrumento adota o estilo de pergunta em formato de *puzzles* ao invés da escolha entre múltiplas respostas apresentadas em questionários tradicionais, e contempla a experimentação e teste de hipóteses por meio da captura das interações do estudante com o teste.

Em sua pesquisa, Couto (2018) realizou um levantamento sobre as ferramentas utilizadas para abordar o PC nas escolas em todo o mundo, e quais métodos de avaliação eram adotados. A linguagem de programação *Scratch* se revelou uma das ferramentas mais utilizadas para a abordagem do PC.

A maioria dos trabalhos investigados utilizou métodos de avaliação por meio de questionários, variando entre respostas de múltipla escolha, escala, respostas abertas, verdadeiro ou falso, entre

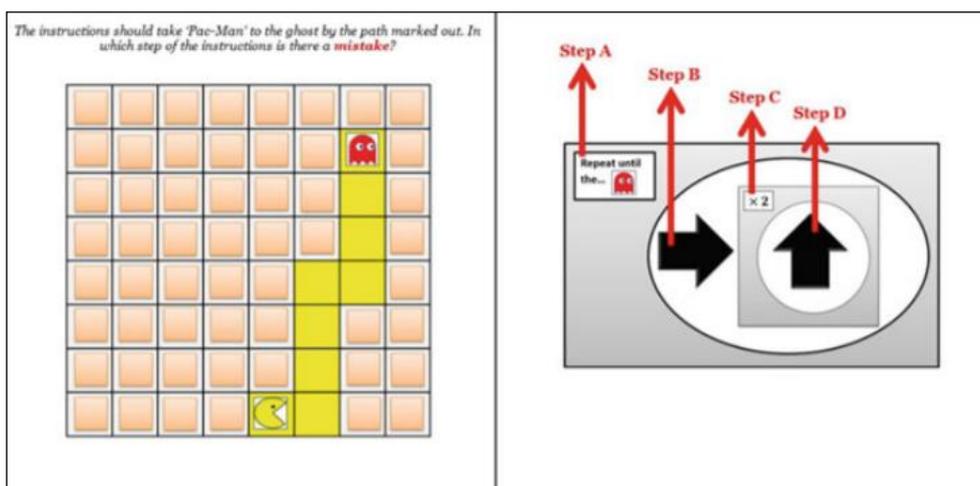
outros. Outros métodos de avaliação incluem entrevistas em grupo e individuais, observação, ou análise dos artefatos produzidos. Em geral, os instrumentos são apresentados pelos seus autores como um experimento, aplicados em contextos variados com o objetivo de se propor uma ferramenta de apoio ao processo de ensino-aprendizagem dos estudantes, normalmente adotada em atividades extracurriculares, *workshops* ou minicursos. O maior número dos instrumentos de avaliação observados não passou por processos formais para atestar sua confiabilidade e validade. Alguns, no entanto, são de interesse desta pesquisa em função do trabalho de validação realizado.

2.2.1 Computational Thinking Test

O *Computational Thinking Test* – *CTt* (ROMÁN-GONZÁLEZ, 2015) é uma ferramenta de avaliação diagnóstica que consiste em um instrumento de múltipla escolha composto por 28 itens, administrados on-line em um tempo máximo de 45 minutos. Cada item aborda um ou mais dos seguintes conceitos computacionais, que aparecem em dificuldade crescente ao longo do teste: instruções e sequências básicas; *loops* (“repetir vezes”, “repetir até”, “repetir enquanto”); condicionais simples (“se”), condicionais compostas (“se/senão”) e funções simples. Esses conceitos estão alinhados com os padrões de aprendizagem estabelecidos pela *CSTA* (2011) para o ensino da ciência da computação. Especificamente, em seu Nível 2 (correspondente às séries 6-9 no sistema norte-americano), o padrão definido com relação aos conceitos e habilidades computacionais que devem ser dominados nessas idades se caracteriza pela implementação de soluções de problemas usando uma linguagem de programação, incluindo estruturas de repetição, estruturas condicionais, expressões lógicas, variáveis e funções. Em sua pesquisa, Brackmann (2017) mapeou os itens do *CTt* para os quatro pilares do PC (abstração, reconhecimento de padrões, decomposição e algoritmos).

Em cada item, as respostas são representadas em um dos dois estilos: “setas visuais” ou “blocos visuais”. Para ser resolvido, cada item exige ao sujeito uma das seguintes tarefas cognitivas: sequenciar um algoritmo, concluir um algoritmo incompleto ou depurar um algoritmo incorreto. A Figura 1 apresenta um item do tipo “setas” que aborda os conceitos de “repetir até” e “repetir vezes”.

Figura 1. CTt, item 11: *loops* “repetir até + repetir vezes”; “setas visuais”

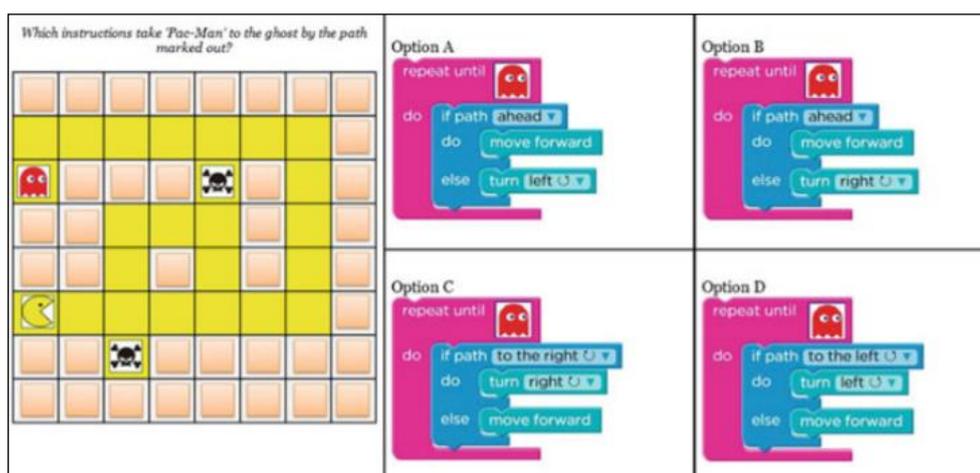


Fonte: Román-González et al. (2019)

Nesse item, o estudante precisa analisar no tabuleiro qual a sequência de passos que o personagem deve percorrer. Ao lado direito, um aninhamento de comandos de repetição determina a automatização do movimento do personagem. Porém, uma das instruções está incorreta, e o estudante deve ser capaz de identificar o erro.

A Figura 2 apresenta um item do tipo “blocos” que aborda os conceitos “repetir até” e condicionais compostas “se/senão”.

Figura 2. CTt, item 18: *loops* “repetir até” + condicional *if/else*; “blocos visuais”



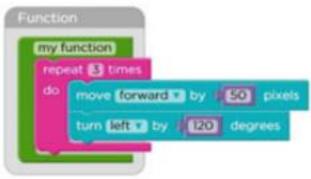
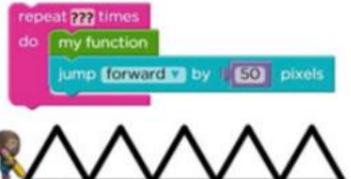
Fonte: Román-González et al. (2019)

Neste item, novamente o estudante precisa interpretar o caminho que o personagem deve percorrer. Porém, neste caso, as instruções são apresentadas em formato de bloco de comandos, e o

movimento do personagem agora depende dos obstáculos no caminho, que devem ser evitados por meio das estruturas condicionais. Apenas um dos quatro algoritmos está correto, por isso o estudante deve depurar cada opção para identificar a correta.

A Figura 3 apresenta um exemplo de item do tipo “bloco” no qual é necessário que o estudante conclua o algoritmo incompleto.

Figura 3. CTt, item 18: loops “repetir até” + condicional *if/else*; “blocos visuais”

<p>The following set of instructions is called 'my function', and draws one triangle of 50 pixels each side:</p> 	<p>Option A</p> <p>15</p>	<p>Option B</p> <p>5</p>
<p>The instructions below should make the artist draw the following design. Each side of each triangle measures 50 pixels. What is missing in the instructions?</p> 	<p>Option C</p> <p>4</p>	<p>Option D</p> <p>3</p>

Fonte: Román-González et al. (2019)

Nesse item, é proposto um padrão de desenho que o algoritmo deve produzir, porém falta a informação da quantidade de repetições em uma das etapas do programa. O estudante deve ser capaz de identificar qual o valor correto da repetição para garantir a execução do desenho, levando em consideração a implementação do conceito de “funções”.

O CTt demonstrou evidências de validade para avaliar o PC em indivíduos entre 10 e 16 anos ao ser submetido a um processo completo de validação relatado por Román-González (2016). Algumas etapas desse processo foram relatadas em publicações específicas, como a validação de conteúdo (ROMÁN-GONZÁLEZ, 2015), a validação de critério (ROMÁN-GONZÁLEZ et al., 2017b) e a validação preditiva (ROMÁN-GONZÁLEZ et al., 2018). A validade de conteúdo foi estabelecida pelo julgamento de 20 especialistas que refinou o teste de 40 para 28 itens. A confiabilidade do instrumento foi considerada boa ao ser medida por meio do cálculo do coeficiente alfa de Cronbach de aproximadamente 0.80. A validade de critério foi verificada ao analisar as correlações altas entre o CTt e outros instrumentos psicométricos, e os autores concluíram que foram fornecidas evidências suficientes de confiabilidade e validade do CTt. Por fim, a validade convergente

foi medida pelo cálculo da correlação entre o *CTt* e outros dois instrumentos, o *Dr. Scratch* e o *Bebras Tasks*. Mesmo com pontos fortes, os autores também reconhecem algumas limitações do *CTt*.

Como pontos fortes, os autores destacam que o *CTt* pode ser administrado em condições de pré-teste para medir o nível de desenvolvimento inicial do PC em estudantes sem experiência prévia em programação, pode ser administrado coletivamente para ser usado em exames maciços e detecção precoce de alunos com altas habilidades (ou necessidades especiais) para tarefas de programação, e pode ser utilizado para coletar dados quantitativos em avaliações pré-pós da eficácia de currículos ou programas destinados a promover o PC. Essas características levaram Couto (2018) a escolher o *CTt* para a comparação com o *CT Puzzle Test* em sua defesa de dissertação.

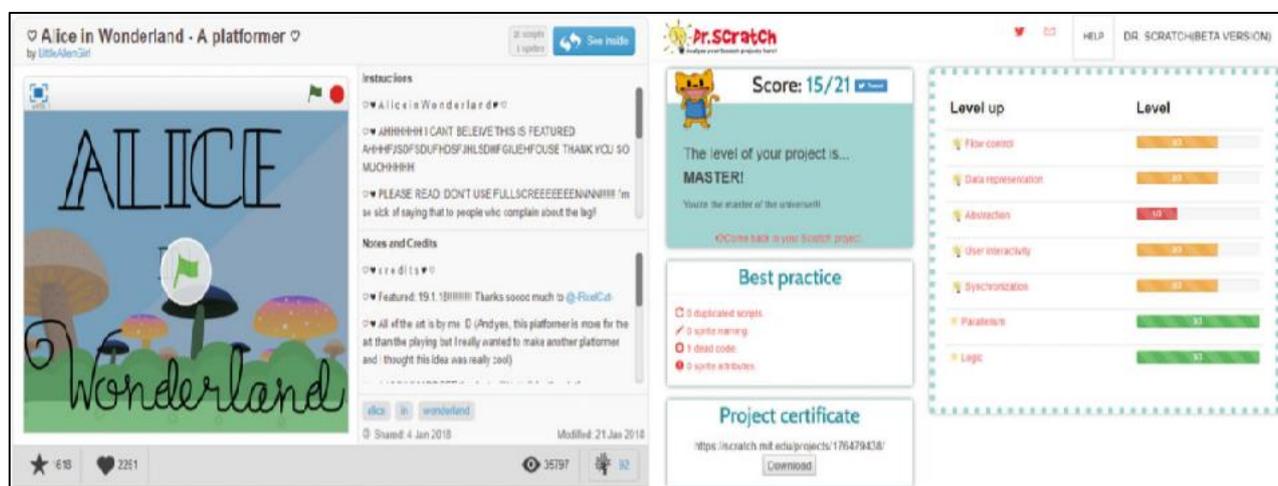
Como limitações, os autores apontam que o *CTt* fornece uma avaliação estática e descontextualizada, é excessivamente focado em “conceitos computacionais”, cobre apenas parcialmente “práticas computacionais” e ignora “perspectivas computacionais” (BRENNAN e RESNICK, 2012). Por fim, como o *CTt* é inteiramente projetado com itens de múltipla escolha, ele pode estar medindo o PC em seus níveis mais baixos de complexidade cognitiva (“reconhecer” e “entender”)(GOWS et al., 2013). Um instrumento destinado a medir o PC também em níveis mais altos de complexidade (“Aplicar” e “Assimilar”) (GOWS et al., 2013) deve incluir itens que requerem não apenas reconhecimento, mas também evocam o algoritmo correto; bem como abrir problemas complexos cuja resolução exige que os alunos transfiram de forma criativa o PC para diferentes domínios (ROMÁN-GONZÁLEZ et al., 2017b).

2.2.2 *Dr. Scratch*

O *Dr. Scratch* (MORENO-LEÓN et al., 2015) é um aplicativo *web* gratuito e de código aberto que analisa, de maneira automatizada, projetos programados em linguagem *Scratch*. A pontuação que o *Dr. Scratch* atribui a um projeto é baseada no grau de desenvolvimento de sete dimensões do PC: abstração e decomposição de problemas, pensamento lógico, sincronização, paralelismo, noções algorítmicas de controle de fluxo, interatividade do usuário e representação de dados. Essas dimensões são avaliadas estaticamente pela inspeção do código-fonte do projeto analisado e recebem uma pontuação de 0 a 3, resultando em uma avaliação total (“pontuação de domínio”) que varia de 0 a 21 quando todas as sete dimensões são agregadas. Além disso, o *Dr. Scratch* gera um relatório de *feedback* que inclui ideias e propostas para aprimorar as habilidades de PC dos alunos. O relatório de

feedback (Figura 4) também incentiva os alunos a experimentar novos blocos e estruturas de rascunho, a fim de melhorar seus próximos projetos.

Figura 4. *Dr. Scratch*: relatório de avaliação e *feedback* para o projeto “Alice in Wonderland”



Fonte: Román-González et al. (2019)

A figura acima apresenta o escore atingido pelo estudante, sugere dicas de melhores práticas para aprimorar sua produção, e apresenta a porcentagem de aderência do projeto a cada conceito computacional abordado.

A validade convergente do *Dr. Scratch* com relação às notas fornecidas por especialistas (MORENO-LEÓN et al., 2017) e com relação a várias métricas de complexidade de engenharia de software (MORENO-LEÓN et al., 2016), já foi relatada. Finalmente, o *Dr. Scratch* também demonstrou validade discriminante para distinguir entre diferentes tipos de projetos do *Scratch*, como animações, projetos de arte, projetos de música, histórias e jogos (MORENO-LEÓN et al., 2017b).

Por fim, o *Dr. Scratch* complementa o *CTt*, pois o primeiro inclui “práticas computacionais” (BRENNAN; RESNICK, 2012) que o *CTt* não apresenta, como iteração, teste, remixagem ou modularização. No entanto, o *Dr. Scratch* não tem a possibilidade de ser usado em condições puras de pré-teste, pois é aplicado aos projetos do *Scratch* depois que o aluno aprender pelo menos alguma codificação por um certo tempo (ROMÁN-GONZÁLEZ et al., 2017).

2.2.3 *Bebras Tasks*

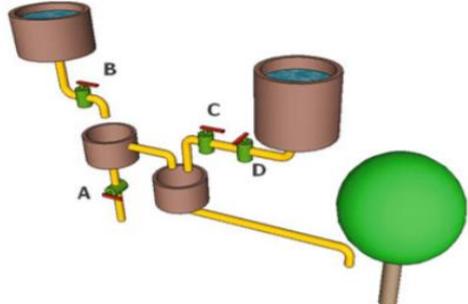
O Concurso Internacional *Bebras* é uma competição nascida na Lituânia em 2003, que visa promover o interesse e a excelência de estudantes do ensino fundamental e médio em todo o mundo

no campo da Ciência da Computação do ponto de vista do PC (CARTELLI et al., 2012; DAGIENE; FUTSCHEK, 2008; DAGIENE; STUPURIENE, 2014). A cada ano, o concurso propõe um conjunto de Tarefas (“*Bebras Tasks*”), cuja abordagem geral é a resolução de problemas reais, significativos para os alunos, por meio da transferência e projeção de seu pensamento computacional sobre os mesmos. As tarefas são independentes de qualquer software ou hardware específico e podem ser administradas a indivíduos sem nenhuma experiência anterior em programação.

O Concurso Internacional *Bebras* é um evento de promoção do PC, não um instrumento de medição. Isso porque, entre outras considerações, o concurso não é composto por um conjunto estável e determinado de itens de tarefa, mas por um conjunto que varia de ano para ano, com pequenas modificações entre cada país. No entanto, sua crescente expansão despertou o interesse de pesquisadores em psicométrica, que começaram a investigar suas possíveis virtudes como instrumento de medida do PC (ROMÁN-GONZÁLEZ et al., 2017b). A seguir, são apresentados alguns exemplos das tarefas do *Bebras Tasks*.

A Figura 5 apresenta a tarefa “abastecimento de água”, a qual propõe o desafio lógico de identificar quais registros entre 4 reservatórios de água precisam ser abertos para molhar uma árvore. Neste desafio, são apresentadas opções com alternativas de abertura e fechamento dos registros dos reservatórios de água, e o estudante precisa apontar qual das opções resulta no sucesso em fazer a água chegar até a árvore.

Figura 5. *Bebras Tasks* item 1: abastecimento de água (estrutura lógica/binária)

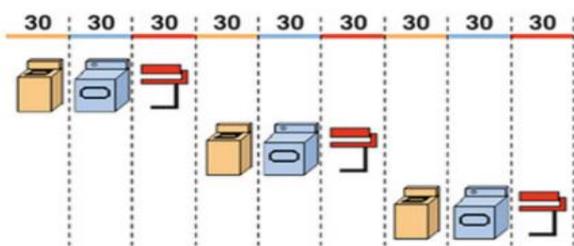
<p>Beaver has constructed a pipeline system to water his apple tree. The expressions contain variables A, B, C, D, which may be true or false. A variable has the value true, if the corresponding gate is open, and false, if it is closed.</p> <p>In which case the apple tree gets water?</p> 	<p>Option 1 A = false, B = true, C = false, D = false</p> <hr/> <p>Option 2 A = true, B = true, C = false, D = false</p> <hr/> <p>Option 3 A = true, B = false, C = false, D = true</p> <hr/> <p>Option 4 A = false, B = false, C = false, D = true</p>
--	---

Fonte: Román-González et al. (2019)

Para resolver a tarefa, o sujeito deve aplicar um pensamento condicional sobre a referida estrutura.

A Figura 6 apresenta a tarefa “lavanderia rápida”. A tarefa propõe a resolução de um cálculo para que uma lavanderia que dispõe de três máquinas com tarefas distintas termine de lavar as roupas de três clientes simultâneos no mínimo tempo possível. Para responder corretamente a essa tarefa, o sujeito deve entender que várias ações podem ser realizadas simultaneamente e em paralelo.

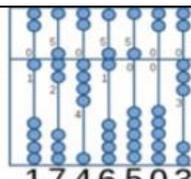
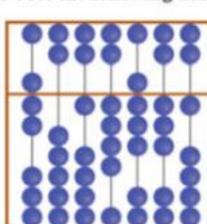
Figura 6. *Bebras Tasks* item 2: lavanderia rápida (paralelismo / algoritmos)

<p>Beaver Joe has started a new laundry business. He has got three machines: a washer, a dryer and a pressing iron. Every machine is connected through its own timer which provides for half an hour of electricity.</p> <p>So, when a client arrives, he needs 90 minutes for all of the three procedures. And three clients using the machinery consequently need 270 minutes.</p>  <p>30 30 30 30 30 30 30 30 30</p> <p>But now, there are three beavers arriving which are really busy. Each one of them has enough clothes for a load of its own. But they agree that they want to finish as quickly as possible.</p> <p>How many minutes does it take for all three of them to finish their laundry?</p>	<p>Option A 90 minutes</p> <hr/> <p>Option B 120 minutes</p> <hr/> <p>Option C 150 minutes</p> <hr/> <p>Option D 270 minutes</p>
--	--

Fonte: Román-González et al. (2019)

A Figura 7 apresenta a tarefa “ábaco”, na qual o estudante precisa identificar qual número está sendo representado no ábaco.

Figura 7. *Bebras Tasks* item 3: ábaco (abstração, decomposição, algoritmos)

<p>A number is represented on a Chinese abacus by the position of its beads. The value of a bead on the top part is 5; the value of a bead on the bottom part is 1. The abacus is reset to zero by pushing the beads away from the centre.</p> <p>To represent the number 1 746 503 the appropriate beads are moved towards the centre of the abacus:</p>  <p>1 7 4 6 5 0 3</p>			
<p>What number does the following abacus represent?</p> 			
<p>Option A 3014431</p>	<p>Option B 7514831</p>	<p>Option C 3514431</p>	<p>Option D 7014831</p>

Fonte: Román-González et al. (2019)

Nesta tarefa, o estudante precisa decompor as informações apresentadas, reconhecer os padrões e sequenciar o raciocínio para identificar o número solicitado.

O *Bebras Tasks* fornece uma avaliação naturalista e significativa, contextualizada em problemas da “vida real” que podem ser usados não apenas para medir, mas também para ensinar e aprender CT. No entanto, “as propriedades psicométricas dessas tarefas ainda estão longe de serem demonstradas, e algumas delas correm o risco de serem muito tangenciais ao núcleo do PC” (Román-ROMÁN-GONZÁLEZ et al., 2017b).

2.3 COMPUTATIONAL THINKING PUZZLE TEST

O *CT Puzzle Test* (GONÇALVES, 2015) é composto por itens no formato de *puzzles*, que abordam atividades de resolução de problemas, planejadas para que não seja necessário conhecimento prévio. Os itens que compõem o teste são classificados em fases. Uma fase é grupo de itens que apresenta o mesmo tipo de *puzzle*, porém com níveis de dificuldade diferentes. Assim, cada fase é composta por um ou mais itens. Inicialmente, Gonçalves implementou 9 (nove) tipos de *puzzle* (fases).

Para estimar a validade do instrumento, Gonçalves conduziu aplicações do teste na própria universidade. Para a validade aparente e de conteúdo, o pesquisador utilizou as seguintes amostras: especialistas do Grupo de Informática na Educação (GIE) – UNIVALI, composto por 11 integrantes entre doutores, mestres, doutorandos e mestrandos nas áreas da Computação, Educação, Matemática e Psicologia; 11 estudantes do 1º e 2º ano do ensino médio do Colégio de Aplicação da Universidade e uma turma do sexto período do curso superior de Ciência da Computação – UNIVALI. Para a validade concorrente, as amostras foram compostas por estudantes do 7º período do curso superior de Ciência da Computação, do 6º período do curso superior de Engenharia da Computação e do 1º período do curso superior de Psicologia. Para a validade preditiva, 16 estudantes do 9º ano do ensino fundamental da rede municipal de ensino foram expostos ao instrumento. Para estimativa de fidedignidade foi utilizado o método da consistência interna e o coeficiente alfa para todas as aplicações do teste. Os métodos de validade convergente e discriminante não foram abordados pelo fato de não haver outro instrumento à época para comparação de resultados. Os esforços de Gonçalves na investigação da validade e fidedignidade do *CT Puzzle Test* indicaram evidências fortes de validade do instrumento, mas o próprio autor relata que, “porém, para fins estatísticos, a aplicação com mais indivíduos se faz necessária” (Gonçalves, 2015).

Entre 2017 e 2018, durante a implantação da disciplina obrigatória de PC, o *CT Puzzle Test* foi utilizado por Couto (2018) para diagnosticar o desenvolvimento dos pilares do PC nos estudantes do CAU. Couto (2018) investigou o impacto da disciplina obrigatória de PC na percepção dos alunos com relação às suas próprias habilidades e suas opiniões sobre a importância das questões relacionadas ao PC. Couto fez uma análise qualitativa realizando entrevistas, e outra análise quantitativa aplicando o *CT Puzzle Test*.

Durante sua pesquisa, a autora descobriu que alguns itens do teste apresentavam inconsistências na aferição de seus escores, como erro na previsão de melhor resposta que influencia no escore de cada item. A partir dessa descoberta, Couto revisou e ajustou as fórmulas de cálculo de escore de todos os itens do instrumento, e calibrou o peso que cada tipo de interação exerce sobre o escore do referido item. Além disso, a pesquisadora percebeu que a fase 7 era de conhecimento comum pelos estudantes, o que demonstrou que o escore obtido nessa fase era resultado da memória sobre a resolução daquele tipo de *puzzle*, e não das etapas de testagem de hipótese para resolução de problemas. Por esse motivo, Couto eliminou todos os níveis dessa fase do instrumento.

Outras alterações importantes foram realizadas pela pesquisadora no instrumento visando diminuir o tempo para sua resolução e viabilizar sua aplicação em um período regular de aula. A fase 8 também foi totalmente removida por ser considerada de dificuldade excessiva, e as fases 1, 2, 3 e 7 tiveram a quantidade de níveis diminuídas.

Com alterações tão substanciais, Couto decidiu realizar a validação convergente do instrumento modificado, adotando como critério o *CTt* (ROMÁN-GONZÁLEZ, 2015). Além disso, essa etapa de validação ainda não havia sido executada no primeiro processo de validação do *CT Puzzle Test* por falta de outro instrumento semelhante para usar como critério à época, como apontou Gonçalves. A escolha do *CTt* para essa nova etapa de validação se deu por duas razões principais. A primeira é que o *CTt* foi o teste com mais características em comum com o *CT Puzzle Test* – é um instrumento independente de conhecimento prévio do sujeito, independente de linguagem de programação ou tecnologia, pode ser facilmente aplicado por ser um instrumento *on-line* e gera um escore que pode ser comparado com os resultados do *CT Puzzle Test*. A segunda razão é que o *CTt* já passou por testes para confirmar sua validade e fidedignidade (ROMÁN-GONZÁLEZ, 2015; ROMÁN-GONZÁLEZ et al., 2017b).

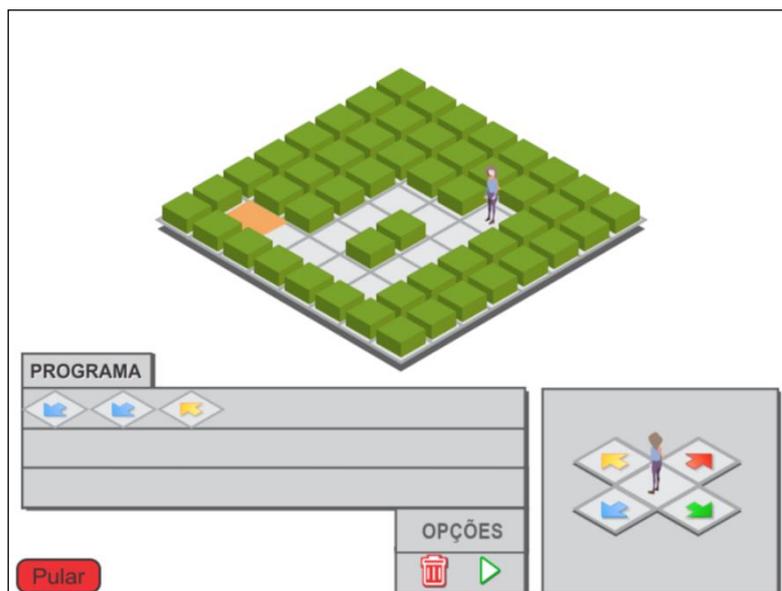
Nesta etapa de sua pesquisa, Couto decidiu adotar pesos para cada pilar do PC no cálculo dos escores dos itens do *CT Puzzle Test*. Couto adotou a metodologia proposta por Brackmann (2017). Brackmann também adotou o *CTt* em sua pesquisa, e sugeriu o mapeamento dos itens do *CTt* para os quatro pilares do PC. Para que a comparação fosse feita pelos mesmos critérios, Couto também mapeou os itens do *CT Puzzle Test* para cada pilar do PC. Em seguida, a autora utilizou um algoritmo para calcular a correlação entre os resultados dos dois testes e calibrar os pesos para alcançar a maior correlação possível. Os pesos propostos por Couto para o cálculo de escore de cada fase do *CT Puzzle Test* são detalhados na seção 2.3.8. A versão aprimorada do *CT Puzzle Test* é descrita a seguir.

Os diferenciais desse instrumento com relação aos instrumentos já existentes residem na avaliação por meio de *puzzles*, que estimula o raciocínio lógico e o pensamento crítico, e se caracteriza pela independência de domínio e conhecimento prévio, podendo ser aplicado nas condições de pré-teste e pós-teste, sem a necessidade de treinamento dos estudantes. O teste também considera a experimentação e teste de hipóteses por meio da captura das interações do estudante com o instrumento para o cálculo do escore. O teste é composto de sete tipos de fases, e algumas fases apresentam mais de um item no decorrer do teste com variações dos desafios conforme o nível de dificuldade aumenta. Algumas fases oferecem três botões de dicas de resolução que podem ser requisitadas pelo estudante e indicam o próximo passo necessário à conclusão do *puzzle*. Todas as fases oferecem a opção de “pular” para que o estudante avance para próxima fase sem completar a fase atual, ato que influencia no cálculo do escore final. Cada tipo de interação avaliada possui um peso e gera descontos ou acréscimos na pontuação de cada item, pontuação que pode variar de 0 a 1. Tanto as quantidades absolutas de cada tipo de interação, quanto os cálculos de escore de cada item são armazenados em banco de dados para permitir sua análise pelo aplicador do teste. Para cada fase, são atribuídos pesos aos quatro pilares do PC, e uma pontuação para cada pilar é mensurada. Ao final do teste, o aplicador dispõe dos dados brutos sobre as interações do estudante, dos escores de cada item e da pontuação geral de cada pilar do PC.

2.3.1 Primeira Fase

Para resolver o desafio da primeira fase, o indivíduo deve informar as direções para fazer com que o personagem se movimente pelos quadros de um tabuleiro partindo de um ponto inicial até um ponto final. Cada passo do personagem é representado por um comando que deve ser arrastado para o bloco PROGRAMA. A figura 8 apresenta essa fase.

Figura 8. Fase 1 do *CT Puzzle Test*



Fonte: Autor.

O estudante tem a sua disposição quatro comandos que representam as quatro direções para as quais o personagem pode caminhar, sendo cada comando equivalente a um quadro no tabuleiro. O estudante deve, portanto, arrastar para a sequência de comandos quantos passos forem necessários para fazer o personagem se locomover pelos quadros do tabuleiro, evitando os obstáculos, até que o personagem chegue ao local designado. A equação para calcular o escore da primeira fase é apresentada em (1):

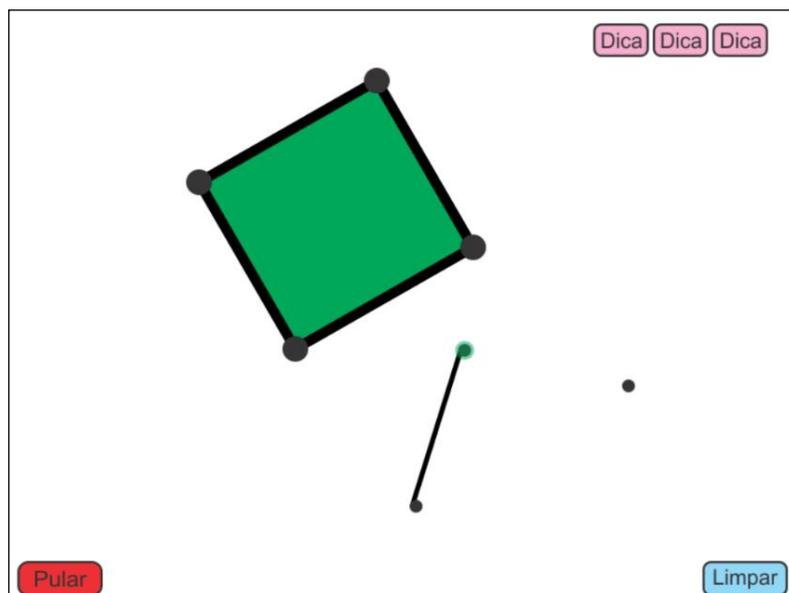
$$F1 = 1 - ((I - I_{min}) * 0.001) - (A * 0.05) - (L * 0.001) - ((P - P_{max}) * 0.01) - (T - 0.001) \quad (1)$$

Para aferir o escore dessa fase, são contabilizados o número de instruções utilizadas “I”, o número de instruções mínimo para resolver o desafio “I_{min}”, a quantidade de instruções apagadas “A”, a quantidade de vezes que utilizou o botão Limpar “L”, a quantidade de vezes que apertou o botão *Play* “P”, a quantidade de vezes máxima que pode pressionar *Play* sem ser penalizado “P_{max}” e “T” é o tempo em segundos para completar a fase.

2.3.2 Segunda Fase

A segunda fase (figura 9) é baseada na atividade de conectar os pontos para formar figuras.

Figura 9. Fase 2 do *CT Puzzle Test*



Fonte: Autor.

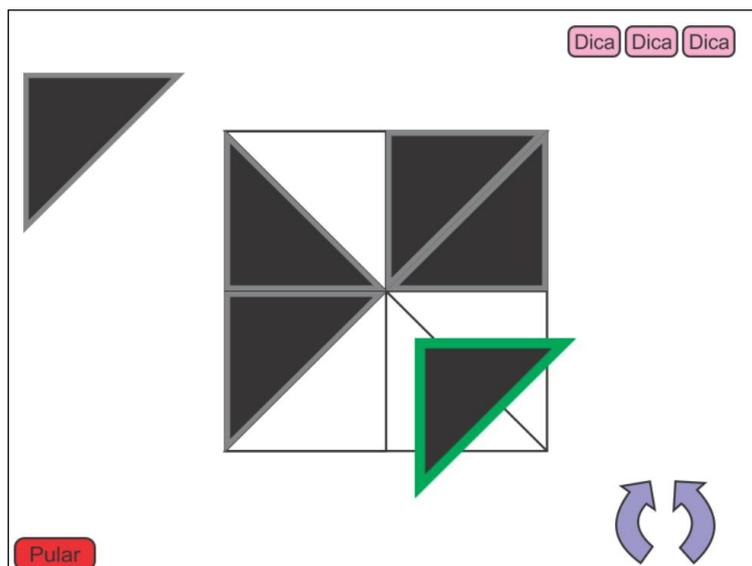
São apresentadas as figuras que deverão ser encontradas e em seguida as figuras desaparecem e pontos dispersos são espalhados na tela. O indivíduo deve então conectar os pontos de forma a criar o formato das figuras solicitadas. Para aferir o escore dessa fase, é utilizada a equação apresentada em (2):

$$F2=1-((C- Cmin)0.001)-(D*0.5)-(L*0.001)-(T*0.01) \quad (2)$$

Na equação, “C” representa o número de cliques efetuados, “Cmin” o número de cliques mínimo para resolver o desafio, “D” é a quantidade de dicas solicitadas, “L” a quantidade de vezes que utilizou o botão Limpar e “T” é o tempo em segundos para completar a fase.

2.3.3 Terceira fase

A terceira fase é composta por duas figuras onde a menor é uma peça que pode ser encontrada “n” vezes dentro da figura maior. A figura 10 apresenta essa fase.

Figura 10. Fase 3 do *CT Puzzle Test*

Fonte: Autor.

Para resolver este desafio o indivíduo deve arrastar as peças menores para dentro da figura maior, encaixando-as nos espaços disponíveis, até que as peças se esgotem. A equação para calcular o escore da terceira fase é apresentada em (3):

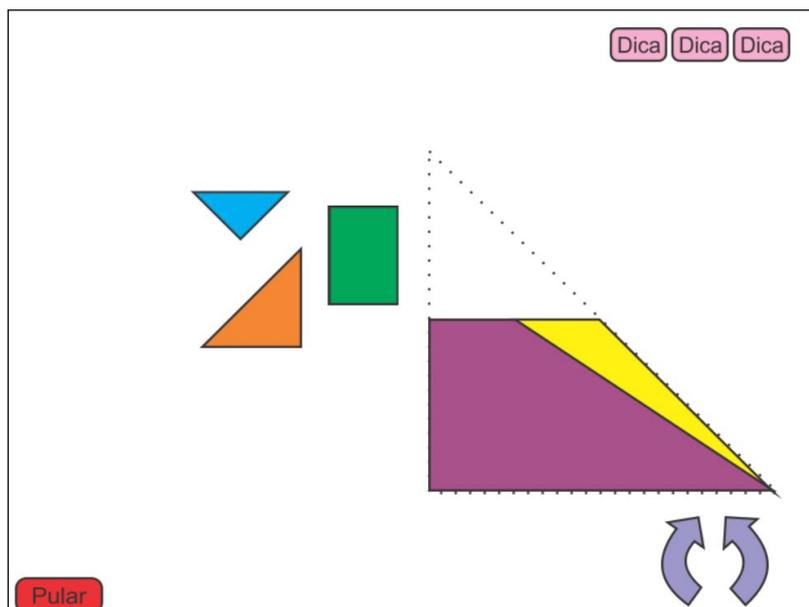
$$F3=1-((C- Cmin)0.001)-((G- Gmin)0.001)-(D*0.05)-(T*0.01) \quad (3)$$

Para aferir o escore dessa fase, são contabilizados a quantidade de cliques efetuados “C”, a quantidade mínima de cliques necessária para a resolução do *puzzle* “Cmin”, a quantidade de vezes que o estudante utilizou a opção Girar “G”, a quantidade mínima de giros necessários para concluir o desafio “Gmin”, a quantidade de dicas utilizadas “D”, e o tempo despendido “T”.

2.3.4 Quarta fase

A quarta fase é baseada no quebra-cabeça Tangram, onde partes de um triângulo retângulo estão dispersas e precisam ser agrupadas para formar a figura completa. A Figura 11 apresenta o *puzzle* com as peças dispersas na tela.

Figura 11. Fase 4 do *CT Puzzle Test*



Fonte: Autor.

O *puzzle* é composto de uma área pontilhada para onde as peças devem ser arrastadas, setas de rotação das peças e os botões de dicas que podem ser invocados pelo sujeito. O estudante deve selecionar uma peça, rotacioná-la se julgar necessário e arrastá-la para seu devido lugar, repetindo este processo até que a figura principal fique completa. Para aferir o escore dessa fase, é utilizada a equação apresentada em (4):

$$F4=1-((C- Cmin)0.001)-((G- Gmin)0.001)-(D*0.2)-(T*0.005) \quad (4)$$

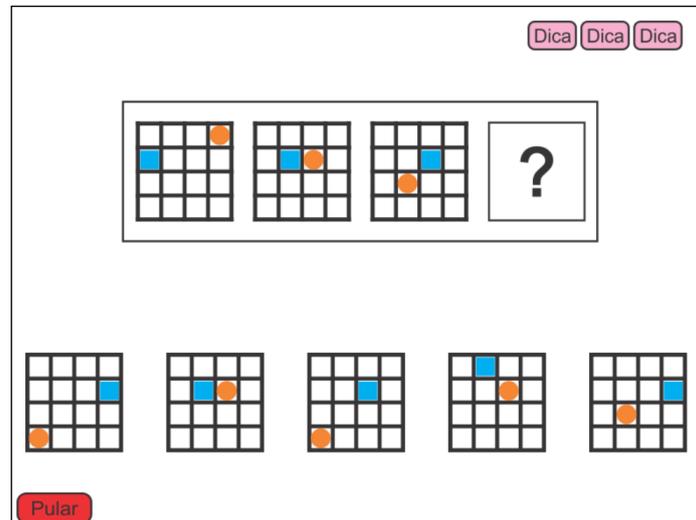
Semelhante à fase 3, para aferir o escore da quarta fase, são contabilizados a quantidade de cliques efetuados “C”, a quantidade mínima de cliques necessária para a resolução do *puzzle* “Cmin”, a quantidade de vezes que o estudante utilizou a opção Girar “G”, a quantidade mínima de giros necessários para concluir o desafio “Gmin”, a quantidade de dicas utilizadas “D”, e o tempo despendido “T”.

2.3.5 Quinta fase

Na elaboração da quinta fase foram utilizados os conceitos de análise de padrões. Para resolver este desafio é necessário identificar qual das opções fornecidas corresponde à próxima figura da sequência apresentada, de acordo com o padrão observado nas figuras que já compõem a sequência.

A Figura 12 apresenta o *puzzle* com a sequência de três figuras e um ponto de interrogação que representa a figura faltante. Logo abaixo, são exibidas as quatro opções dentre as quais o estudante deve optar e arrastar até o espaço identificado pelo ponto interrogação.

Figura 12. Fase 5 do *CT Puzzle Test*



Fonte: Autor.

A equação para calcular o escore da primeira fase é apresentada em (5):

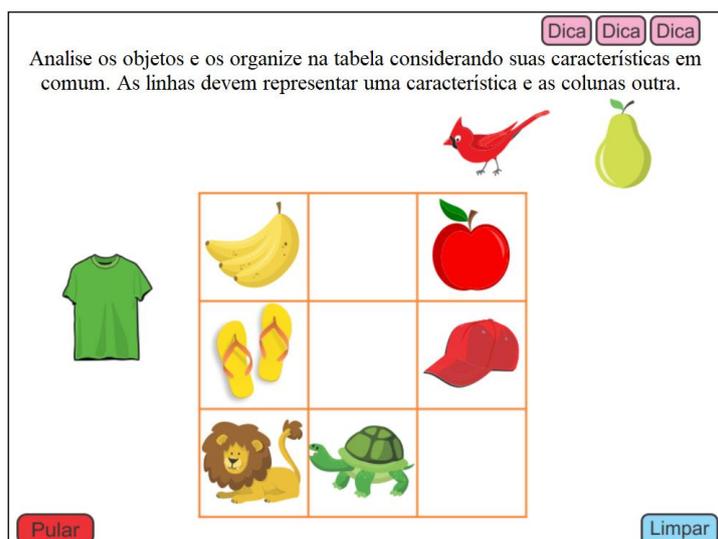
$$F5=1-(E*0.001)-(D*0.5)-(T*0.01) \quad (5)$$

Para aferir o escore dessa fase, são contabilizados a quantidade de tentativas “E”, a quantidade de dicas utilizadas “D” e o tempo despendido “T”.

2.3.6 Sexta fase

A sexta fase foi criada baseada no conceito de classificação e organização de dados. O indivíduo deve encontrar a relação entre os objetos espalhados pela tela e organizá-los em linhas e colunas de uma tabela considerando suas características (cor e classe), conforme ilustra a figura 13.

Figura 13. Fase 6 do *CT Puzzle Test*



Fonte: Autor.

O estudante pode organizar os objetos em diversas sequências desde que o resultado seja o agrupamento correto dos objetos. Caso complete a tabela e os objetos não estejam agrupados corretamente, os objetos são automaticamente retirados da tabela para que o estudante reinicie o *puzzle*. A equação para calcular o escore da primeira fase é apresentada em (6):

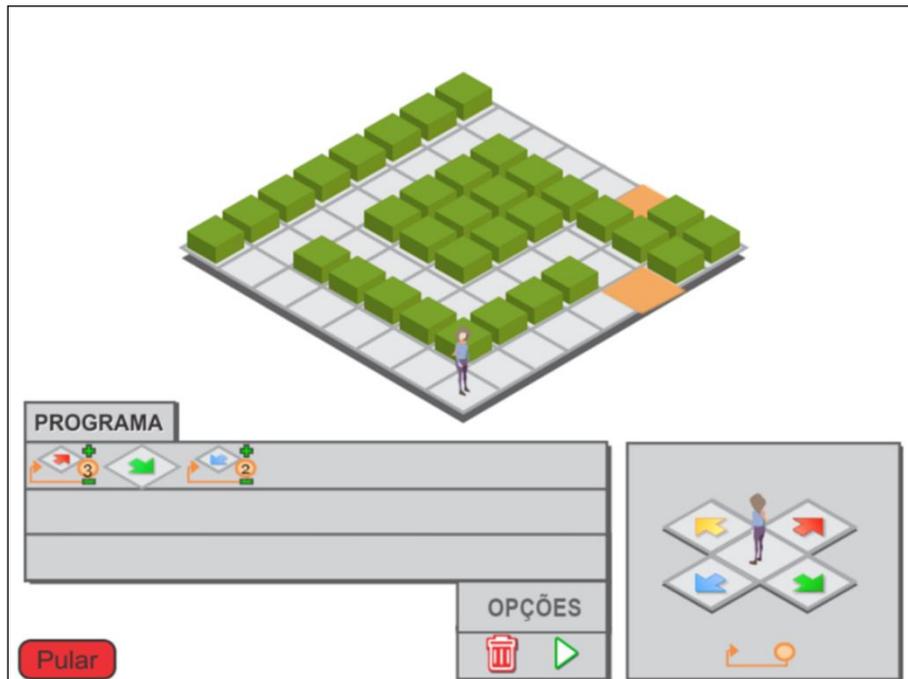
$$F6=1-(E*0.1)-(L*0.05)-(D*0.1)-(T*0.0005) \quad (6)$$

Para aferir o escore dessa fase, são contabilizados a quantidade de vezes que o estudante completou a tabela (quantidade de tentativas) “E”, a quantidade de vezes que o estudante utilizou a opção Limpar “L”, a quantidade de dicas utilizadas “D” e o tempo despendido “T”.

2.3.7 Sétima fase

A última fase do jogo é muito similar à primeira, porém é acrescido o comando *loop*. Apesar do comando *loop* estar disponível para utilização, é possível completar o *puzzle* sem ele. A Figura 14 apresenta a fase 7 do *CT Puzzle Test*.

Figura 14. Fase 7 do *CT Puzzle Test*



Fonte: Autor.

Como pode-se observar na Figura 10, abaixo dos comandos de direção está disponível o comando *loop*. Ao arrastar-se este comando para o bloco PROGRAMA, o estudante pode incluir comandos direcionais dentro do *loop* e indicar a quantidade de vezes que o comando deve ser repetido automaticamente. O cálculo de escore desse tipo de fase é mais complexo que os cálculos das demais fases. Para aferir o escore dessa fase, são aplicadas duas equações distintas. Os estudantes que não utilizaram o recurso de *Loop* efetivamente possuem seus escores calculados por (7):

$$F7 = 0.7 - (I - I_{min}) * 0.001 - (A * 0.001) - (L * 0.001) - ((P - P_{max}) * 0.001) - (T - 0.001) - (Lo * 0.001) \quad (7)$$

Já os estudantes que utilizaram o recurso de *Loop* possuem seus escores calculados por (8):

$$F9 = M - ((I - I_{min}) * 0.001) - (A * 0.001) - (L * 0.001) - ((P - I) * 0.001) - ((Lo - L_{min}) * 0.001) - ((ILO - ILO_{min}) * 0.001) - (T * 0.001) \quad (8)$$

Na equação (7), são considerados o número de instruções utilizadas “I”, o número de instruções mínimo para resolver o desafio “I_{min}”, a quantidade de instruções apagadas “A”, a quantidade de vezes que utilizou o botão Limpar “L”, a quantidade de vezes que apertou o botão *Play*

“P”, a quantidade de vezes máxima que pode pressionar *Play* sem ser penalizado “Pmax” e o tempo em segundos para completar a fase “T”. Quando o estudante não utiliza o recurso de *Loop*, ele é penalizado de forma que o escore máximo possível é 0.7. Adicionalmente, uma penalidade é adicionada para o uso de *Loops* vazios “Lo”, ou seja, sem instruções dentro.

Na equação (8), além de todas as variáveis da equação (7), foi inserida uma nova variável “M” que define o escore máximo a ser alcançado: os estudantes que utilizaram o *Loop* para resolver parte do desafio podem alcançar o escore máximo de 0.85, enquanto aqueles que utilizaram o *Loop* para resolver o desafio inteiro podem alcançar o escore máximo de 1. Além disso, são consideradas as variáveis “Lo” para contabilizar os *Loops* utilizados, “Lomin” para a quantidade mínima de *Loops* necessários para resolver o desafio, “ILo” que representa a quantidade de instruções dentro de um *Loop* e “ILomin” que é a quantidade mínima de instruções dentro do *Loop* para resolver o desafio.

2.3.8 Pesos dos Pilares do PC

Para criar um escore para as os pilares do PC, cada tipo de fase do *CT Puzzle Test* recebeu diferentes pesos aos seus escores com relação a cada pilar. A Tabela 1 apresenta os pesos atribuídos a cada fase em relação cada pilar do PC.

Tabela 1. Distribuição dos pesos das fases do *CT Puzzle Test* em relação aos pilares do PC

Pilares do PC	Pesos						
	Fase 1	Fase 2	Fase 3	Fase 4	Fase 5	Fase 6	Fase 7
Abstração	0.1	0.4	0.2	0.2	0.2	0.4	0.1
Algoritmo	0.4	0	0	0	0	0	0.4
Reconhecimento de padrões	0.1	0.4	0.4	0.4	0.4	0.3	0.1
Decomposição	0.45	0.2	0.4	0.4	0.4	0.3	0.4

Fonte: Adaptado de Couto (2018).

No processamento dos resultados do teste, os pesos apresentados na Tabela 1 são aplicados a ao escore de cada item, e o resultado total do escore de cada pilar do PC é finalmente calculado.

2.4 VALIDAÇÃO DE INSTRUMENTOS DE AVALIAÇÃO

Para uma utilização apropriada do instrumento de estudo com intuito de mensurar um traço latente, é preciso levar em consideração duas características essenciais: sua confiabilidade (ou fidedignidade) e sua validade.

2.4.1 CONFIABILIDADE

A confiabilidade (ou fidedignidade) refere-se, principalmente, “à estabilidade, consistência interna e equivalência de uma medida” (MARTINS, 2006). Estimativas de confiabilidade são afetadas por diversos aspectos do ambiente de avaliação (avaliadores, características da amostra, tipo de instrumento, método de administração) e pelo método estatístico utilizado. Portanto, os resultados de uma pesquisa utilizando instrumentos de medida só podem ser interpretados quando “as condições de avaliação e a abordagem estatística são apresentadas de maneira clara” (KOTTNER et al., 2011).

A confiabilidade de um teste diz respeito à capacidade do teste de medir sem erros aquilo a que se propõe. Medir sem erros significa que “o mesmo teste, medindo os mesmos sujeitos em ocasiões diferentes, ou testes equivalentes, medindo os mesmos sujeitos na mesma ocasião, produzem resultados idênticos” (PASQUALI, 2009). Entretanto, como o erro está sempre presente em qualquer medida, a análise da precisão de um instrumento psicológico visa mostrar precisamente o quanto ele se afasta da correlação ideal. A seguir, serão descritos os critérios de estabilidade, consistência interna e equivalência que serão abordados neste trabalho.

2.4.1.1 Estabilidade

A estabilidade de uma medida é o grau em que resultados similares são obtidos em dois momentos distintos (POLIT; BECK, 2011). A avaliação da estabilidade pode ser realizada pelo método de teste-reteste. Nesse tipo de método, o coeficiente de estabilidade é obtido ao se aplicar um mesmo teste aos mesmos sujeitos em duas ocasiões diferentes, que costumam ser separadas por um intervalo de tempo. Esse tipo de coeficiente é dado pela correlação entre os escores observados na primeira e segunda aplicações do teste (HOGAN, 2006). O intervalo de tempo entre as medições influenciará a interpretação da confiabilidade do teste-reteste, e a confiabilidade do teste-reteste tende a diminuir à medida que o tempo de reaplicação do teste é prolongado. Quanto à interpretação dos resultados, valores mínimos de 0,70 são considerados satisfatórios (NUNNALLY; BERNSTEIN, 1994).

2.4.1.2 Consistência Interna

A consistência interna – ou homogeneidade – indica se todas as subpartes de um instrumento medem a mesma característica (STREINER, 2003). Para alcançar consistência interna, todos os itens de um domínio devem realmente medir um determinado construto e não outro construto diferente.

Trata-se de uma importante propriedade de medida para instrumentos que avaliam um único construto, utilizando, para isso, uma diversidade de itens. Uma estimativa de consistência interna baixa pode significar que os itens medem construtos diferentes ou que as respostas às questões do instrumento são inconsistentes. A maioria dos pesquisadores avalia a consistência interna de instrumentos por meio do coeficiente alfa de Cronbach (STREINER; KOTTNER, 2014). O coeficiente alfa de Cronbach reflete o grau de covariância entre os itens de uma escala. Dessa forma, “quanto menor a soma da variância dos itens, mais consistente é considerado o instrumento” (PASQUALI, 2013). Apesar de o coeficiente alfa de Cronbach ser o mais utilizado na avaliação da consistência interna, ainda não há consenso quanto a sua interpretação. Embora estudos determinem que valores superiores a 0,7 sejam os ideais, algumas pesquisas consideram valores abaixo de 0,70 – mas próximos a 0,60 – como satisfatórios (STREINER, 2003).

Quanto à correlação média entre os itens, se esta for baixa, o valor do coeficiente alfa de Cronbach também será baixo. À medida que o coeficiente alfa aumenta, a correlação média acompanha essa elevação. Portanto, se as correlações forem altas, há evidência de que os itens medem o mesmo construto, satisfazendo a avaliação da confiabilidade (STREINER, 2003; CORTINA, 1993).

2.4.1.3 Equivalência

A equivalência refere-se ao “grau de concordância entre dois ou mais observadores quanto aos escores de um instrumento” (POLIT; BECK, 2011). A forma mais comum de avaliar a equivalência é a confiabilidade interobservadores, que “envolve a participação independente de dois ou mais avaliadores” (HEALE; TWYXCROSS, 2015). Nesse caso, o instrumento é submetido a respondentes treinados, e existe equivalência quando as pontuações obtidas forem as mesmas. A confiabilidade interobservadores depende, principalmente, “de um treinamento adequado dos avaliadores e de uma padronização da aplicação do teste” (ROUSSON et al., 2002). Quando existe elevada concordância entre os avaliadores, infere-se que os erros de medição foram minimizados.

2.4.2 VALIDADE

A validade refere-se ao fato de um instrumento medir exatamente o que se propõe a medir (ROBERTS et al., 2006; PASQUALI, 2017). Hogan (2006) comenta que foram distinguidos três aspectos de validade: validade de conteúdo, validade de critério e a validade de construto.

2.4.2.1 Validade de Conteúdo

A validade de conteúdo de um teste consiste em “verificar se o teste constitui uma amostra representativa de um universo finito de comportamentos (domínio)” (PASQUALI, 2009). A validade de conteúdo refere-se ao grau em que o conteúdo de um instrumento reflete adequadamente o construto que está sendo medido, ou seja, é a avaliação do quanto uma amostra de itens é representativa de um universo definido ou domínio de um conteúdo. Pasquali (2017) afirma que a etapa de validação de conteúdo inclui a análise teórica dos itens do teste, que visa verificar a compreensão das tarefas propostas no teste por parte dos testandos por meio de uma análise semântica, e a avaliação da pertinência do item à unidade correspondente, bem como o processo cognitivo envolvido por meio da análise de juízes.

A análise semântica envolve a compreensão dos itens pela população-alvo, ou seja, a população a que se destina o instrumento, tendo relevância quanto à inteligibilidade dos itens e a validade aparente (PASQUALI, 2010). Segundo Pelegrino et al. (2012) essa análise deve sempre ser realizada antes do teste, com uma pequena amostra da população.

Kimberlin e Winterstein (2008) afirmam que, como não existe um teste estatístico específico para avaliação da validade de conteúdo, geralmente utiliza-se primeiro uma abordagem qualitativa, por meio da avaliação de um comitê de especialistas, e depois uma abordagem quantitativa, com utilização do índice de validade de conteúdo (IVC). O IVC mede “a proporção ou porcentagem de juízes em concordância sobre determinados aspectos de um instrumento e de seus itens” (ALEXANDRE; COLUCI, 2011). Este método consiste de uma escala com pontuação de 1 a 4, em que: 1 = item não equivalente; 2 = item necessita de grande revisão para ser avaliada a equivalência; 3 = item equivalente, necessita de pequenas alterações; e 4 = item absolutamente equivalente (COLUCI et al., 2015). Os itens que receberem pontuação de 1 ou 2 devem ser revisados ou eliminados. Para calcular o IVC de cada item do instrumento, basta somar as respostas 3 e 4 dos participantes do comitê de especialistas e dividir o resultado dessa soma pelo número total de respostas, conforme fórmula a seguir (ALEXANDRE; COLUCI, 2011):

$$\text{IVC} = \text{N}^{\circ} \text{ de respostas "3" ou "4"} / \text{N}^{\circ} \text{ total de respostas}$$

O índice de concordância aceitável entre os membros do comitê de especialistas deve ser de no mínimo 0,80 e, preferencialmente, maior que 0,90.

Outro coeficiente utilizado para medir a concordância interobservadores é o coeficiente *Kappa*, aplicado a variáveis categóricas. Trata-se de uma “medida de concordância entre os avaliadores e assume valor máximo igual a 1,00. Quanto maior o valor de *Kappa*, maior a concordância entre os observadores. Valores próximos ou abaixo de 0,00 indicam a inexistência de concordância” (SALMOND, 2008).

2.4.2.2 Validade de Critério

A validade de critério trata do “grau de eficácia em que um teste prediz um desempenho específico de uma pessoa” (PASQUALI, 2009). A mensuração do coeficiente de validade é feita por meio do cálculo da correlação entre os escores do teste “e uma outra medida presente ou futura de desempenho das pessoas, que se chama critério” (HOGAN, 2006).

Costuma-se distinguir dois tipos de validade de critério: validade preditiva e validade concorrente. Segundo Pasquali (2009), “a diferença fundamental entre os dois tipos é basicamente uma questão do tempo que ocorre entre a coleta da informação pelo teste a ser validado e a coleta da informação sobre o critério.” O autor explica que a validade concorrente se dá quando as coletas forem realizadas com um curto intervalo de tempo, “mais ou menos” simultâneas. Caso os dados sobre o critério sejam coletados com um intervalo de tempo maior, quando a segunda coleta acontece em um momento futuro, fala-se em validade preditiva. O autor ainda destaca que o fato de a informação ser obtida simultaneamente ou posteriormente à do próprio teste não é um fator tecnicamente relevante à validade do teste. Relevante, sim, é a determinação de um critério válido. Assim, o que realmente importa para a validade do teste é a definição de um critério adequado contra o qual o instrumento avaliado será comparado, e que a medição desse critério seja feita de forma independente do instrumento que se está buscando validar.

Vários critérios podem ser utilizados para essa técnica de validação, segundo Pasquali (2009):

- **Desempenho Acadêmico:** consiste na obtenção do nível de desempenho escolar dos alunos, seja através das notas dadas pelos professores, seja pela média acadêmica geral do aluno, seja pelas honrarias acadêmicas que o aluno recebeu, etc. Esse tipo de critério é suscetível a subjetividade na avaliação dos alunos por seus professores.
- **Desempenho em treinamento especializado.** Trata-se do desempenho obtido em cursos de treinamento em situações específicas, como no caso de músicos, pilotos, atividades

mecânicas ou eletrônicas especializadas, etc. No final deste treinamento há tipicamente uma avaliação, a qual produz dados úteis para servirem de critério de desempenho do aluno.

- Desempenho profissional. Comparação dos resultados do teste com o sucesso/fracasso ou o nível de qualidade do sucesso dos sujeitos na própria situação de trabalho.
- Diagnóstico psiquiátrico. Muito utilizado para validar testes de personalidade/psiquiátricos.
- Diagnóstico subjetivo. Avaliações feitas por colegas e amigos podem servir de base para estabelecer grupos-critério. É utilizada esta técnica, sobretudo, em testes de personalidade, onde é difícil encontrar avaliações mais objetivas.
- Outros testes disponíveis. Os resultados obtidos por meio de outro teste válido, que prediga o mesmo desempenho que o teste a ser validado, servem de critério para determinar a validade do novo teste.

Ao explicar a utilização de outros testes disponíveis para a validação de critério de um instrumento, Pasquali (2009) chama atenção para o fato de que o teste selecionado precisa ser um teste já validado, mas que possua alguma desvantagem com relação ao teste que está sendo avaliado. Caso o teste usado como critério seja de alguma forma considerado superior ao teste que se deseja validar, não faria sentido construir um novo teste. Da mesma forma, caso o teste usado como critério seja inferior ao novo teste, também não faria sentido usá-lo como critério uma vez que o critério deve ser o objetivo que se deseja alcançar, a referência de qualidade que se deseja pôr à prova o novo instrumento proposto. Dessa forma, o critério adotado deve ser um teste já validado com qualidade comprovada, mas que, de alguma forma, possua limitações que podem ser compensadas pelo novo instrumento proposto. Nesse caso, os dois instrumentos acabam por se complementar.

2.4.2.3 Validade de Construto

A validade de construto é a extensão em que um conjunto de variáveis realmente representa o construto a ser medido (MARTINS, 2006; HAIR JUNIOR et al., 2009). A fim de estabelecer a validade de construto, geram-se “previsões com base na construção de hipóteses, e essas previsões são testadas para dar apoio à validade do instrumento” (HAIR JUNIOR et al., 2009). Dentre as

técnicas de investigação da validade de construto, as mais importantes são o teste de hipóteses, a validade transcultural e a validade estrutural ou fatorial.

O teste de hipóteses é uma fonte de evidência da validade de construto obtida pelas correlações entre o teste a se validar e outros testes. Existem diversas estratégias para confirmação da validade de construto pelo teste de hipóteses. Uma delas é a técnica de grupos conhecidos. Nesta abordagem, grupos diferentes de indivíduos preenchem o instrumento de pesquisa. Em seguida, os resultados dos grupos são comparados. Espera-se que resultados entre grupo semelhantes sejam similares, e que resultados entre grupos diferentes sejam divergentes e o instrumento se mostre sensível a ponto de detectar essas diferenças (KIMBERLIN; WINTERSTEIN, 2008). Outra técnica para verificar a validade de construto é obtê-la pelas avaliações da validade convergente e da validade discriminante. O objetivo é demonstrar que o teste alcança uma correlação consistentemente alta entre medidas delineadas para avaliar um mesmo construto ou construto similar (validade convergente), e que ao mesmo tempo apresenta correlações consistentemente baixas entre medidas para avaliar construtos distintos (validade discriminante) (POLIT, 2015).

A validade transcultural diz respeito à medida em que as evidências suportam a inferência de que o instrumento original e um adaptado culturalmente são equivalentes. Um instrumento traduzido e adaptado para outro contexto cultural, deve possuir um desempenho similar ao da versão original.

A validade estrutural ou fatorial é uma técnica muito utilizada entre os pesquisadores para “verificação da validade de construto para auxílio na construção, revisão e avaliação de instrumentos psicológicos, assim como na elaboração de teorias psicológicas” (LAROS et al., 2010). Segundo o autor, a análise fatorial é extremamente útil quando utilizada para análise de escalas que consistem em um bom número de itens utilizados para mensurar personalidade, formas de comportamento ou atitudes. A análise fatorial fornece ferramentas para “avaliar as correlações entre um grande número de variáveis, definindo os fatores, ou seja, as variáveis fortemente relacionadas entre si” (ALEXANDRE; COLUCI, 2011). Há dois tipos de análise fatorial: confirmatória e exploratória.

A análise fatorial exploratória (*exploratory factor analysis – EFA*) proporciona ao pesquisador a quantidade de fatores necessários para representar os dados, ou seja, é uma ferramenta para explorar a dimensionalidade de um conjunto de itens. Já a análise fatorial confirmatória (*confirmatory factor analysis – CFA*) “é um modo de confirmar quão bem as variáveis analisadas representam um número menor de construtos” (HAIR JUNIOR et al., 2009). Na *EFA*, as variáveis produzem cargas para todos

os fatores, enquanto na *CFA* as variáveis só produzem cargas nos fatores indicados no modelo. Dessa forma, o modelo confirmatório “é muito mais rigoroso e muito mais restritivo, motivo pelo qual é fortemente indicado para validação de questionários” (POLIT, 2015).

3 VALIDAÇÃO DO *CT PUZZLE TEST*

Conforme justificado nos capítulos anteriores, a validade é o aspecto mais importante para a avaliação da efetividade de um instrumento de avaliação. Um instrumento validado assegura a qualidade de seus escores e mensura o que de fato se propõe a mensurar. A validação de instrumentos de avaliação do PC é extremamente relevante para sua consolidação nos currículos educacionais.

O foco deste trabalho foi a realização da validação de conteúdo por meio das técnicas de avaliação por juízes e de análise semântica do *CT Puzzle Test*. As demais formas de verificação de confiabilidade e validade devem compor trabalhos futuros.

Neste capítulo são apresentadas as etapas de validação realizadas. A seção 3.1 descreve a revisão técnica do instrumento e a sistematização de sua aplicação, a fim de abordar os objetivos específicos 1, 2 e 3. As demais seções do capítulo abordam os objetivos específicos 4 e 5. Na seção 3.2.1, é apresentado o processo de validação de conteúdo do *CT Puzzle Test* por meio da avaliação de juízes convidados. A fase seguinte, relatada na seção 3.2.2, trata da aplicação dos testes para coleta de dados que servem de base para a análise semântica do teste.

3.1 REVISÃO DO INSTRUMENTO

O *CT Puzzle Test* começou a ser desenvolvido por Gonçalves em 2015. Nesse período, o teste foi planejado, construído e testado, e os primeiros resultados de sua aplicação foram divulgados. Em 2018, ao continuar o trabalho de Gonçalves, Couto submeteu o instrumento a novas amostras e adotou o *CTt* como critério para a análise dos resultados obtidos. Durante a análise, a pesquisadora identificou inconsistências no cálculo de escores do *CT Puzzle Test*. A partir dessa descoberta, Couto aprimorou o instrumento e implementou alterações importantes em sua estrutura por meio de revisão de fórmulas que resultaram na reclassificação de resultados.

Durante a pesquisa de Couto, a aplicação do *CT Puzzle Test* sofreu algumas limitações com relação a qualidade das amostras. Isso se deu ao fato de que as amostras sofreram influência para cada análise devido à disponibilidade dos estudantes e professores durante a aplicação dos testes. Os alunos participantes foram voluntários e a coleta de dados foi feita de forma anônima, o que pode ter influenciado o rigor de testagem. Ainda, houve um distanciamento no período entre a aplicação dos dois testes, visto que o *CT Puzzle Test* foi aplicado em um ano (2017), enquanto o *CTt* foi aplicado no ano seguinte (2018), quando situações como a progressão de série das turmas e a mobilidade dos

estudantes entre escolas – situações naturais na troca de ano letivo – também podem ter afetado as amostras.

Sobre os aspectos técnicos do instrumento, o *CT Puzzle Test* foi inicialmente desenvolvido por meio da linguagem *Action Script* para o ambiente *Flash*, uma tecnologia atualmente em desuso e sem suporte nos sistemas e navegadores de *Internet* atuais. Sobre o armazenamento dos dados, os resultados da aplicação do teste eram armazenados em uma tabela simples. Em sua pesquisa, Couto utilizou a versão do instrumento desenvolvida com a tecnologia *Flash*, e hospedou o projeto em um servidor que permitisse seu acesso a partir da *Internet*. Após suas análises, como uma etapa adicional da sua pesquisa, a pesquisadora aprimorou o instrumento ao converter o projeto para a linguagem *HTML5*. Porém, essa última versão não foi colocada em produção, ou seja, não foi testada com amostras reais. Além disso, ainda não havia sido implementada nessa versão a persistência de dados. Os resultados do teste eram apresentados em tela ao final da sua aplicação, e o sistema não armazenava esses resultados em banco de dados.

Fez-se necessário revisar o *CT Puzzle Test* com relação as tecnologias utilizadas em seu desenvolvimento, sua estrutura e arquitetura enquanto *software*. A respeito da aplicação prática do teste, uma revisão que contemplasse a definição de diretrizes para aplicação do teste e adicionalmente uma reestruturação da arquitetura de disponibilidade e acessibilidade do teste e de seus resultados se fez necessária para oportunizar a adoção do teste por pesquisadores, educadores e profissionais da saúde que tenham interesse em investigar o desenvolvimento do PC em indivíduos na sua área de atuação.

3.1.1 Revisão Técnica

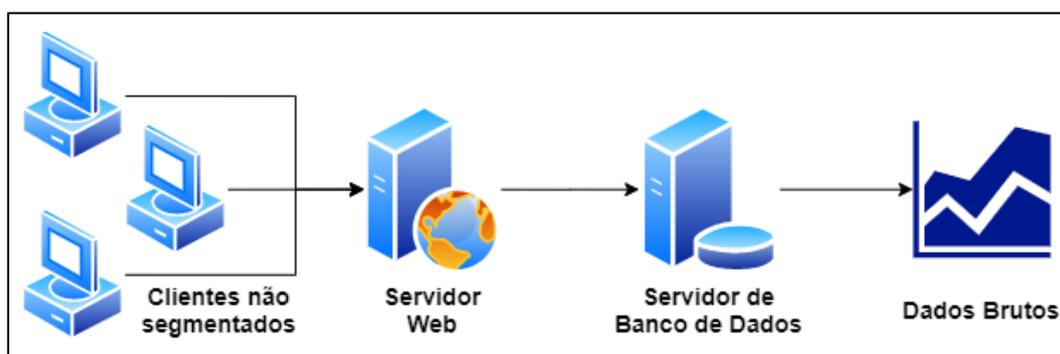
Nesta seção, é apresentada a contribuição deste trabalho para o aprimoramento técnico na arquitetura de sistema do instrumento. É discutida a necessidade de se aprimorar aspectos do desenvolvimento do projeto. Como resultado, são relatadas as modificações realizadas para garantir a acessibilidade e a disponibilidade do teste e dos resultados.

3.1.1.1 Evolução da Arquitetura do Sistema

A versão do *CT Puzzle Test* utilizada por Couto em sua pesquisa, desenvolvida com a tecnologia *Flash*, era hospedada em um servidor *web* para que o teste fosse acessível via *Internet*. Essa versão também armazenava em banco de dados os dados brutos sobre as interações dos

estudantes com o instrumento. O instrumento não calculava os escores dos estudantes de forma automática – os cálculos e análises dos resultados deveriam ser feitos pelo aplicador do teste de forma complementar. Além disso, os dados dos estudantes eram armazenados em uma tabela única, na forma de uma grande lista, o que não é ideal para o gerenciamento de múltiplas aplicações do teste. A Figura 15 apresenta o modelo de arquitetura dessa versão do *CT Puzzle Test*.

Figura 15. Modelo de arquitetura do *CT Puzzle Test* versão *Flash*



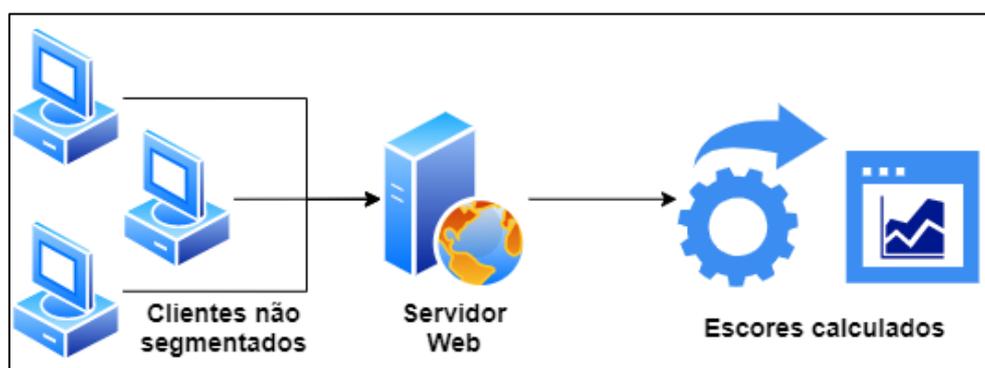
Fonte: Autor.

Nesse modelo, o teste é acessível por meio da *Internet* pois está hospedado em um servidor *web*. Os computadores dos estudantes, no papel de clientes do servidor *web*, requisitam o teste, que é transferido e apresentado em um navegador *web*, e assim a aplicação do teste é iniciada. O sistema solicita somente o nome do estudante para criar um registro no banco de dados e distinguir o resultado do estudante em relação aos demais registros armazenados. São armazenadas no banco de dados as interações dos estudantes para cada um dos níveis do teste, como quantidade de tentativas, quantidade de erros, quantidade de dicas utilizadas etc. Após a conclusão da aplicação do teste, o pesquisador precisa recuperar os dados armazenados no banco de dados e analisar os resultados por meio da compilação manual dos dados brutos. Os estudantes não são segmentados no banco de dados, o que significa que o teste armazena uma lista única de estudantes, sem distinguir as amostras. Esse modelo é suficiente para a pesquisa por um único pesquisador, mas a única forma de distinguir as amostras é a data e hora de realização do teste. Isso representa um obstáculo para o gerenciamento de múltiplas aplicações do teste, e torna inviável a distribuição do sistema para que outras instituições ou avaliadores façam uso da mesma instância do teste. Por utilizar a tecnologia *Flash*, em desuso e sem suporte em sistemas atuais, essa versão foi descontinuada após Couto finalizar sua pesquisa.

Mesmo tendo utilizado os resultados da versão *Flash* durante o período da pesquisa, Couto converteu o *software* para a tecnologia *HTML5* (amplamente compatível com os sistemas atuais) e

disponibilizou seu código fonte. Essa versão recebeu a implementação do cálculo automatizado de escores, mas não foi colocada em produção, ou seja, não foi aplicada com estudantes. Essa versão também não teve implementada a funcionalidade de armazenamento dos resultados em banco de dados. A figura 16 apresenta o modelo de arquitetura dessa versão do *CT Puzzle Test*.

Figura 16. Modelo de arquitetura do *CT Puzzle Test* versão *HTML5*



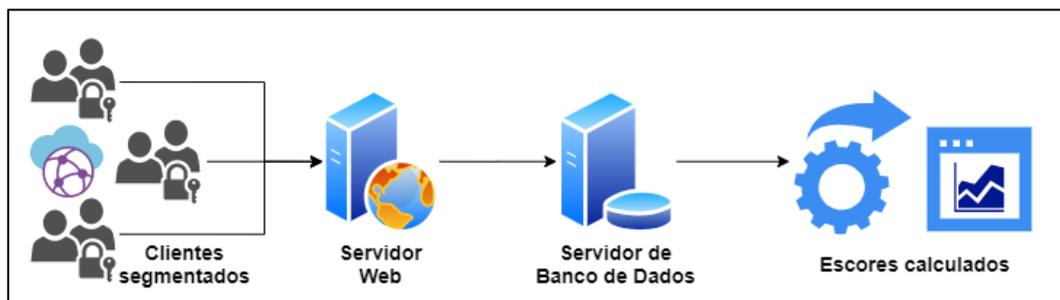
Fonte: Autor.

A versão *HTML5* de Couto também está hospedada em um servidor *web*, e a exemplo da versão anterior, o acesso ao teste é realizado por meio da *Internet*. Essa versão também adota somente o nome do usuário como critério de distinção, apesar de ainda não contar com a implementação do armazenamento em banco de dados. O teste processa as interações dos estudantes para cada um de seus níveis, e após a conclusão da aplicação, apresenta um relatório final que demonstra o desempenho do estudante, com os escores em todas as fases, além dos escores totais de acordo com os quatro pilares do PC. No entanto, conforme supracitado, o instrumento não armazena essas informações, que são perdidas se o teste for fechado ou reiniciado.

3.1.1.2 Contribuição deste trabalho

Para a evolução do *CT Puzzle Test*, foi implementado um sistema de cadastro que permite a criação de segmentos de usuários, para que cada aplicador interessado em adotar o teste seja capaz de gerenciar suas amostras. No modelo desenvolvido, aplicadores podem criar acessos próprios para suas amostras e têm acesso somente aos resultados correspondentes. Dessa forma, o sistema pode ser distribuído para uso simultâneo para quem quer que seja, e faz a distinção de seus usuários. A Figura 17 apresenta o modelo proposto.

Figura 17. Novo modelo de arquitetura do *CT Puzzle Test*

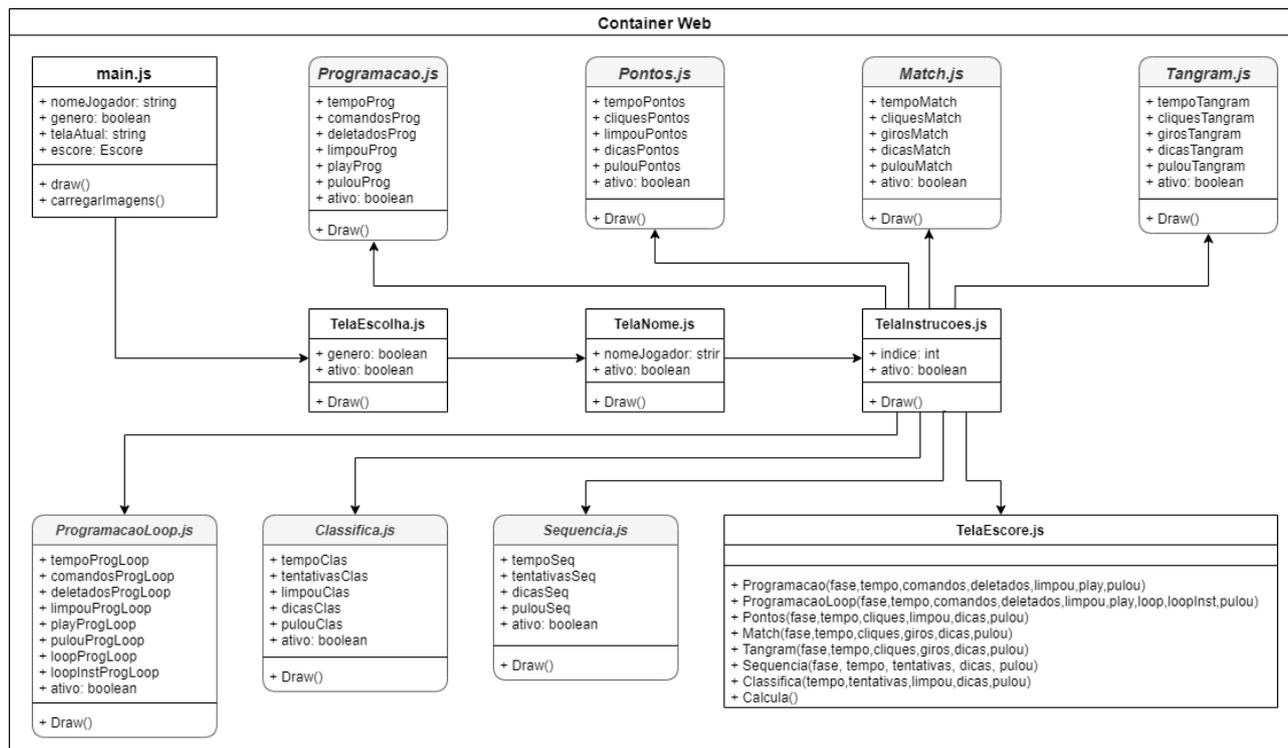


Fonte: Autor.

No novo modelo, o sistema permite que aplicadores cadastrem um ambiente composto por amostras que são distintas, protegidas por credenciais de segurança, o que garante acesso somente aos dados referentes às suas pesquisas. Esse modelo permite a ampliação do acesso ao teste e sua disseminação entre os pesquisadores e comunidade interessada no PC. O teste continua disponível por meio da *Internet*, e processa os resultados das aplicações para fornecer aos aplicadores os escores já calculados pela metodologia sugerida por Couto, assim como também fornece os dados brutos de interação dos estudantes com o instrumento para fins de análises estatísticas.

Na atual versão, que continua a adotar a tecnologia *HTML5*, o *CT Puzzle Test* é uma aplicação composta por uma página *Web* que implementa códigos em linguagem de *scripts*. A página *web* é uma página *HTML*, e a linguagem de *script* é o *JavaScript*. A página *web* serve de contêiner para os *scripts*, responsáveis pelas animações e imagens de cada fase do teste. Em cada tela ou módulo, o usuário interage com os elementos visuais das fases, como figuras de objetos, animações, personagens, formas geométricas entre outros. A cada nível, essas interações são armazenadas em memória para que ao final da aplicação o sistema efetue os cálculos de escore. Cada tipo de *puzzle* possui atributos específicos, que demandam comportamentos distintos do usuário para a execução de cada tarefa. A figura 18 apresenta o modelo que representa as dependências entre os módulos do sistema.

Figura 18. Dependência entre os módulos do *CT Puzzle Test*



Fonte: Autor.

Como pode ser observado na figura, o teste é composto de vários arquivos *JavaScript* (com extensão “.js”). Os módulos com o nome realçado na cor cinza correspondem às fases compostas por *puzzles*, enquanto os demais módulos representam as outras telas que compõem o teste. O teste é iniciado por um módulo principal que aciona todos os demais. A primeira tela mostrada ao usuário solicita que ele selecione um personagem que o acompanhará durante todo o teste, como tutor que dá instruções no início de cada nível, e que também participa do cenário em alguns casos. A partir daí o sistema começa a propor os *puzzles* e monitorar as ações do usuário que influenciam na avaliação do PC. Antes de cada fase, uma tela de instruções é apresentada. Ao final dos níveis de todas as fases, o sistema compila todas as interações do usuário em cada nível e calcula seus escores, armazenando os resultados no banco de dados para a disponibilização de relatórios ao aplicador do teste.

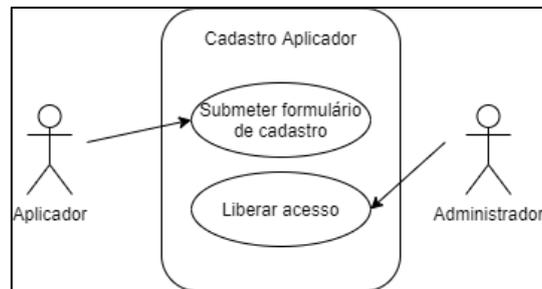
Para a implantação da funcionalidade de gestão das aplicações do teste foi desenvolvido um *website*². Este *website* permite que o aplicador se cadastre e cadastre suas amostras, criando uma chave de acesso única para cada amostra.

² <http://lite.acad.univali.br/ctpuzzle/>

3.1.1.3 Diagramas de Casos de Uso

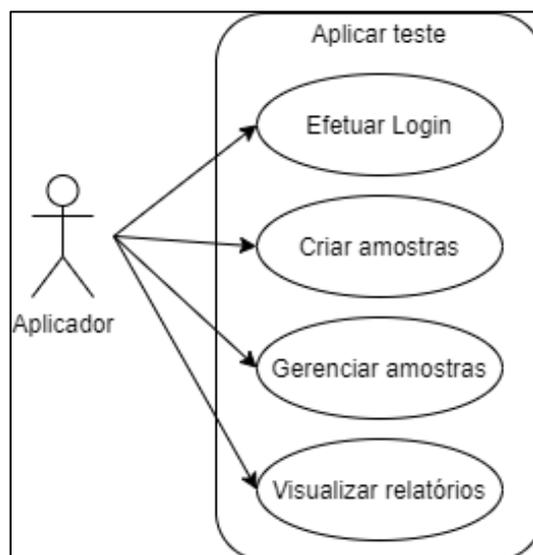
Os tipos de usuário possíveis são “administrador”, “aplicador” e “estudante”. A figura 19 apresenta o caso de uso “Cadastro de Aplicador”:

Figura 19. Caso de Uso “Cadastro de Aplicador”



Fonte: Autor.

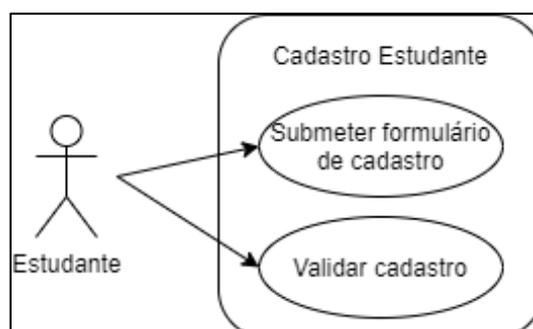
O usuário “aplicador” se aplica ao profissional interessado em aplicar o teste e avaliar o PC e indivíduos de seu interesse. Pode ser um professor que aplica o teste a seus estudantes, um pesquisador que aplica o teste a uma amostra ou um profissional clínico que aplica o teste a seus pacientes. O usuário “aplicador” deve acessar o *website* e se cadastrar na opção “Cadastre-se - Área do Aplicador”. Após preencher o formulário e submetê-lo, seu cadastro fica pendente de validação. O usuário “administrador” recebe um e-mail notificando sobre a submissão, e por meio de um *link* ele valida o cadastro. O aplicador recebe um e-mail informando que seu acesso está liberado e que ele pode efetuar login no sistema. Após ter seu cadastro validado, o aplicador pode criar amostras para coleta de resultados dos seus estudantes. A figura 20 apresenta o caso de uso “Aplicar Teste”:

Figura 20. Caso de Uso “Aplicar Teste”

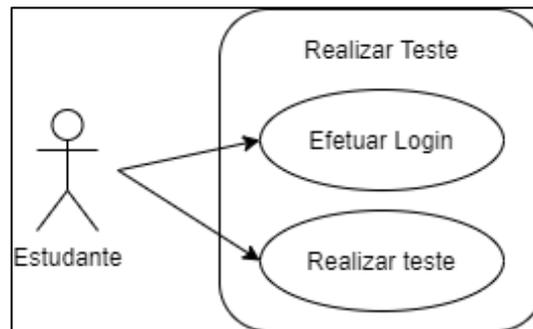
Fonte: Autor.

O usuário aplicador pode cadastrar novas amostras, editar suas informações incluindo a chave de acesso única da amostra, e configurar cada amostra como “aberta” permitindo que os estudantes utilizem a chave de acesso e executem o teste, ou “fechada”, impedindo que os estudantes acessem o teste. Este recurso permite o controle de início e fim das aplicações. Além disso, o aplicador pode visualizar relatórios contendo os resultados de cada uma das suas amostras.

O *website* também permite que estudantes se cadastrem por meio da opção “Cadastre-se – Área do Estudante”. Em seguida, basta utilizar a chave de acesso fornecida previamente pelo aplicador para ter acesso ao teste, conforme os casos de uso apresentados pelas figuras 21 e 22:

Figura 21. Caso de Uso “Cadastro Estudante”

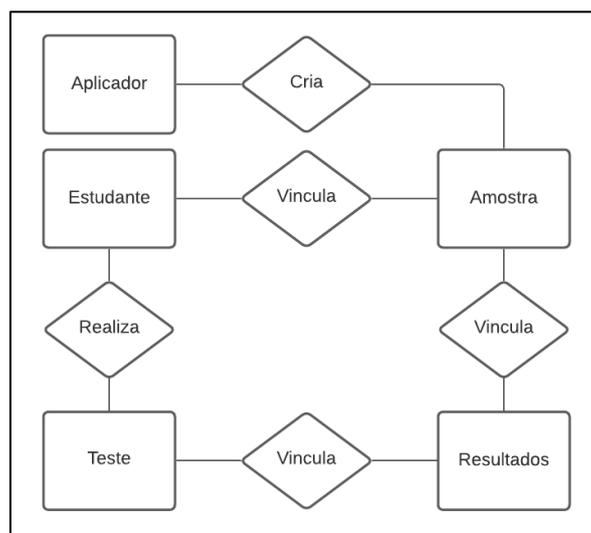
Fonte: Autor.

Figura 22. Caso de Uso “Realizar Teste”**Fonte:** Autor.

Diferente do usuário “aplicador” que precisa que sua condição de aplicador seja validada pelo administrador, o usuário “estudante” só precisa ativar sua conta para ter seu cadastro automaticamente validado. Após se cadastrar, o estudante recebe um e-mail com o link de ativação. Após isso, o estudante já pode efetuar *login* no sistema e utilizar a chave de acesso da amostra da qual faz parte para realizar o teste.

3.1.1.4 Modelo conceitual simplificado

Para ilustrar o funcionamento do sistema, é apresentado na figura 23 um modelo conceitual simplificado, omitindo-se seus atributos, que podem ser verificados na próxima seção.

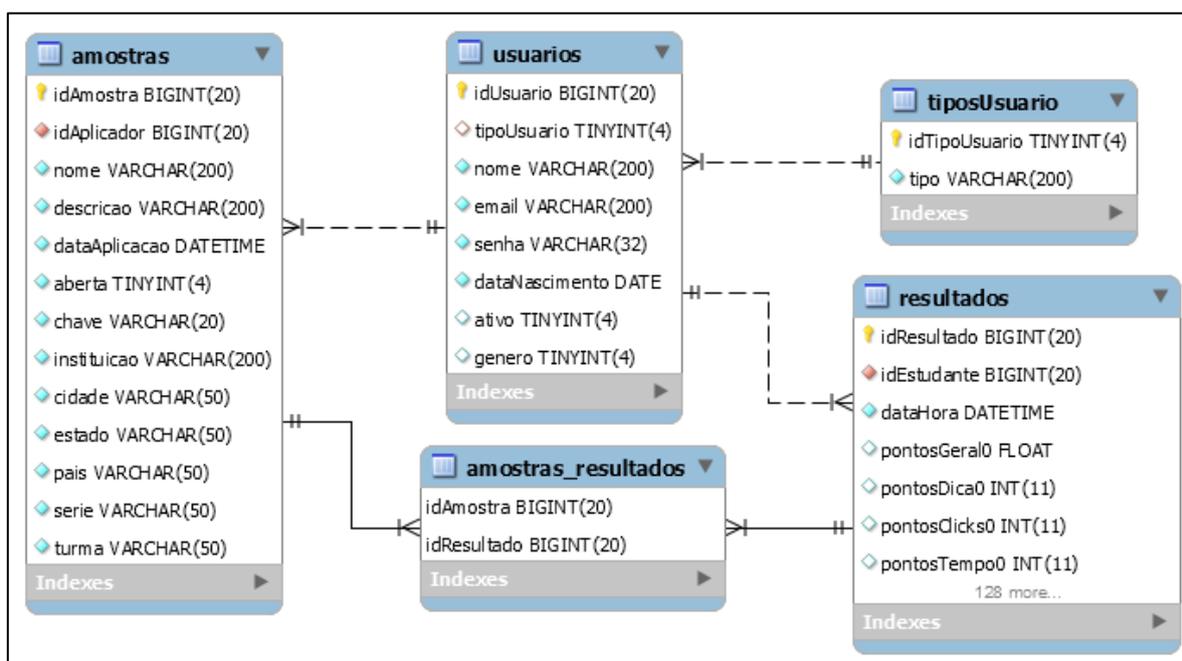
Figura 23. Modelo conceitual simplificado**Fonte:** Autor.

O usuário aplicador cria uma amostra e define uma chave de acesso. O aplicador deve então informar essa chave de acesso aos estudantes de sua amostra no ato de aplicação do teste. Ao utilizar uma chave de acesso, o resultado do teste do estudante fica vinculado a amostra correspondente. O aplicador então pode acessar os resultados de cada uma das suas amostras.

3.1.1.5 Banco de Dados

A primeira versão do teste armazenava todas as interações do indivíduo com o instrumento em uma única tabela, identificando cada registro pelo nome informado pelo estudante no momento da aplicação do teste. A nova versão implementou a identificação de cada usuário distintamente, por meio de identificador único, além de credenciais de login e senha. Outros recursos implementados foram a segmentação dos tipos de usuário (administrador, aplicador e estudante), segmentação de amostras (cada aplicador cria e gerencia suas próprias amostras), vinculação dos estudantes a amostras específicas por meio de chaves de acesso, e relatórios específicos sobre os resultados de cada amostra. A figura 24 apresenta o diagrama entidade relacionamento (DER) do banco de dados implementado:

Figura 24. Diagrama Entidade Relacionamento (DER)



Fonte: Autor

3.1.1.6 Correções da última versão

Durante a atualização de arquitetura e implementação de novas funcionalidades, diversos aspectos do teste foram atualizados. Alguns elementos visuais foram melhorados, enquanto outros pontos de caráter técnico foram aprimorados e/ou corrigidos. Diversos *bugs* foram corrigidos e as mensagens do teste de um modo geral foram revisadas. A lista a seguir apresenta algumas atualizações gerais de implementação:

- Os dados das interações que eram exibidos em tempo real durante as fases foram ocultados para não influenciar o estudante;
- As fases que avançavam automaticamente após a resolução do *puzzle* foram alteradas para que o usuário tenha controle sobre o avanço ao clicar no botão “Continuar”;
- A tela final que mostrava os resultados dos escores foi ocultada da etapa de validação do instrumento para ser redesenhada nas próximas versões, e foi substituída por uma tela de sucesso na conclusão do teste;

As fases de programação receberam as atualizações mais importantes. As principais alterações são descritas a seguir:

- As fases de programação sequencial foram totalmente reformuladas: foram definidos 4 níveis com escala crescente de dificuldade:
 - Nível 1: um objetivo com solução em linha reta;
 - Nível 2: um objetivo, uma curva necessária para solução;
 - Nível 3: dois objetivos, duas curvas necessárias para solução;
 - Nível 4: dois objetivos, três curvas necessárias para solução.
- As fases de programação com *loop* foram totalmente reformuladas: foram definidos 4 níveis com escala crescente de dificuldade:
 - Nível 1: dois objetivos com necessidade de comandos de repetição de um passo;
 - Nível 2: dois objetivos com necessidade de comandos de repetição de dois passos;
 - Nível 3: três objetivos com necessidade de comandos de repetição de três passos;
 - Nível 4: três objetivos com necessidade de comandos de repetição de três passos e melhor solução que explora a colisão do personagem nas barreiras, tornando a solução menos evidente.

- Outras atualizações das fases de programação:
 - A inclusão de passos no programa das fases de programação pode ser feita só com um clique, não é mais obrigatório arrastar;
 - O personagem agora volta para o início do caminho ao final de cada execução para permitir uma melhor depuração da sequência de passos já inserida em tentativas anteriores;
 - É exibida uma notificação ao final de uma execução sem sucesso para alertar o usuário a tentar novamente;
 - As cores das setas que indicam os passos do personagem foram modificadas para distinguir cada direção e permitir uma melhor visualização do algoritmo;
 - Adicionadas novas instruções para as fases de programação.

A mecânica de dicas de algumas fases também foi ajustada. As mudanças mais significativas são apresentadas a seguir:

- Cada botão de dica agora representa uma dica individual (antes as dicas possuíam uma única sequência, independente do botão selecionado);
- Os botões de dicas não desaparecem mais após serem utilizados, ficam somente marcados com outra cor;
- As dicas já usadas podem ser revistas clicando novamente sobre o botão já utilizado;
- Atualização das mensagens de dicas das fases Sequência, Classificar, Encaixe, Ligar Pontos e Tangram tornando as dicas mais fáceis de compreender.

As atualizações e aprimoramentos realizados no *CT Puzzle Test* neste trabalho dão continuidade à sua evolução, mas não encerram esse processo. Muitos aspectos ainda podem evoluir, e as principais contribuições futuras são apresentadas na seção 4.2.

3.1.2 Material de apoio

Nesta seção serão apresentados o material de apoio para aplicadores e o material de apoio para estudantes. Esse material terá o objetivo de instruir os usuários do teste sobre sua dinâmica, sua usabilidade entre outros aspectos.

Ressalta-se que devido às restrições impostas pelas medidas de combate à pandemia de COVID-19, houve prejuízos à definição e cumprimento das diretrizes de aplicação do teste. Em decorrência da impossibilidade de aplicação e acompanhamento *in loco*, os materiais de apoio sofreram adaptações no que diz respeito aos procedimentos de realização do teste. Essas adaptações

levam em conta um ambiente sem controle no qual a aplicação e as respostas ao instrumento acontecem em momentos assíncronos, sem a presença do aplicador, e os estudantes atendem ao teste a partir de seus ambientes residenciais, com seus próprios equipamentos e recursos tecnológicos, sem a possibilidade de pedir ajuda ou tirar qualquer dúvida. Portanto, as definições ora apresentadas somente se aplicam a este contexto.

Para que o teste seja aplicado em condições normais em trabalhos futuros, essas definições precisarão ser revisadas.

3.1.2.1 Material de apoio ao aplicador

Nesta seção serão apresentadas as instruções de aplicação do teste. A fim de orientar o aplicador interessado em adotar o *CT Puzzle Test*, o material de apoio é composto de informações a respeito de:

- O que é o *CT Puzzle Test* e quais seus objetivos;
- Como acessar o endereço eletrônico do teste e se cadastrar no sistema;
- Como criar e gerenciar amostras;
- Como organizar uma aplicação do teste;
- As condições de aplicação (local, tempo, recursos de informática necessários etc.);
- A utilização dos resultados.

Após se cadastrar no endereço eletrônico do teste, o aplicador/professor responsável pode criar amostras/turmas e definir chaves de acesso. Quando um estudante utiliza uma chave de acesso, fica automaticamente vinculado à amostra/turma correspondente. Depois que os estudantes realizam o teste, o responsável pode verificar os resultados dos escores e os dados brutos sobre as interações realizadas por cada estudante com o instrumento. As diretrizes de aplicação definidas para o teste são listadas a seguir:

- O teste deve ser individual;
- Deve preferencialmente ser realizado na presença do responsável;
- O tempo máximo para realização do teste é de 45 minutos;

- O público-alvo do teste são jovens de 14 a 17 anos, mas a depender do objetivo do responsável, o teste pode ser estendido a outras faixas etárias.
- Antes de iniciar o teste, os estudantes devem acessar o guia de introdução que é apresentado na página do teste após o login.
- O *CT Puzzle Test* é um teste on-line, e pode ser acessado de qualquer computador ou smartphone com conexão à Internet. No entanto, é indicado o uso de um computador para melhor visualização e precisão nos controles.

O material foi diagramado em formato PDF e está disponível no endereço eletrônico³ do teste.

A figura 25 apresenta algumas das instruções que compõem o material:

Figura 25. Material de apoio ao aplicador – instruções de utilização do teste

O infográfico apresenta as seguintes seções e conteúdos:

- UTILIZAÇÃO DO TESTE:**
 - O CT Puzzle Test é um teste on-line, e pode ser acessado de qualquer computador ou smartphone com conexão à Internet. No entanto, é indicado o uso de um computador para melhor visualização e precisão nos controles.
 - Após se cadastrar no endereço eletrônico do teste, o aplicador pode criar amostras e definir chaves de acesso. Quando um estudante utiliza uma chave de acesso de uma amostra, fica automaticamente vinculado à ela.
 - Depois que os estudantes realizam o teste, o aplicador pode verificar os resultados dos escores e os dados brutos sobre as interações realizadas por cada estudante com o instrumento.
- CADASTRO:**
 - Acesse o site do teste e faça um cadastro rápido: <http://lite.acad.univali.br/ctpuzzle>
 - No canto superior direito, clique em "Cadastre-se" e escolha "Área do Aplicador".
 - Preencha o formulário e envie.
 - Um administrador do sistema receberá uma notificação com seus dados para validar sua condição de aplicador. Após ter seu cadastro validado, você receberá um e-mail avisando que já pode acessar o sistema.
- AMOSTRAS:**
 - Após efetuar login no sistema com seu e-mail e senha anteriormente cadastrados, utilize o menu da página para acessar a opção "Minhas Amostras".
 - Depois, escolha "Nova Amostra".
 - Preencha o formulário e envie.
- CHAVE DE ACESSO:**
 - Ao cadastrar uma amostra o aplicador deve definir uma "chave de acesso".
 - Essa chave de acesso deve ser única, pois qualquer estudante que utilizá-la é vinculado à amostra correspondente.
 - No momento da aplicação do teste, o aplicador deve informar a chave de acesso aos estudantes.
- LIBERAÇÃO DAS AMOSTRAS:**
 - Ao cadastrar uma nova amostra, ela é automaticamente bloqueada. Isso significa que os estudantes não podem utilizar sua chave de acesso ainda.
 - Para liberar a utilização da chave de acesso da nova amostra, o aplicador deve acessar o menu "Minhas Amostras" e marcar a caixa de seleção "Liberada".
 - Após encerrar a aplicação do teste, basta desmarcar a mesma opção para bloquear novamente a chave de acesso da amostra correspondente.

Fonte: Autor.

3.1.2.2 Material de apoio ao estudante

Essas instruções são importantes para que o respondente compreenda as regras do instrumento antes de começar a respondê-lo, e orienta o estudante sobre:

- O que é o *CT Puzzle Test* e qual seu objetivo;
- Como acessar o endereço eletrônico do teste se cadastrar no sistema;

³ http://lite.acad.univali.br/ctpuzzle/CT_Puzzle_Test_Guia_Aplicador.pdf

- Como iniciar o teste;
- A organização do teste em fases;
- As ações possíveis de clicar, arrastar e soltar;
- Como funcionam os tutoriais;
- Como funcionam as dicas.

O material foi diagramado no formato *PDF* e está disponível no *website*⁴ do teste. A figura 25 apresenta algumas das instruções que compõem o material:

Figura 26. Material de apoio ao estudante - passos para iniciar o teste



Fonte: Autor.

3.2 VALIDAÇÃO DE CONTEÚDO

Como relatado na seção 1.1.4, as restrições impostas pelas medidas de combate à pandemia de COVID-19 resultaram na redução de escopo desta pesquisa. Inicialmente foi prevista a realização de mais etapas de validação, como a validação de critério e de construto. No entanto, sem acesso a públicos para aplicação do teste e sem controle sobre aplicações à distância, optou-se pela realização da etapa de validação de conteúdo, que poderia ser realizada sob essas condições.

⁴ http://lite.acad.univali.br/ctpuzzle/CT_Puzzle_Test_Guia_Instrucoes.pdf

Para realizar a validação de conteúdo do *CT Puzzle Test*, foram utilizadas as técnicas de avaliação por juízes e análise semântica. A seguir, serão apresentados o detalhamento das amostras, a metodologia de validação, os resultados e o impacto no instrumento.

Esta seção pretende testar as hipóteses de pesquisa H1 e H2.

3.2.1 Avaliação por juízes

O objetivo de submeter o teste a avaliação por juízes é julgar a pertinência de cada item do teste sobre a mensuração do PC com relação ao seu nível de dificuldade, nível de aderência a cada pilar do PC e sua relevância para diagnosticar o PC de um modo geral. Adicionalmente, esta etapa também contempla a avaliação dos juízes com relação à quantidade de itens do teste, à adequação das instruções que precedem os itens e ao tempo estimado para conclusão do teste.

3.2.1.1 Amostra

Para o processo de avaliação de juízes, 17 especialistas foram convidados para colaborar com o julgamento do instrumento. Seguindo as recomendações de Buffum et al. (2015), os especialistas convidados são principalmente professores da Área de Ciência da Computação com experiência e/ou publicações e envolvimento em projetos de pesquisa e implantação do PC na Educação Básica.

Dos 17 especialistas convidados, 8 aceitaram colaborar com esta pesquisa. O quadro 5 apresenta seus perfis profissionais:

Quadro 7. Perfis profissionais dos especialistas

#	Titulação	Instituição certificadora	Instituição onde atua	Área de pesquisa
1	Doutor	Institute of Education (IoE) University of London, UK	Universidade Federal Rural de Pernambuco (UFRPE)	Tecnologia para Educação
2	Doutor	Universidade Federal do Rio Grande do Sul (UFRGS)	Instituto Federal Farroupilha (IFFAR)	Informática na Educação
3	Doutor	Universidade Federal de Pernambuco (UFPE)	Universidade Federal de Pernambuco (UFPE)	Ciência da Computação
4	Doutor	Universidade Federal do Paraná (UFPR)	Universidade Tecnológica Federal do Paraná (UTFPR)	Inteligência Artificial aplicada à Educação
5	Mestre	Universidade Federal de Pernambuco (UFPE)	Universidade Federal de Pernambuco (UFPE)	Ciência da Computação

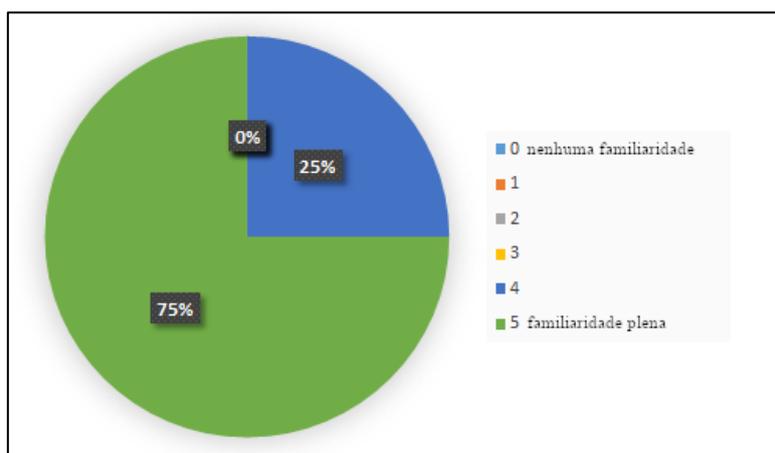
6	Doutor	Universidade Federal de Campina Grande (UFCG)	Universidade Estadual de Feira de Santana (UEFS)	Ciência da Computação
7	Doutor	Aalto University - School of Science and Technology, FI	Instituto Federal do Mato Grosso do Sul (IFMT)	Ciência da Computação
8	Doutor	Universidade Federal de Campina Grande (UFCG)	Universidade Federal da Paraíba (UFPB)	Ciência da Computação

Fonte: Autor.

Sobre a experiência dos especialistas na área de pesquisa do PC, foram realizadas duas perguntas no questionário de avaliação descrito na seção 3.2.2. Essas perguntas precedem a avaliação do teste em si.

A primeira pergunta foi: “Indique o seu nível de familiaridade com as definições de Pensamento Computacional (antes de responder a este questionário), onde 0 representa nenhuma familiaridade e 5 representa familiaridade plena”. O resultado é apresentado na figura 27:

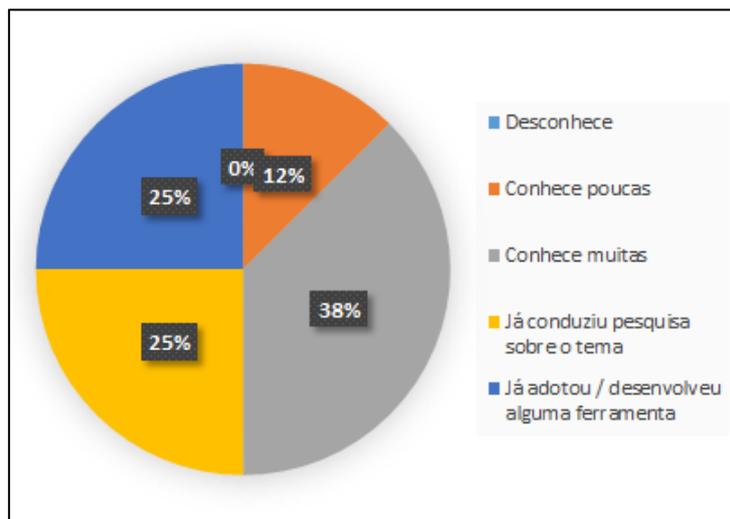
Figura 27. Nível de familiaridade dos especialistas com as definições do PC



Fonte: Autor.

De acordo com os resultados, 75% dos especialistas informou que possui familiaridade plena com as definições do PC. A segunda pergunta diz respeito à experiência do especialista sobre ferramentas de avaliação do PC: “Indique seu nível de conhecimento em termos de ferramentas de avaliação do Pensamento Computacional”, e as respostas possíveis foram: “desconhece”, “conhece poucas”, “conhece muitas”, “já conduziu pesquisa sobre o tema” e “já adotou / desenvolveu alguma ferramenta”. Os resultados são apresentados na figura 28:

Figura 28. Nível de conhecimento dos especialistas sobre ferramentas de avaliação do PC



Fonte: Autor.

Conforme se observa no gráfico, a maioria dos especialistas conhecem muitas ferramentas de avaliação do PC, e 50% deles já pesquisaram, adotaram ou desenvolveram ferramentas desse tipo.

Vale ressaltar que também há evidências de que as medidas de restrição para o combate à pandemia de COVID-19 podem ter prejudicado a participação dos convidados. O primeiro convite à lista de 17 especialistas foi realizado em novembro de 2020, quando foi sugerido um prazo de 10 dias para resposta. Ao fim do prazo, apenas 5 especialistas haviam respondido. Com o objetivo de estimular mais respostas, o convite foi reiterado algumas vezes. Ao final, o prazo foi prorrogado por mais vinte dias, totalizando trinta dias totais, até que os últimos três juízes se juntassem ao grupo inicial.

3.2.1.2 Instrumentos

Para que os especialistas realizassem a avaliação do *CT Puzzle Test*, foi formulado um questionário com perguntas referentes a características do instrumento. As perguntas do questionário, conforme recomendado por Buffum et al. (2015), combinam escalas, perguntas objetivas e perguntas abertas. Especificamente, as avaliações solicitadas dizem respeito a três aspectos do instrumento: características de cada item individual, características de cada tipo de fase e características gerais do instrumento.

Sobre as características de cada item individual foi questionado sobre o nível de dificuldade do item por meio de escala a variar de 0 (“muito fácil”) a 5 (“muito difícil”).

A respeito das características de cada tipo de fase, foram feitos os seguintes questionamentos:

- Relevância de cada item em medir o PC como um todo – escala a variar de 0 (“irrelevante”) a 5 (“totalmente relevante”);
- Dimensionamento da quantidade de itens de cada fase – resposta objetiva com as opções “Insuficiente”, “Adequada” e “Excessiva”;
- Nível de adequação das instruções apresentadas pelos tutoriais de cada fase – escala a variar de 0 (“inadequado”) a 5 “totalmente adequado”;
- Aderência do item a cada um dos quatro pilares do PC – escala a variar de 0 (“ausência completa de aderência”) a 5 (“aderência total”);
- Comentários e sugestões abertas para o aprimoramento de cada fase;

Os questionamentos relacionados às características gerais do instrumento foram:

- Dimensionamento da quantidade de fases (7 fases) – resposta objetiva com as opções “Insuficiente”, “Adequada” e “Excessiva”;
- Dimensionamento da quantidade total de itens (19 itens) – resposta objetiva com as opções “Insuficiente”, “Adequada” e “Excessiva”;
- Tempo estimado para realização do teste (45 minutos) – resposta objetiva com as opções “Insuficiente”, “Adequada” e “Excessiva”;
- Nível da capacidade do *CT Puzzle Test* em avaliar o PC em jovens de 14 a 17 anos – escala a variar de 0 (“incapaz”) a 5 “totalmente capaz”;
- Nível de adequação da abordagem do teste em utilizar *puzzles* em alternativa a abordagens de múltipla escolha – escala a variar de 0 (“inadequado”) a 5 “totalmente adequado”;
- Nível de adequação da interface gráfica e da linguagem adotada no teste ao público-alvo (jovens de 14 a 17 anos) – escala a variar de 0 (“inadequado”) a 5 “totalmente adequado”;

- Comentários e sugestões abertas sobre o instrumento em geral.

3.2.1.3 Metodologia

A avaliação foi conduzida por meio dos seguintes passos:

- Formação da lista de especialistas;
- Contato com cada especialista por correio eletrônico, serviços de mensagens e/ou ligação telefônica solicitando sua colaboração na avaliação do instrumento;
- Confirmação da lista de especialistas que aceitarem o convite;
- Envio de correio eletrônico aos especialistas com informações de apresentação do teste, instruções de acesso ao teste, instruções de acesso ao questionário de avaliação, instruções para preenchimento do questionário de avaliação e prazo para preenchimento;
- Recolhimento das respostas e compilação dos resultados.

Ao acessar o questionário, os juízes receberam instruções sobre como fundamentar sua avaliação. São apresentados o público-alvo do teste, a definição de Pensamento Computacional que deve ser considerada para a avaliação do instrumento, e uma explicação sobre cada pilar do PC para que os juízes mensurem o peso de cada pilar nos itens do *CT Puzzle Test*. Além disso, é apresentado o *link* de acesso ao teste com uma recomendação para que o(a) juiz(a) acesse o instrumento em paralelo ao questionário para resolver um item no teste e responder sobre este item no questionário.

3.2.1.4 Resultados da avaliação das fases

Para a análise dos resultados, inicialmente considerou-se a adoção do índice *Kappa* de concordância interobservadores. No entanto, para o presente trabalho, o coeficiente *Kappa* se mostrou inadequado, resultando em valores baixos. O *kappa* é usado principalmente para testar a concordância entre observadores que classificam categorias dicotômicas de dados (LANDIS; KOCH, 1977; SUEN; ARY, 1989). O uso de *kappa* com dados categóricos politômicos ou dados ordinais não é recomendado porque *kappa* mede a frequência de concordância exata *versus* concordância aproximada, e o valor de *kappa* é altamente dependente das definições das categorias. Se houver mais

de duas categorias de dados nominais, como é o caso deste trabalho, as diferenças entre os pares de dados causarão vários níveis de discordância entre observadores ou juízes (HUTCHINSON, 1993; MACLURE; WILLETT, 1987). Quando dados ordinais são usados, as distâncias entre as categorias podem contribuir para discordâncias entre os juízes.

Diante deste contexto, foi adotado o índice de validade de conteúdo (IVC), quando aplicável, para determinar o grau de concordância dos juízes sob cada aspecto do teste. O IVC permite que dois ou mais avaliadores revisem e avaliem independentemente a relevância de uma amostra de itens para o domínio de conteúdo representado em um instrumento. O pesquisador então registra a proporção de casos em que os avaliadores concordam e determina a estabilidade de sua concordância. O IVC considera uma escala de 1 a 4. Para o cálculo, considera-se que os itens que receberem pontuação 1 ou 2 devem ser revisados ou eliminados. O índice é calculado pela soma da quantidade de respostas 3 e 4, dividida pelo número total de respostas. Dessa maneira, um alto IVC demonstra que os juízes concordam que o aspecto avaliado está presente no item, enquanto um IVC baixo demonstra que os juízes concordam que o aspecto avaliado está ausente no item. Na presente pesquisa, foram feitas perguntas no formato escala cujas respostas variam de 0 a 5, e uma conversão para a faixa do IVC foi necessária. O quadro 6 apresenta essa metodologia:

Quadro 8. Conversão da escala aplicada para o IVC

Faixa da escala aplicada	Faixa do IVC	Representatividade
0 a 2	1 e 2	Aspecto com representatividade nula ou baixa
3 a 5	3 e 4	Aspecto com representatividade alta ou muito alta

Fonte: Autor.

Pela conversão apresentada, o IVC foi calculado somando-se a quantidade de respostas 3, 4 e 5, e dividindo-se o total pelo número de juízes (8) para cada pergunta.

Nos aspectos onde o IVC não foi aplicável, foram calculados somente a média e o desvio padrão das respostas.

3.2.1.4.1 Dificuldade dos itens

A primeira pergunta sobre os itens do teste foi a respeito da dificuldade de cada item. O objetivo do teste é aumentar a dificuldade gradativamente em cada tipo de fase, sem deixá-lo tão

difícil a ponto de torná-lo improvável de se alcançar a pontuação máxima. Dessa forma, caso algum item se apresentasse extremamente difícil, poderia ser eliminado do teste.

A tabela 2 apresenta os resultados das perguntas sobre a dificuldade dos itens do *CT Puzzle Test*:

Tabela 2. Dificuldade dos itens do *CT Puzzle Test*

Itens	Dificuldade (de 0 – muito fácil a 5 – muito difícil)									Média	Desvio Padrão
	Juíz 1	Juíz 2	Juíz 3	Juíz 4	Juíz 5	Juíz 6	Juíz 7	Juíz 8			
Fase 1 – Item 1	0	0	1	1	0	1	0	1	0,50	0,53	
Fase 1 – Item 2	1	1	1	2	0	2	1	2	1,25	0,71	
Fase 1 – Item 3	1	2	4	3	0	3	1	3	2,13	1,36	
Fase 1 – Item 4	1	2	3	3	0	4	1	3	2,13	1,36	
Fase 2 – Item 1	0	0	0	0	0	0	0	0	0,00	0,00	
Fase 2 – Item 2	0	1	1	0	5	1	1	2	1,38	1,60	
Fase 2 – Item 3	0	2	3	2	5	2	1	4	2,38	1,60	
Fase 2 – Item 4	1	3	4	1	5	3	2	4	2,88	1,46	
Fase 3 – Item 1	0	0	0	1	5	3	0	0	1,13	1,89	
Fase 3 – Item 2	0	1	1	0	0	4	0	1	0,88	1,36	
Fase 4 – Item 1	1	2	1	3	0	2	2	1	1,50	0,93	
Fase 4 – Item 2	1	3	1	3	0	3	3	2	2,00	1,20	
Fase 5 – Item 1	1	1	3	4	0	3	1	1	1,75	1,39	
Fase 5 – Item 2	2	1	3	3	0	2	1	3	1,88	1,13	
Fase 6 – Item 1	3	2	4	4	0	4	2	2	2,63	1,41	
Fase 7 – Item 1	0	2	3	3	0	3	3	2	2,00	1,31	
Fase 7 – Item 2	1	2	3	4	0	4	4	3	2,63	1,51	
Fase 7 – Item 3	1	3	3	4	0	4	3	3	2,63	1,41	
Fase 7 – Item 4	1	3	4	4	0	5	3	4	3,00	1,69	

Fonte: Autor.

A partir dos resultados das avaliações dos juízes, foi possível observar que, de fato, a dificuldade do teste se demonstra progressiva, iniciando em um nível fácil e terminando em um nível mais difícil, mas não difícil demais. O item 1 da fase 2 foi apontado como o mais fácil do teste. De fato, este item, que faz parte da fase de ligar pontos, apresenta pontos de tamanhos diferentes,

tornando a solução altamente intuitiva. No entanto, essa fase é composta de quatro itens, e percebe-se o aumento gradativo da dificuldade, o que atinge o objetivo do teste.

3.2.1.4.2 Relevância dos itens para avaliar o PC

A segunda pergunta sobre os itens do teste foi a respeito da relevância de cada tipo de fase na avaliação do PC. Caso a pertinência de alguma fase do teste fosse muito próxima de zero, ela poderia ser eliminada do instrumento. Neste caso, foi possível aplicar o IVC para medir o grau de concordância dos juízes sobre a relevância de cada item, onde quanto mais próximo de 1, maior a concordância de que o item é relevante para a avaliação do PC.

A tabela 3 apresenta os resultados das perguntas sobre a relevância dos itens do *CT Puzzle Test* na avaliação do PC:

Tabela 3. Relevância dos itens do *CT Puzzle Test* na avaliação do PC

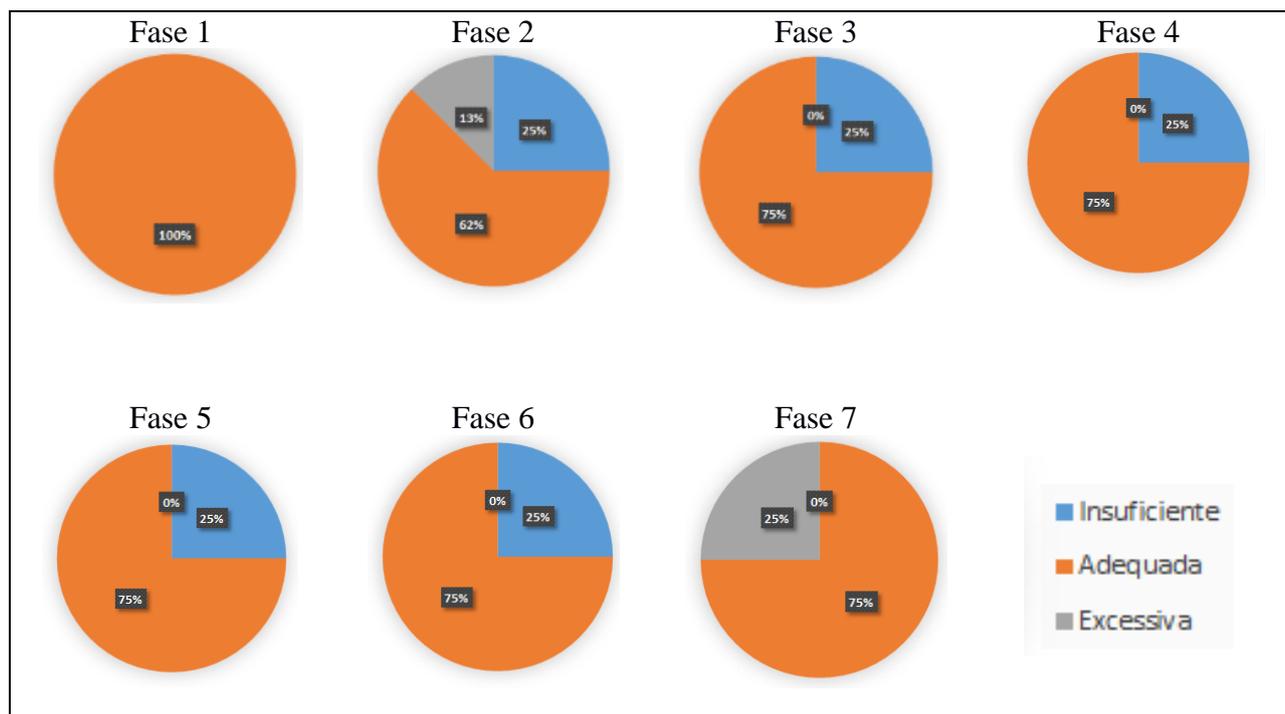
Fases	Relevância (de 0 – irrelevante a 5 – totalmente relevante)										IVC
	Juíz 1	Juíz 2	Juíz 3	Juíz 4	Juíz 5	Juíz 6	Juíz 7	Juíz 8	Média	Desvio Padrão	
Fase 1	3	5	5	5	5	5	3	4	4,38	0,92	1,00
Fase 2	3	4	4	2	2	4	2	3	3,00	0,93	0,63
Fase 3	2	4	3	5	2	4	2	3	3,13	1,13	0,63
Fase 4	3	5	3	5	2	5	2	3	3,50	1,31	0,75
Fase 5	3	3	3	5	2	5	2	3	3,25	1,16	0,75
Fase 6	3	5	4	5	4	5	2	4	4,00	1,07	0,88
Fase 7	3	5	5	5	5	5	3	4	4,38	0,92	1,00

Fonte: Autor.

Conforme observa-se nos resultados, todas as fases do teste alcançaram uma média superior a 3, o que indica que os juízes consideraram todas as fases como relevantes para a avaliação do PC. O IVC calculado reforça essa afirmação, com valores próximos de 1, e inclusive com o valor máximo alcançado nas fases de programação (fases 1 e 7).

3.2.1.4.3 Adequação da quantidade de itens de cada fase

Sobre a quantidade de itens para cada fase, a pergunta apresentava apenas três opções de resposta: insuficiente, adequada e excessiva. Neste caso, foram utilizados gráficos para apresentar os resultados, como se observa na figura 29:

Figura 29. Adequação da quantidade de itens de cada fase

Fonte: Autor.

A maioria dos juízes avaliou como adequadas as quantidades de itens em cada fase do *CT Puzzle Test*, sendo que em nenhum caso o grau de concordância foi menor que 62%.

3.2.1.4.4 Adequação das instruções ao estudante

Cada tipo de fase do teste é precedido por tutoriais que orientam o estudante sobre o que fazer. Assim, foi perguntado aos juízes sobre a adequação dessas instruções, conforme a tabela 4.

Tabela 4. Adequação das instruções (tutoriais) de cada fase

Fases	Adequação das instruções (de 0 – inadequado a 5 – totalmente adequado)										IVC
	Juíz 1	Juíz 2	Juíz 3	Juíz 4	Juíz 5	Juíz 6	Juíz 7	Juíz 8	Média	Desvio Padrão	
Fase 1	4	5	4	5	3	5	5	4	4,38	0,74	1,00
Fase 2	5	4	1	5	0	5	3	3	3,25	1,91	0,75
Fase 3	3	5	4	5	0	5	4	4	3,75	1,67	0,88
Fase 4	5	5	3	5	4	4	4	4	4,25	0,71	1,00
Fase 5	5	4	3	5	3	5	4	4	4,13	0,83	1,00
Fase 6	4	3	2	5	4	4	3	1	3,25	1,28	0,75
Fase 7	5	5	1	4	4	3	5	4	3,88	1,36	0,88

Fonte: Autor.

Conforme observa-se nos resultados, todas as fases do teste alcançaram uma média superior a 3, o que indica que os juízes consideraram que as orientações que antecedem as fases são adequadas para instruir o estudante na resolução do item. O IVC calculado demonstra um grau elevado de concordância entre os juízes.

3.2.1.4.5 Nível de aderência de cada pilar do PC

Com relação à aderência de cada tipo de fase a cada um dos quatro pilares do PC, os juízes avaliaram conforme é apresentado nas tabelas 5 a 11. Quanto maior o IVC sobre um pilar, maior é a concordância dos juízes de que o item avalia aquele pilar. Portanto, um IVC alto demonstra indícios sobre o peso que aquele pilar exerce sobre a fase, pois considera apenas as respostas sobre uma alta aderência do pilar na fase. Essa informação pode ser utilizada para validar os pesos aplicados ao cálculo de escore de cada tipo de fase, propostos por Couto (2018).

Tabela 5. Nível de aderência da fase 1 para cada pilar do PC

Pilar do PC	Aderência da fase 1 a cada pilar do PC (de 0 – ausência de aderência a 5 – aderência total)										
	Juíz 1	Juíz 2	Juíz 3	Juíz 4	Juíz 5	Juíz 6	Juíz 7	Juíz 8	Média	Desvio Padrão	IVC
Algoritmos	5	5	5	5	5	5	4	5	4,88	0,35	1,00
Recon. de Padrões	2	5	2	1	0	3	1	1	1,88	1,55	0,25
Decomposição	3	5	4	3	0	3	2	5	3,13	1,64	0,75
Abstração	3	5	2	4	3	1	1	3	2,75	1,39	0,63

Fonte: Autor.

Tabela 6. Nível de aderência da fase 2 para cada pilar do PC

Pilar do PC	Aderência da fase 2 a cada pilar do PC (de 0 – ausência de aderência a 5 – aderência total)										
	Juíz 1	Juíz 2	Juíz 3	Juíz 4	Juíz 5	Juíz 6	Juíz 7	Juíz 8	Média	Desvio Padrão	IVC
Algoritmos	1	3	1	2	0	5	3	5	2,50	1,85	0,50
Recon. de Padrões	4	5	3	5	0	5	4	5	3,88	1,73	0,88
Decomposição	4	2	2	2	0	3	1	2	2,00	1,20	0,25
Abstração	4	2	5	5	0	1	1	4	2,75	1,98	0,50

Fonte: Autor.

Tabela 7. Nível de aderência da fase 3 para cada pilar do PC

Pilar do PC	Aderência da fase 3 a cada pilar do PC (de 0 – ausência de aderência a 5 – aderência total)										
	Juíz 1	Juíz 2	Juíz 3	Juíz 4	Juíz 5	Juíz 6	Juíz 7	Juíz 8	Média	Desvio Padrão	IVC
Algoritmos	1	3	1	1	0	5	3	1	1,88	1,64	0,38
Recon. de Padrões	5	5	4	5	0	5	2	5	3,88	1,89	0,75
Decomposição	4	5	4	4	0	5	1	2	3,13	1,89	0,63
Abstração	4	3	4	5	0	2	1	5	3,00	1,85	0,63

Fonte: Autor.

Tabela 8. Nível de aderência da fase 4 para cada pilar do PC

Pilar do PC	Aderência da fase 4 a cada pilar do PC (de 0 – ausência de aderência a 5 – aderência total)										
	Juíz 1	Juíz 2	Juíz 3	Juíz 4	Juíz 5	Juíz 6	Juíz 7	Juíz 8	Média	Desvio Padrão	IVC
Algoritmos	0	2	1	1	0	5	2	2	1,63	1,60	0,13
Recon. de Padrões	5	5	3	5	0	5	4	5	4,00	1,77	0,88
Decomposição	4	5	4	5	5	5	1	2	3,88	1,55	0,75
Abstração	4	3	4	5	0	3	4	5	3,50	1,60	0,88

Fonte: Autor.

Tabela 9. Nível de aderência da fase 5 para cada pilar do PC

Pilar do PC	Aderência da fase 5 a cada pilar do PC (de 0 – ausência de aderência a 5 – aderência total)										
	Juíz 1	Juíz 2	Juíz 3	Juíz 4	Juíz 5	Juíz 6	Juíz 7	Juíz 8	Média	Desvio Padrão	IVC
Algoritmos	0	1	3	1	0	5	1	0	1,38	1,77	0,25
Recon. de Padrões	5	5	5	5	5	5	4	5	4,88	0,35	1,00
Decomposição	4	4	2	4	0	4	1	1	2,50	1,69	0,50
Abstração	5	3	2	5	0	5	3	5	3,50	1,85	0,75

Fonte: Autor.

Tabela 10. Nível de aderência da fase 6 para cada pilar do PC

Pilar do PC	Aderência da fase 6 a cada pilar do PC (de 0 – ausência de aderência a 5 – aderência total)										
	Juíz 1	Juíz 2	Juíz 3	Juíz 4	Juíz 5	Juíz 6	Juíz 7	Juíz 8	Média	Desvio Padrão	IVC
Algoritmos	0	4	1	2	0	5	3	0	1,88	1,96	0,38
Recon. de Padrões	5	5	5	5	5	5	3	5	4,75	0,71	1,00
Decomposição	4	4	3	5	0	4	2	4	3,25	1,58	0,75
Abstração	5	5	5	5	0	5	2	5	4,00	1,93	0,75

Fonte: Autor.

Tabela 11. Nível de aderência da fase 7 para cada pilar do PC

Pilar do PC	Aderência da fase 7 a cada pilar do PC (de 0 – ausência de aderência a 5 – aderência total)										
	Juíz 1	Juíz 2	Juíz 3	Juíz 4	Juíz 5	Juíz 6	Juíz 7	Juíz 8	Média	Desvio Padrão	IVC
Algoritmos	5	5	5	5	5	5	4	5	4,88	0,35	1,00
Recon. de Padrões	2	5	5	5	5	5	3	4	4,25	1,16	0,88
Decomposição	3	5	4	5	4	5	3	3	4,00	0,93	1,00
Abstração	4	5	2	5	4	3	2	3	3,50	1,20	0,75

Fonte: Autor.

Os resultados apresentados nas tabelas demonstram quais pilares do PC e em qual grau esses pilares estão presentes nos tipos de fase do teste. Dessa forma, é possível comparar os níveis de aderência dos pilares do PC aos pesos que cada pilar exerce no cálculo de escore da fase, de acordo com a definição de Couto (2018), como é demonstrado pela tabela 12:

Tabela 12. Peso dos pilares sobre os escores *versus* IVC da aderência aos pilares

Pilares do PC	Fase 1		Fase 2		Fase 3		Fase 4		Fase 5		Fase 6		Fase 7	
	Peso	IVC												
Algoritmos	0,4	1,00	0,0	0,50	0,0	0,38	0,0	0,13	0,0	0,25	0,0	0,38	0,4	1,00
Recon. de padrões	0,1	0,25	0,4	0,88	0,4	0,75	0,4	0,88	0,4	1,00	0,3	1,00	0,1	0,88
Decomposição	0,45	0,75	0,2	0,25	0,4	0,63	0,4	0,75	0,4	0,50	0,3	0,75	0,4	1,00
Abstração	0,1	0,63	0,4	0,50	0,2	0,63	0,2	0,88	0,2	0,75	0,4	0,75	0,1	0,75
Coefficiente de correlação de Pearson	0,76		0,46		0,90		0,81		0,67		0,79		0,89	

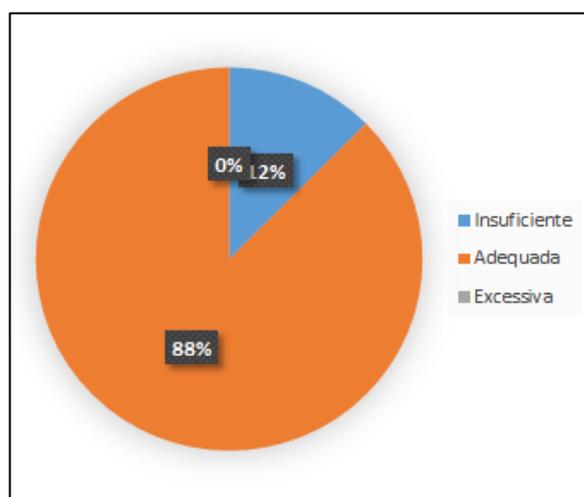
Fonte: Autor.

Como é possível observar, existe relação entre os pesos sugeridos por Couto (2018) e a avaliação dos juízes, o que demonstra evidência de que os pesos são válidos. Foi calculado o coeficiente de correlação de Pearson, que mede a relação estatística entre duas variáveis contínuas. Quanto mais distante de 0 e mais próximo de 1, maior é a correlação entre as variáveis. Na maioria das fases, o cálculo do coeficiente resultou em valores na faixa que indica correlação forte (entre 0,7 e 0,9). A fase 5 se aproximou muito desta mesma faixa (0,67), e a única fase que apresentou uma correlação mais baixa entre os pesos apresentados por Couto e os pesos apontados pelos juízes foi a fase 2 (0,46). Trabalhos futuros podem sugerir ajustes nos pesos das fórmulas de cálculo de escore para tentar aumentar essas correlações.

3.2.1.5 Resultados da avaliação geral do instrumento

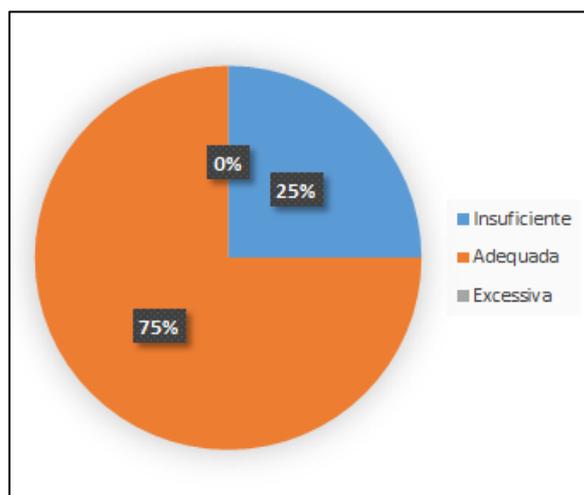
As avaliações dos juízes sobre características gerais do instrumento apontam que o *CT Puzzle Test* é um teste adequado para a avaliação do PC. 88% dos juízes consideraram a quantidade total de fases (7 fases) adequada. Segundo os resultados, 75% consideraram que a quantidade total de itens do teste (19 itens) também é adequada, conforme figuras 30 e 31.

Figura 30. Nível de adequação da quantidade total de fases do *CT Puzzle Test*



Fonte: Autor.

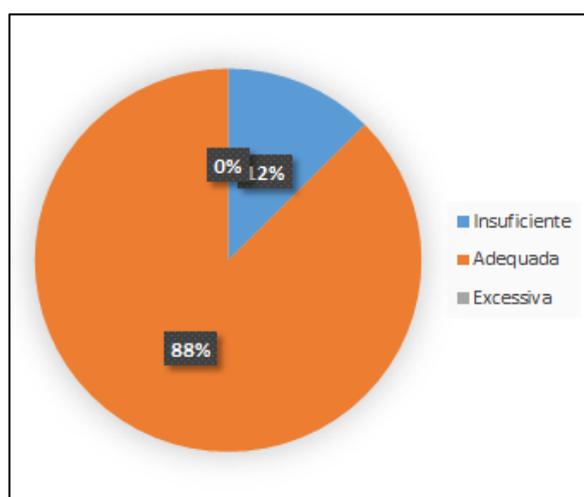
Figura 31. Nível de adequação da quantidade total de itens do *CT Puzzle Test*



Fonte: Autor.

O tempo de duração prevista para a conclusão do teste (45 minutos) também é apropriado segundo 88% dos juízes, conforme a figura 32.

Figura 32. Nível de adequação do tempo previsto para conclusão do *CT Puzzle Test*



Fonte: Autor.

Sobre a capacidade do teste em avaliar o desenvolvimento do PC nos estudantes, 75% dos juízes classificaram o teste como capaz, com avaliações variando entre os valores 3 e 4, conforme aponta a tabela 13.

Tabela 13. Capacidade do *CT Puzzle Test* em avaliar o PC

Capacidade do teste em avaliar o PC (de 0 – incapaz a 5 – totalmente capaz)										
Juíz 1	Juíz 2	Juíz 3	Juíz 4	Juíz 5	Juíz 6	Juíz 7	Juíz 8	Média	Desvio Padrão	IVC
2	4	4	4	2	3	3	4	3,25	0,89	0,75

Fonte: Autor.

Sobre a abordagem do instrumento por meio de *puzzles*, 100% dos juízes apontaram que essa abordagem é altamente adequada, com valores entre 4 e 5 de acordo com a tabela 14.

Tabela 14. Adequação da abordagem por meio de *puzzles*

Adequação da abordagem por meio de <i>puzzles</i> (de 0 – inadequada a 5 – totalmente adequada)										
Juíz 1	Juíz 2	Juíz 3	Juíz 4	Juíz 5	Juíz 6	Juíz 7	Juíz 8	Média	Desvio Padrão	IVC
5	5	4	5	4	4	5	4	4,50	0,53	1,00

Fonte: Autor.

Por fim, sobre a interface gráfica e a linguagem adotada para o público-alvo (jovens de 14 a 17 anos), 75% dos juízes indicaram um nível de adequação alto (valores entre 3 e 5), conforme apresentado na tabela 15.

Tabela 15. Adequação da interface gráfica e da linguagem adotada no teste

Adequação da interface gráfica e da linguagem adotada no teste (de 0 – inadequada a 5 – totalmente adequada)										
Juíz 1	Juíz 2	Juíz 3	Juíz 4	Juíz 5	Juíz 6	Juíz 7	Juíz 8	Média	Desvio Padrão	IVC
3	3	2	3	0	5	3	4	2,88	1,46	0,75

Fonte: Autor.

Nas perguntas abertas, várias observações foram registradas, principalmente com relação a detalhes de interface e usabilidade, e de instruções ao estudante. Todas essas sugestões serão abordadas em trabalhos futuros.

3.2.2 Análise Semântica

Com o objetivo de verificar se o instrumento de pesquisa foi elaborado de maneira compreensível para a população-alvo, foram convidados estudantes com idade entre 14 e 17 anos para participar da análise semântica. Destaca-se que as condições de aplicação sofreram influência

das restrições impostas pelas medidas de enfrentamento da pandemia mundial da doença COVID-19. A pandemia teve seu curso iniciado durante o período de realização desta pesquisa e exigiu que as diretrizes de aplicação do teste fossem adaptadas a esse cenário. Portanto, os resultados apresentados são válidos apenas sob esse contexto, e novas aplicações em condições normais devem ser consideradas em trabalhos futuros.

3.2.2.1 Amostra

Para esta etapa de validação, foram convidados estudantes do Instituto Federal Catarinense Campus Brusque, matriculados nos três anos do ensino médio, de dois cursos: Curso Técnico de Informática Integrado ao Ensino Médio e Curso Técnico de Química Integrado ao Ensino Médio. Como o objetivo dessa etapa é a validação aparente, não foram distinguidas amostras por gênero, curso ou outro critério. O convite foi feito via aplicativo de mensagem e via e-mail para todos os estudantes.

Ressalta-se que devido às restrições impostas pelas medidas de combate à pandemia de COVID-19, as escolas encontravam-se em período de atividade remota e os estudantes estavam dispensados de atividades presenciais. Diversos fatores decorrentes dessa situação influenciaram o alcance do pesquisador ao público-alvo, como a adaptação dos estudantes a uma modalidade diferente de ensino de forma emergencial, a tensão pessoal e comunitária que envolve a situação sanitária decorrente de uma pandemia mundial, a dificuldade de contato com todos os estudantes e o prejuízo às diretrizes de aplicação do teste. Sendo assim, de um montante de aproximadamente 400 estudantes convidados, apenas 13 aceitaram fazer o teste. Esse número é satisfatório para a análise semântica, pois seu objetivo é avaliar a compreensão dos itens pela população-alvo, e deve ser aplicada a uma pequena amostra da população. No entanto, o contexto sugere que apenas aqueles(as) estudantes com maior afinidade com o tema abordado pelo teste tenham se engajado a ponto de se voluntariar. Assim, outro potencial prejuízo decorrente das limitações impostas pelas medidas restritivas de combate à pandemia pode ter sido o enviesamento desta amostra.

3.2.2.2 Instrumentos

O convite para participação no teste foi acompanhado do Guia de Instruções apresentado na seção 3.1.2.2. Os estudantes foram convidados a se cadastrar no *website* do teste, utilizar uma chave de acesso criada para esse fim, fazer o teste e responder um questionário composto por três perguntas abertas:

1. Você teve alguma dificuldade com os controles do teste? Algum problema na hora de mexer nos objetos, algo que te atrapalhou?
2. Os tutoriais do teste te ajudaram a entender o que fazer em cada desafio ou você ainda ficou com dúvida em algum item?
3. Agora que já falamos dos controles e dos tutoriais, por favor, conte como foi pra você realizar esse teste de uma maneira geral. Foi fácil ou difícil? Foi chato ou divertido? Se esse teste fosse para te avaliar, você se sentiria mais à vontade com ele do que com um questionário de múltipla escolha?

3.2.2.3 Procedimentos

No guia fornecido aos estudantes, é informado que após se cadastrar no *website* e realizar o teste, seriam apresentadas três perguntas de suma importância, e solicitado que os estudantes as respondessem. Assim, após finalizar o teste, a tela é recarregada e o questionário é apresentado. Ao submeter as respostas, estas são registradas do banco de dados do teste. Em função das restrições impostas pelas medidas de combate à pandemia de COVID-19, os estudantes responderam o teste a partir de suas residências, em horário livre e sem supervisão. Portanto, os resultados só são válidos para este contexto.

3.2.2.4 Resultados

Com relação à primeira pergunta do questionário “Você teve alguma dificuldade com os controles do teste? Algum problema na hora de mexer nos objetos, algo que te atrapalhou?”, 100% dos estudantes informaram que não tiveram problemas com os controles. Dessa forma, a usabilidade do teste se apresenta adequada.

Sobre a segunda pergunta “Os tutoriais do teste te ajudaram a entender o que fazer em cada desafio ou você ainda ficou com dúvida em algum item?”, 84% dos estudantes afirmaram que os tutoriais os ajudaram, foram didáticos e não restaram dúvidas. Cabe registrar que 16% apontaram dificuldades com as instruções da fase de Classificação, que também foram apontadas por uma parcela dos juízes como não sendo totalmente adequadas. Dessa forma, é pertinente a revisão dessas instruções, mesmo que a maioria dos estudantes não tenha tido problemas.

A terceira pergunta era mais aberta e questionava a opinião dos estudantes sobre a dificuldade em realizar o teste, a diversão proporcionada e a sensação sobre a abordagem por *puzzles*. Ao se analisar as respostas, observa-se que 100% dos estudantes avaliaram o teste como “fácil”, 100% julgaram que o teste é “legal/divertido”, e 92% dos estudantes se sentiram mais à vontade de serem avaliados pela abordagem de *puzzles* em comparação a abordagens tradicionais de testes de múltipla escolha.

Sobre os resultados obtidos, salienta-se que as condições de aplicação sofreram influência das restrições impostas pelas medidas de enfrentamento da pandemia mundial da doença COVID-19 que teve seu curso iniciado durante o período de realização desta pesquisa. As adaptações necessárias para a realização das etapas da presente pesquisa e os impactos nos seus resultados são detalhados na seção 3.2.2.1. Portanto, os resultados apresentados são válidos apenas sob esse contexto, e devem ser consideradas novas aplicações em condições normais em trabalhos futuros.

4 CONCLUSÕES

Este trabalho teve como objetivo principal atestar a validade de conteúdo do *CT Puzzle Test* para a avaliação do Pensamento Computacional em estudantes de 14 a 17 anos. Para alcançar este objetivo o trabalho foi dividido em 4 objetivos específicos.

O primeiro objetivo específico “aprimorar os aspectos técnicos do *CT Puzzle Test* para sua disponibilização via *Internet*” foi alcançado por meio da revisão técnica implementada no instrumento e descrita do capítulo 3.1.1.

Para alcançar o segundo objetivo específico “disponibilizar orientações de apoio ao aplicador do teste”, foi elaborado um Guia do Aplicador, contendo todas as informações e instruções necessárias para que qualquer pessoa interessada em utilizar o *CT Puzzle Test* tenha condições de aplicá-lo e de utilizar seus resultados.

O terceiro objetivo “apresentar conteúdo introdutório sobre o teste para o estudante” foi atingido com a elaboração e disponibilização do Guia de Instruções, que orienta os estudantes sobre como se cadastrar no *website*, utilizar o sistema e responder os itens do teste.

O quarto objetivo específico “Submeter o *CT Puzzle Test* a etapas de validação de conteúdo” foi atingido com a avaliação do teste por especialistas, da qual os resultados são apresentados no capítulo 3.2.1, e pela análise semântica com a aplicação de um questionário a um grupo de estudantes que pertencem ao público-alvo do teste, da qual os resultados podem ser verificados no capítulo 3.2.2.

Além de alcançar os objetivos específicos, este trabalho busca responder duas questões de pesquisa baseadas nas hipóteses formuladas. A seguir os resultados alcançados para cada teste de hipótese são descritos:

- H1: “Juízes concordam que a abordagem por puzzles é válida e pertinente na avaliação do PC”. De fato, 100% dos juízes atestaram que a abordagem por *puzzles* do *CT Puzzle Test* é altamente adequada para a avaliação do PC em estudantes.
- H2: “Os estudantes testados preferem a abordagem por *puzzles* às abordagens tradicionais de teste de perguntas e respostas”. A análise semântica conduzida durante a pesquisa apontou que 92% dos estudantes afirmaram que preferem ou se sentem

mais à vontade com essa abordagem em comparação a testes com perguntas objetivas de múltipla escolha.

Dessa forma, a conclusão sobre a QP1 “A abordagem por meio de *puzzles* na construção dos itens de um instrumento de avaliação do PC é válida e pertinente” resulta positiva.

Além das duas hipóteses iniciais, foram formuladas outras, buscando investigar a validade de conteúdo do instrumento. Os resultados esperados para cada teste dessas hipóteses são descritos a seguir:

- H3: “Juízes concordam que é alto o nível da capacidade do *CT Puzzle Test* em avaliar o PC em jovens de 14 a 17 anos”: 75% dos juízes classificaram o teste como capaz de avaliar o PC em estudantes.
- H4: “Juízes concordam que é alto o nível de aderência dos itens do *CT Puzzle Test* aos quatro pilares do PC.”: a avaliação dos juízes indicou, fase a fase do teste, que todos os itens que o compõe estão baseados e aderentes aos pilares do PC. Cada fase possui características próprias, e adere à pilares distintos em cada caso.
- H5: “Os estudantes conseguem acessar e navegar pelo *CT Puzzle Test*, ler e compreender as instruções e exemplos e registrar as respostas correspondentes.”: a análise semântica apontou que todos os estudantes foram capazes de navegar pelo teste, compreender suas instruções e atividades e responder seus itens, sem enfrentar problemas.

Resta respondida, portanto, a QP2: “O *CT Puzzle Test* possui validade de conteúdo, sendo seus itens relevantes e representativos para avaliar o PC?”. Com base na análise de juízes e na análise semântica, é possível afirmar que a validade de construto do *CT Puzzle Test* está comprovada.

4.1 CONTRIBUIÇÕES DA DISSERTAÇÃO

O principal resultado desta pesquisa é contribuir para a evolução do *CT Puzzle Test* e atestar a validade do teste em avaliar o PC de um indivíduo e em utilizar a abordagem por *puzzles*. Um teste de avaliação do PC formalmente validado representa uma contribuição à rede nomológica do PC, colabora com a definição operacional do construto e favorece o ingresso do PC a currículos escolares ao promover componentes e métodos de avaliação adequados para avaliar os estudantes. O resultado

deste trabalho atesta a validade de conteúdo do *CT Puzzle Test* como instrumento de avaliação do PC que aborda o construto por meio de uma abordagem mais abrangente que outros instrumentos conhecidos. Além disso, por meio da disponibilização de material de apoio e sistema de gestão de aplicações, este trabalho permite a disponibilização do teste a qualquer interessado em investigar o desenvolvimento do PC em indivíduos na sua área de atuação.

4.1.1 Publicações em Eventos

Durante o desenvolvimento deste trabalho, 1 artigo foi publicado e outro submetido para publicação, ainda sem confirmação até o fechamento deste texto. O primeiro se trata da apresentação da proposta do *CT Puzzle Test*, publicado no CBIE em 2020 (RAABE et al 2020). O segundo artigo apresenta as contribuições da evolução do *CT Puzzle Test* para a pesquisa do PC.

4.2 TRABALHOS FUTUROS

A continuidade do processo de evolução do instrumento prevê novas etapas de validação das modificações que o instrumento sofreu em suas últimas etapas de aprimoramento com novas amostras de estudantes. O *CT Puzzle Test* será re-submetido a técnicas de investigação de validade de critério e de construto.

Atualmente, o *CT Puzzle Test* é composto por fases de implementação fixa. Por isso, variações na forma de avaliar são possíveis apenas com a alteração no código fonte. Essa solução atende o requisito de um teste estável para avaliação do PC. Mas para situações em que é necessário variar os níveis, adicionar ou remover fases, a solução atual apresenta pouca flexibilidade. Por isso, como extensão do *CT Puzzle Test*, sugere-se ser desenvolvida uma plataforma de avaliação do PC. Através dela, pode ser possível instanciar testes contendo as mesmas fases (ou mecânicas) adotadas. A vantagem de tornar a ferramenta extensível é que outras abordagens de *puzzles* podem ser criadas para avaliar os mesmos pilares do pensamento computacional, e os aplicadores poderão configurar a dificuldade e o dinamismo das fases.

Por último, os pesos aplicados aos cálculos de score também devem ser revisitados buscando-se uma maior correlação com os pesos apontados pelos juízes.

REFERÊNCIAS

- ALEXANDRE, N. M. C.; COLUCI, M. Z. O. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. *Ciência & Saúde Coletiva*, v. 16, n. 7, p. 3061–3068, jul. 2011.
- BALANSKAT, A.; ENGELHARDT, K. Computing our Future. 2014. DOI 10.13140/RG.2.1.5029.9048. Disponível em: <http://rgdoi.net/10.13140/RG.2.1.5029.9048>.
- BALL, M. A.; GARCIA, D. D. Autograding and Feedback for Snap! In: THE 47TH ACM TECHNICAL SYMPOSIUM, 2016. Anais [...]. [s.l.]: ACM Press, 2016.
- Bell, T., Fellows, M. R., and Witten, I. (1998). Computer Science Unplugged: offline activities and games for all ages. *Computer Science Unplugged*.
- BLIKSTEIN, P. O pensamento computacional e a reinvenção do computador na educação. 2008. Disponível em: http://www.blikstein.com/paulo/documents/online/ol_pensamento_computacional.html. Acesso em: 29 setembro 2020.
- BOALER, J. Alternative approaches to teaching, learning and assessing mathematics. *Evaluation and Program Planning*, v. 21, n. 2, p. 129–141, maio 1998.
- BOCCONI, S. et al. Developing computational thinking in compulsory education-implications for policy and practice. Join Research Center (European Commission). Sevilha. 2016.
- BRACKMANN, Christian Puhmann. Desenvolvimento do pensamento computacional através de atividades desplugadas na educação básica. 2017. Tese (Programa de Pós-Graduação em Informática na Educação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul.
- BRACKMANN, C. P. et al. Development of Computational Thinking Skills through Unplugged Activities in Primary School. In: THE 12TH WORKSHOP, 2017. Anais [...]. [s.l.]: ACM Press, 2017.
- BRENNAN, K.; RESNICK, M. New frameworks for studying and assessing the development of computational thinking. In: Proceedings of the 2012 annual meeting of the American Educational Research Association, Vancouver, Canada. 2012. p. 1-25.
- BUFFUM, P. S. et al. A Practical Guide to Developing and Validating Computer Science Knowledge Assessments with Application to Middle School. In: THE 46TH ACM TECHNICAL SYMPOSIUM, 2015. Anais [...]. [s.l.]: ACM Press, 2015.
- CARTELLI, A.; DAGIENE, V.; FUTSCHEK, G. Bebras Contest and Digital Competence Assessment. Current Trends and Future Practices for Digital Literacy and Competence. [s.l.]: IGI Global, 2012. p. 35–46. DOI 10.4018/978-1-4666-0903-7.ch004. Disponível em: <http://dx.doi.org/10.4018/978-1-4666-0903-7.ch004>
- COLUCI, M. Z. O.; ALEXANDRE, N. M. C.; MILANI, D. Construção de instrumentos de medida na área da saúde. *Ciência & Saúde Coletiva*, v. 20, n. 3, p. 925–936, mar. 2015.

CORTINA, J. M. What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, v. 78, n. 1, p. 98–104, 1993.

COUTO, N. E. R. Impacto da implantação de uma disciplina de pensamento computacional em estudantes. Dissertação (Mestrado em Computação Aplicada) – Universidade do Vale de Itajaí. 2018.

CSTA. K–12 Computer Science Standards. ACM. New York, 2011.

DAGIENE, V.; FUTSCHEK, G. Bebras International Contest on Informatics and Computer Literacy: Criteria for Good Tasks. *Lecture Notes in Computer Science*. [s.l.]: Springer Berlin Heidelberg, 2008. p. 19–30. DOI 10.1007/978-3-540-69924-8_2. Disponível em: http://dx.doi.org/10.1007/978-3-540-69924-8_2

DAGIENE, V.; STUPURIENE, G. Informatics education based on solving attractive tasks through a contest. In: *KEY COMPETENCIES IN INFORMATICS AND ICT*, 2014., 2014, Potsdam. Preliminary Proceedings. Potsdam: University Of Potsdam, 2014. p. 51 - 62.

FALKNER, N.; SOORIAMURTHI, R.; MICHALEWICZ, Z. Puzzle-Based Learning for Engineering and Computer Science. *Computer*, v. 43, n. 4, p. 20–28, abr. 2010.

GARCIA, J. D. R. et al. Developing Computational Thinking at School with Machine Learning: An exploration. In: *2019 INTERNATIONAL SYMPOSIUM ON COMPUTERS IN EDUCATION (SIIE)*, nov. 2019. Anais [...]. [s.l.]: IEEE, nov. 2019.

Guerra, Ana Cervigni Guerra; Colombo, Regina Maria Thienne. *Tecnologia da Informação: Qualidade de Produto de Software*. Brasília: MCT/SEPIN, 2009. 429 p.

GONÇALVES, F. Um instrumento para diagnóstico do pensamento computacional. 2015. 128 f. Dissertação (Mestrado) - Curso de Computação Aplicada, Universidade do Vale do Itajaí, Itajaí, 2015.

GOUWS, L. A.; BRADSHAW, K.; WENTWORTH, P. Computational thinking in educational activities. *Proceedings of the 18th ACM conference on Innovation and technology in computer science education - ITiCSE '13*. Anais... In: *THE 18TH ACM CONFERENCE*. ACM Press, 2013.

GROVER, S. *Robotics and Engineering fo Middle and High School Students to Develop Computational Thinking*. American Educatinal Research Association. New Orleans. 2011.

GROVER, S. Systems of assessments for deeper learning of computational thinking in k-12. In: *Annual meeting of the American Educational Research Association*, Chicago: 2015.

HAIR JUNIOR, J. et al. *Análise multivariada de dados*. 6 ed. Porto Alegre: Bookman. 2009.

HAYNES, Stephen N.; RICHARD, David C. S.; KUBANY, Edward S. Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, vol. 7, nº 3, p. 238–247, set. 1995.

HEALE, R.; TWYLCROSS, A. Validity and reliability in quantitative studies. *Evidence Based Nursing*, v. 18, n. 3, p. 66–67, 15 maio 2015.

Hinkle, D., Wiersma, W., and Jurs, S. (2003). *Applied Statistics for the Behavioral Sciences*. Houghton Mifflin, Boston, MA, 5 edition.

HOGAN, T. *Introdução À Prática de Testes Psicológicos*. Rio de Janeiro: Ltc, 2006.

ISTE, CSTA. Operational definition of computational thinking for K-12 education. International Society for Technology in Education and Computer Science Teachers Association. 2011.

KALELIOGLU, F.; GULBAHAR, Y.; KUKUL, V. A Framework for Computational Thinking Based on a Systematic Research Review. *Baltic Journal Of Modern Computing*. Letônia, p. 583-596. 2016.

KIMBERLIN, C. L.; WINTERSTEIN, A. G. Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, v. 65, n. 23, p. 2276–2284, 1 dez. 2008.

KONG, S.-C.; ABELSON, H.; LAI, M. *Introduction to Computational Thinking Education*. Computational Thinking Education. Springer. Singapura, 2019.

KOTTNER, J. et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Journal of Clinical Epidemiology*, v. 64, n. 1, p. 96–106, jan. 2011.

Kraemer, E. W., Lombardo, S. V., and Lepkowski, F. J. (2007). The librarian, the machine, or a little of both: A comparative study of three information literacy pedagogies at oakland university. *College & Research Libraries*, 68(4):330–342.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*.

LAROS, J. A.; REIS, R. F.; TELLEGEN P. J. Indicações da validade convergente do teste Não-Verbal de Inteligência Son-R 2 1/2 – 7[A]. *Avaliação Psicológica*, Porto Alegre , v.9, n. 1, Abril, 2010.

MAIORANA, F.; GIORDANO, D.; MORELLI, R. Quizly: A live coding assessment platform for App Inventor. In: 2015 IEEE BLOCKS AND BEYOND WORKSHOP (BLOCKS AND BEYOND), out. 2015. *Anais [..]. [s.l.]: IEEE*, out. 2015.

MALHOTRA, N.K. *Pesquisa de Marketing: uma orientação aplicada*. 3. ed. Porto Alegre: Bookman, 2001.

MARTINS, G. Sobre confiabilidade e validade. *Revista Brasileira de Gestão de Negócios*, v. 8, n. 20, p. 1-12, 2006.

MERRICK, K. E. An Empirical Evaluation of Puzzle-Based Learning as an Interest Approach for Teaching Introductory Computer Science. *IEEE Transactions on Education*, v. 53, n. 4, p. 677–680, nov. 2010.

MICHALEWICZ, M. *Puzzle-based learning: an introduction to critical thinking, mathematics, and problem solving*. Melbourne: Hybrid Publishers; 2008

MORENO-LEÓN, J.; ROBLES, G.; ROMÁN-GONZÁLEZ, M. Dr. Scratch: Automatic Analysis of Scratch Projects to Assess and Foster Computational Thinking. *Revista de Educación A Distancia, Espanha*, v. 10, n. 46, p.1-23, set. 2015.

MORENO-LEON, J.; ROBLES, G.; ROMAN-GONZALEZ, M. Comparing computational thinking development assessment scores with software complexity metrics. In: 2016 IEEE GLOBAL ENGINEERING EDUCATION CONFERENCE (EDUCON), abr. 2016. Anais [...]. [s.l.]: IEEE, abr. 2016.

MORENO-LEÓN, J. et al. On the Automatic Assessment of Computational Thinking Skills. In: THE 2017 CHI CONFERENCE EXTENDED ABSTRACTS, 2017. Anais [...]. [s.l.]: ACM Press, 2017.

MORENO-LEON, J.; ROBLES, G.; ROMAN-GONZALEZ, M. Towards Data-Driven Learning Paths to Develop Computational Thinking with Scratch. *IEEE Transactions on Emerging Topics in Computing*, v. 8, n. 1, p. 193–205, jul 2017.

MORENO-LEÓN, J. et al. No es lo mismo: un análisis de red de texto sobre definiciones de pensamiento computacional para estudiar su relación con la programación informática. *Revista Interuniversitaria de Investigación en Tecnología Educativa*, 23 dez. 2019.

MÜHLING, A.; RUF, A.; HUBWIESER, P. Design and First Results of a Psychometric Test for Measuring Basic Programming Abilities. In: THE WORKSHOP IN PRIMARY AND SECONDARY COMPUTING EDUCATION, 2015. Anais [...]. [s.l.]: ACM Press, 2015.

NATIONAL RESEARCH COUNCIL. REPORT OF A WORKSHOP ON THE SCOPE AND NATURE OF COMPUTATIONAL THINKING. [s.l.]: National Academies Press, 2010. Disponível em: <http://dx.doi.org/10.17226/12840>.

NATIONAL RESEARCH COUNCIL. REPORT OF A WORKSHOP ON THE PEDAGOGICAL ASPECTS OF COMPUTATIONAL THINKING. [s.l.]: National Academies Press, 2011. Disponível em: <http://dx.doi.org/10.17226/13170>.

NUNNALLY, J.; BERNSTEIN, I. *Psychometric theory*. 3rd ed. New York: McGraw-Hill, 1994. 752 p.

OTA, G.; MORIMOTO, Y.; KATO, H. Ninja code village for scratch: Function samples/function analyser and automatic assessment of computational thinking concepts. In: 2016 IEEE SYMPOSIUM ON VISUAL LANGUAGES AND HUMAN-CENTRIC COMPUTING (VL/HCC), set. 2016. Anais [...]. [s.l.]: IEEE, set. 2016.

PAPERT, S.; HAREL, I. *Constructionism*. Norwood: Ablex Publishing. 1991.

PAPERT, S. *Teaching children thinking*. MIT Artificial Intelligence Laboratory. Memo no. 2247, Logo Memo no. 2. Cambridge. 1971.

PAPERT, S. *Mindstorms: children, computers and powerful ideas*. Nova York: Basic Books. 1980.

PASQUALI, L. *Psicometria*. *Revista da Escola de Enfermagem da USP*, v. 43, p. 992–999, dez. 2009.

PASQUALI L. *Psicometria: teoria dos testes na psicologia*

e na educação. Rio de Janeiro: Vozes. 2013."

PASQUALI, L. Validade dos Testes. *Examen: Política, Gestão e Avaliação da Educação*, v. 1, n. 1, p. 36, 1 set. 2017.

PELEGRINO, F. M. et al. Cross cultural adaptation and psychometric properties of the Brazilian portuguese version of the duke anticoagulation satisfaction scale. *Journal of Clinical Nursing*, Boston, v. 21, n. 17–18, p. 2509–17, 2012.

POLIT, D.; BECK, C. *Fundamentos de pesquisa em enfermagem: métodos, avaliação e utilização*. 7 ed. Porto Alegre: Artmed. 2011.

POLIT, D. F. Assessing measurement in health: Beyond reliability and validity. *International Journal of Nursing Studies*, v. 52, n. 11, p. 1746–1753, nov. 2015.

PRECKEL, F.; BRUNNER, M. Nomological Nets. In: *Encyclopedia of Personality and Individual Differences*. [S. l.]: Springer International Publishing, 2017.

RAABE, André; VIANA, Cassiano; CALBUSCH, Leonardo. CT Puzzle Test: Em direção a uma avaliação interativa do pensamento computacional. In: *SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO*, 31. , 2020, Online. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2020 . p. 1683-1692.

ROBERTS, P.; PRIEST, H.; TRAYNOR, M. Reliability and validity in research. *Nursing Standard*, v. 20, n. 44, p. 41–45, 12 jul. 2006.

ROBLES, G. et al. On Tools that Support the Development of Computational Thinking Skills: Some Thoughts and Future Vision. In: *INTERNATIONAL CONFERENCE ON COMPUTATIONAL THINKING EDUCATION*, 2018., 2018, Hong Kong. Proceedings [...]. Hong Kong: The Education University Of Hong Kong, 2018. p. 129 - 132.

ROMÁN-GONZÁLEZ, M. *Codigoalfabetización y Pensamiento Computacional en Educación Primaria y Secundaria: Validación de un Instrumento y Evaluación de Programas*. 2016. Tese (Programa de Doctorado en Educación) – Universidad Nacional de Educación a Distancia, Madrid, España.

ROMÁN-GONZÁLEZ, M. et al. Does computational thinking correlate with personality? In: *THE FOURTH INTERNATIONAL CONFERENCE*, 2016. Anais [...]. [s.l.]: ACM Press, 2016.

ROMÁN-GONZÁLEZ, M.; MORENO-LEÓN, J.; ROBLES, G. Complementary Tools for Computational Thinking Assessment. *Conference Proceedings of International Conference on Computational Thinking Education 2017*. Hong Kong: The Education University of Hong Kong, jun. 2017.

ROMÁN-GONZÁLEZ, M.; PÉREZ-GONZÁLEZ, J.-C.; JIMÉNEZ-FERNÁNDEZ, C. Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Computers in Human Behavior*, v. 72, p. 678–691, jul. 2017.

ROMÁN-GONZÁLEZ, M.; MORENO-LEÓN, J.; ROBLES, G. Combining Assessment Tools for a Comprehensive Evaluation of Computational Thinking Interventions. *Computational Thinking Education*. [s.l.]: Springer Singapore, 2019.

ROMÁN-GONZÁLEZ, M. Computational thinking test: Design guidelines and content validation. In: *Proceedings of the 7th Annual International Conference on Education and New Learning Technologies (EDULEARN)*, 2015, Barcelona. *Proceedings.. Barcelona: IATED*, 2015. p. 2436–2444.

ROUSSON, V.; GASSER, T.; SEIFERT, B. Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Statistics in Medicine*, v. 21, n. 22, p. 3431–3446, 2002.

SALMOND, S. S. Evaluating the Reliability and Validity of Measurement Instruments. *Orthopaedic Nursing*, v. 27, n. 1, p. 28–30, jan. 2008.

SEITER, L.; FOREMAN, B. Modeling the learning progressions of computational thinking of primary grade students. In: *THE NINTH ANNUAL INTERNATIONAL ACM CONFERENCE*, 2013, San Diego. *Anais [...]. San Diego: ACM Press*, 2013.

SILVA, E. L.; MENEZES, E. M. *Metodologia da pesquisa e elaboração de dissertação*. 3. ed. Florianópolis: Laboratório de Ensino à Distância da UFSC, 2001.

SMITH, L.; CORDOVA, J. Weighted Primary Trait Analysis for Computer Program Evaluation. *Journal Of Computing Sciences In Colleges*. Evansville, p. 14-19. jun. 2005.

SOMMERVILLE, I. *Engenharia de Software*. 9 ed. Pearson. São Paulo, 2011.

SRIKANT, S.; AGGARWAL, V. Automatic Grading of Computer Programs: A Machine Learning Approach. In: *2013 12TH INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND APPLICATIONS (ICMLA)*, dez. 2013. *Anais [...]. [s.l.]: IEEE*, dez. 2013.

STREINER, D. L.; KOTTNER, J. Recommendations for reporting the results of studies of instrument and scale development and testing. *Journal of Advanced Nursing*, v. 70, n. 9, p. 1970–1979, 30 mar. 2014.

STREINER, D. L. Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment*, v. 80, n. 1, p. 99–103, fev. 2003.

TANNER, K.; ALLEN, D. Approaches to Biology Teaching and Learning: Learning Styles and the Problem of Instructional Selection—Engaging All Students in Science Courses. *Cell Biology Education*, v. 3, n. 4, p. 197–201, dez. 2004.

WANGENHEIM, C. et al. CodeMaster - Automatic Assessment and Grading of App Inventor and Snap! Programs. *Informatics in Education*, v. 17, n. 1, p. 117–150, 14 abr. 2018.

WEINTROP, D.; WILENSKY, U. Using Commutative Assessments to Compare Conceptual Understanding in Blocks-based and Text-based Programs. In: *THE ELEVENTH ANNUAL INTERNATIONAL CONFERENCE*, 2015. *Anais [...]. [s.l.]: ACM Press*, 2015. DOI 10.1145/2787622.2787721. Disponível em: <http://dx.doi.org/10.1145/2787622.2787721>.

WING, J. M. Computational thinking. *Communications of the ACM*, v. 49, n. 3, p. 33, 1 mar. 2006.

WING, J. M. Research notebook: Computational thinking - What and why? *The link magazine*, Carnegie Mellon University, Pittsburgh. 2011.

ZUR-BARGURY, I.; PÂRV, B.; LANZBERG, D. A nationwide exam as a tool for improving a new curriculum. In: *THE 18TH ACM CONFERENCE*, 2013. Anais [...]. [s.l.]: ACM Press, 2013. DOI 10.1145/2462476.2462479. Disponível em: <http://dx.doi.org/10.1145/2462476.2462479>.