



## DESENVOLVIMENTO E APLICAÇÃO DE UM MODELO DE AVALIAÇÃO DE JOGOS EDUCACIONAIS DIGITAIS

Alexandre Vaz

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientador: Geraldo Xexéo

Rio de Janeiro  
Janeiro de 2025

DESENVOLVIMENTO E APLICAÇÃO DE UM MODELO DE AVALIAÇÃO DE  
JOGOS EDUCACIONAIS DIGITAIS

Alexandre Vaz

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO  
ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE  
ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO  
GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E  
COMPUTAÇÃO.

Orientador: Geraldo Xexéo

Aprovada por: Prof. Jano Moreira de Souza  
Prof. Tadeu Moreira de Classe

RIO DE JANEIRO, RJ – BRASIL  
JANEIRO DE 2025

Vaz, Alexandre

Desenvolvimento e Aplicação de um Modelo de Avaliação de Jogos Educacionais Digitais/Alexandre Vaz.  
– Rio de Janeiro: UFRJ/COPPE, 2025.

XIX, 181 p.: il.; 29, 7cm.

Orientador: Geraldo Xexéo

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2025.

Referências Bibliográficas: p. 109 – 121.

1. Avaliação de jogos.
  2. Análise fatorial.
  3. Análise consensual fuzzy.
- I. Xexéo, Geraldo.  
II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

*Dedico essa dissertação ao meu  
avô, Murillo Vaz.*

# Agradecimentos

Gostaria de agradecer principalmente a minha família: Christiano, Grace e Murilo. Sou muito grato pelo apoio incondicional que recebo de vocês. Agradeço também a Sofia, que tanto me motivou nessa reta final do mestrado e sempre me trata com muito amor e companheirismo.

Agradeço ao meu orientador, Professor Geraldo Xexéo, por ter me guiado nesta pesquisa e pelos ensinamentos desde a disciplina que cursei antes mesmo de me matricular no mestrado. Sou grato também aos especialistas que validaram o conteúdo do formulário, aos participantes da avaliação e aos responsáveis pelos jogos avaliados nessa pesquisa.

Esse estudo teve apoio da Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Código de Financiamento 001.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

## DESENVOLVIMENTO E APLICAÇÃO DE UM MODELO DE AVALIAÇÃO DE JOGOS EDUCACIONAIS DIGITAIS

Alexandre Vaz

Janeiro/2025

Orientador: Geraldo Xexéo

Programa: Engenharia de Sistemas e Computação

Jogos educacionais de boa qualidade oferecem diversas ferramentas para aprimorar o aprendizado do aluno. Entretanto, há o risco de o jogo não motivar o aluno ao aprendizado satisfatoriamente. A motivação ao aprendizado de um aluno tem influência sobre o seu desempenho acadêmico. Diversos modelos consolidados avaliam o aspecto motivacional de jogos educacionais digitais, porém costumam ser acompanhados de formulários longos e com um contexto de avaliação mais amplo que só a motivação ao aprendizado. Portanto, apresenta-se nesta dissertação um modelo de avaliação de motivação de alunos ao aprendizado via jogos educacionais digitais. Em complemento ao modelo, foi elaborado um formulário para a medição da motivação do aluno. A fim de assegurar a efetividade do formulário, foram avaliados a confiabilidade e as validades de conteúdo e de construtos. A avaliação da validade de conteúdo utilizou um método original a essa pesquisa, que aplica conceitos de tomada de decisão e análise de consenso da lógica fuzzy.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

## DEVELOPMENT AND APPLICATION OF AN EDUCATIONAL DIGITAL GAME EVALUATION MODEL

Alexandre Vaz

January/2025

Advisor: Geraldo Xexéo

Department: Systems Engineering and Computer Science

High quality digital educational games are excellent tools to enhance a student's learning experience. However, it is possible that an educational game does not sufficiently motivate the student towards learning. A student's motivation to learn is a predictor of their academic achievements. Several well-established models evaluate digital educational games' motivational capabilities, however, their questionnaires are lengthy and cover a broader context than the motivation to learn. Therefore, this dissertation presents a model for evaluating motivation to learn through digital educational games. Additionally, a questionnaire was developed to measure the model's criteria. To ensure the questionnaire's effectiveness, reliability, content validity, and construct validity were assessed. The content validity was evaluated by a method originally applied in this thesis, applying fuzzy logic decision-making and consensus analysis concepts.

# Sumário

<b>Lista de Figuras</b>	<b>xi</b>
<b>Lista de Tabelas</b>	<b>xiv</b>
<b>Lista de Símbolos</b>	<b>xvii</b>
<b>Lista de Abreviaturas</b>	<b>xix</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	1
1.2 Como avaliar um Jogo Educacional Digital? . . . . .	3
1.3 Problema . . . . .	5
1.4 Objetivos . . . . .	5
1.5 Organização Trabalho . . . . .	6
<b>2 Avaliação de Jogos</b>	<b>8</b>
2.1 Modelos de Qualidade em Geral e Qualidade de Software . . . . .	9
2.2 Modelos de Avaliação . . . . .	10
<b>3 Conceitos Fundamentais</b>	<b>15</b>
3.1 Avaliação de Jogos Educacionais Digitais . . . . .	15
3.1.1 Motivação ao Aprendizado . . . . .	15
3.1.2 Jogabilidade . . . . .	19
3.2 Fundamentos Matemáticos Abordados . . . . .	21
3.3 Desenvolvimento de Escalas Psicométricas . . . . .	27
3.3.1 Escalas de Medição . . . . .	28
3.4 Avaliação do Instrumento de Medição . . . . .	29
3.4.1 Psicometria . . . . .	29
<b>4 Técnicas e Métodos Utilizados</b>	<b>44</b>
4.1 Desenvolvimento do Modelo . . . . .	44
4.1.1 Design Science Research (DSR) . . . . .	45
4.1.2 Goal-Question-Metric (GQM) . . . . .	45



<b>5</b>	<b>Trabalhos Relacionados</b>	<b>47</b>
5.1	Trabalhos Relacionados . . . . .	47
<b>6</b>	<b>Método Proposto</b>	<b>51</b>
6.1	Desenvolvimento do Modelo . . . . .	52
6.2	Desenvolvimento do Formulário . . . . .	54
6.2.1	Validade . . . . .	56
6.2.2	Estrutura do Formulário de Avaliação . . . . .	58
6.3	Análise de Dados do Desenvolvimento do Formulário . . . . .	58
6.4	Aplicação do Formulário . . . . .	63
6.5	Análise de Dados Após a Aplicação do Formulário . . . . .	64
6.5.1	Carregamento dos Dados . . . . .	65
6.5.2	Tratamento/Pré-processamento dos Dados . . . . .	65
6.5.3	Confiabilidade . . . . .	65
6.5.4	Validade de Construtos . . . . .	66
6.5.5	Comparações entre Grupos Categóricos (Análise de Variância)	72
<b>7</b>	<b>Artefato</b>	<b>74</b>
7.1	Modelo de Avaliação de Jogos Educacionais Digitais Original . . . . .	74
7.2	Modelo de Avaliação de Jogos Educacionais Digitais : Pós-Validação de Conteúdo . . . . .	77
7.3	Modelo de Avaliação de Jogos Educacionais Digitais: Pós-Análise Fa- torial Exploratória . . . . .	79
<b>8</b>	<b>Resultados e Discussões</b>	<b>81</b>
8.1	Resultado . . . . .	81
8.1.1	Resultados Anteriores à Aplicação do Formulário . . . . .	81
8.1.2	Resultados Posteriores à Aplicação do Formulário . . . . .	84
8.2	Discussão . . . . .	98
8.2.1	Validade de Conteúdo . . . . .	98
8.2.2	Confiabilidade . . . . .	101
8.2.3	Análise Fatorial . . . . .	101
8.2.4	Comparação entre Grupos Categóricos . . . . .	103
<b>9</b>	<b>Conclusões</b>	<b>105</b>
9.1	Limitações . . . . .	106
9.2	Trabalhos Futuros . . . . .	107
	<b>Referências Bibliográficas</b>	<b>109</b>
<b>A</b>	<b>Modelo de Avaliação de Jogos Educacionais Digitais</b>	<b>122</b>

A.1	Versão Inicial . . . . .	122
A.2	Área de Avaliação 1: Jogabilidade . . . . .	124
A.3	Área de Avaliação 2: Experiência de Usuário . . . . .	126
A.4	Área de Avaliação 3: Motivação ao Aprendizado . . . . .	127
A.5	Versão Final . . . . .	130
<b>B</b>	<b>Diagramas BPMN</b>	<b>133</b>
B.1	Desenho de Pesquisa de Análises Fatoriais . . . . .	133
B.2	Processo de Desenvolvimento do Modelo e Formulário de Avaliação de Jogos Educacionais Digitais . . . . .	136
<b>C</b>	<b>Itens do Formulário de Avaliação de Jogos Educacionais Digitais</b>	<b>146</b>
C.1	Versão anterior à primeira rodada de Validade de Conteúdo . . . . .	146
C.1.1	Jogabilidade . . . . .	146
C.1.2	Experiência de Usuário . . . . .	147
C.1.3	Motivação ao Aprendizado . . . . .	147
<b>D</b>	<b>Formulário do Modelo de Avaliação de Jogos Educacionais Digitais</b>	<b>148</b>
D.1	Formulários de Validade de Conteúdo . . . . .	148
D.1.1	Respostas . . . . .	149
D.1.2	Área de Avaliação II: Experiência de Usuário (Gráficos) . . . . .	154
D.1.3	Área de Avaliação III: Motivação ao Aprendizado (Gráficos) . . . . .	162
<b>E</b>	<b>Dados Demográficos</b>	<b>167</b>
E.1	Resultados do Teste Kruskal-Wallis Aplicado aos Dados Demográficos	170
<b>F</b>	<b>Hábitos de Jogo dos Respondentes</b>	<b>173</b>
F.1	Resultados do Teste Kruskal-Wallis Aplicado aos Hábitos de Jogo . . . . .	175
<b>G</b>	<b>Análise Fatorial</b>	<b>177</b>
G.1	Análise Fatorial Exploratória . . . . .	177
G.1.1	Overview . . . . .	177
G.1.2	Primeira Rodada . . . . .	177
G.1.3	Segunda Rodada . . . . .	178
G.1.4	Terceira Rodada . . . . .	179
G.1.5	Quarta Rodada . . . . .	181

# Lista de Figuras

1.1	Benefícios da avaliação de jogos sérios para diferentes stakeholders, diagrama adaptado de EMMERICH e BOCKHOLT (2016). . . . .	3
3.1	Diagrama geral de pesquisa de análises fatoriais, adaptado de HAIR <i>et al.</i> (2009) . . . . .	34
6.1	Processo iterativo do desenvolvimento do modelo de avaliação de jogos educacionais digitais. . . . .	52
6.2	Exemplo do formulário de validade de conteúdo, usado para calcular o IVC) . . . . .	57
6.3	Exemplo da avaliação de concisão e clareza dos itens. . . . .	61
6.4	Processo de aplicação do formulário de avaliação de jogos educacionais digitais. . . . .	64
6.5	Subprocesso da análise de adequabilidade dos dados da avaliação a uma análise fatorial. . . . .	67
7.1	Modelo de Reação a Jogos Educacionais Digitais . . . . .	77
7.2	Modelo de Reação a Jogos Educacionais Digitais, exibindo apenas os critérios que permaneceram após a validação de conteúdo. . . . .	78
7.3	Modelo de Avaliação de Jogos Educacionais Digitais original, com destaque aos critérios que permaneceram após a validação de conteúdo. . . . .	79
7.4	Modelo de Avaliação da Motivação ao Aprendizado via Jogos Educacionais Digitais. . . . .	80
8.1	Gráfico de barras dos autovalores da primeira rodada de análise fatorial . . . . .	91
8.2	Visualização da análise paralela, referente à primeira rodada . . . . .	93
8.3	Scatterplot das cargas fatoriais, resultante da primeira rodada de análise de fatorial exploratória, assumindo uma matriz com três fatores. . . . .	93
8.4	Scatterplot dos autovalores da primeira rodada de análise fatorial exploratória, assumindo dois fatores. . . . .	94

8.5	Gráfico de barras dos autovalores da segunda rodada de análise fatorial exploratória, assumindo uma matriz fatorial de três fatores na primeira. . . . .	96
8.6	Visualização da segunda rodada de análise paralela, assumindo três fatores. . . . .	97
A.1	Modelo de Reação a Jogos Educacionais Digitais . . . . .	123
A.2	Modelo de Reação a Jogos Educacionais Digitais, exibindo apenas os critérios que permaneceram após a validação de conteúdo. . . . .	130
A.3	Modelo de Avaliação da Motivação ao Aprendizado via Jogos Educacionais Digitais. . . . .	131
A.4	Modelo do processo <i>Design Science Research Methodology</i> , adaptado de PEFFERS <i>et al.</i> (2007). . . . .	132
B.1	<i>Overview</i> do desenho de pesquisa de análises fatoriais, adaptado de HAIR <i>et al.</i> (2009). . . . .	133
B.2	Parte inicial do desenho de pesquisa de análises fatoriais, adaptado de HAIR <i>et al.</i> (2009). . . . .	134
B.3	Parte final do desenho de pesquisa de análises fatoriais, adaptado de HAIR <i>et al.</i> (2009). . . . .	135
B.4	<i>Overview</i> do processo de desenvolvimento do modelo e formulário de avaliação de jogos educacionais digitais. . . . .	136
B.5	Processo iterativo do desenvolvimento do modelo de avaliação de jogos educacionais digitais. . . . .	137
B.6	Processo do desenvolvimento do formulário de avaliação de jogos educacionais digitais. . . . .	138
B.7	Processo de análise de dados anterior à aplicação do formulário de avaliação de jogos educacionais digitais. . . . .	139
B.8	Processo de aplicação do formulário de avaliação de jogos educacionais digitais. . . . .	140
B.9	Processo de análise de dados posterior à aplicação do formulário de avaliação de jogos educacionais digitais. . . . .	141
B.10	Subprocesso da análise de adequabilidade dos dados da avaliação a uma análise fatorial. . . . .	142
B.11	Diagrama da análise fatorial confirmatória. . . . .	143
B.12	Diagrama da análise fatorial exploratória. . . . .	144
B.13	Diagrama do subprocesso de comparação entre grupos categóricos. . . . .	145
D.1	Carta de apresentação do formulário de validade de conteúdo para a <i>Motivação ao Aprendizado</i> . . . . .	148

D.2	Distribuição das respostas de validade de conteúdo na área de avaliação de <i>Jogabilidade</i> . . . . .	149
D.3	Índice de Validade de Conteúdo da área de avaliação de <i>Jogabilidade</i> . . . . .	150
D.4	Grau de Certeza para a área de avaliação de <i>jogabilidade</i> . . . . .	151
D.5	Distância ao Consenso para a área de avaliação de <i>jogabilidade</i> . . . . .	152
D.6	Ambiguidade dos itens da área de avaliação de <i>jogabilidade</i> . . . . .	153
D.7	Distribuição das respostas de validade de conteúdo na área de <i>Experiência de Usuário</i> . . . . .	154
D.8	Índice de Validade de Conteúdo da área de avaliação de <i>Experiência de Usuário</i> . . . . .	155
D.9	Grau de Certeza para a Área de Avaliação de Experiência de Usuário . . . . .	156
D.10	Ambiguidade dos itens da área de avaliação de <i>Experiência de Usuário</i> . . . . .	157
D.11	Distribuição das respostas de validade de conteúdo na área de avaliação de <i>Motivação ao Aprendizado</i> . . . . .	158
D.12	Índice de Validade de Conteúdo da área de avaliação de <i>Motivação ao Aprendizado</i> . . . . .	159
D.13	Grau de Certeza para a área de avaliação de <i>Motivação ao Aprendizado</i> . . . . .	160
D.14	Distância ao Consenso para a área de avaliação de <i>Motivação ao Aprendizado</i> . . . . .	161
D.15	Ambiguidade dos itens da área de avaliação de <i>Motivação ao Aprendizado</i> . . . . .	162
E.1	Distribuição da área de formação dos respondentes. . . . .	167
E.2	Distribuição da maior titulação acadêmica dos respondentes. . . . .	168
E.3	Distribuição do ano de nascimento dos respondentes. . . . .	168
E.4	Distribuição de faixa etária dos respondentes. . . . .	169
E.5	Distribuição da renda mensal familiar dos respondentes. . . . .	169
E.6	Distribuição do gênero dos respondentes. . . . .	169
F.1	Distribuição da frequência que os respondentes costumam jogar. . . . .	173
F.2	Distribuição da resposta ao questionamento se jogos educacionais são divertidos. . . . .	174
F.3	Distribuição de respondentes que já haviam jogado jogos educacionais. . . . .	174
F.4	Gráfico de correlação das plataformas na qual os respondentes costumam jogar. . . . .	175
G.1	Visão geral das rodadas de análise fatorial exploratória. . . . .	180

# Lista de Tabelas

3.1	Classificação dos valores da métrica Kaiser-Meyer-Olkin de acordo com KAISER e RICE (1974). . . . .	36
5.1	Comparação entre esta pesquisa e trabalhos relacionados . . . . .	49
6.1	Valor mínimo aceitável para a validade de conteúdo ( $\alpha = 0,05$ ). Tabela original de LAWSHE (1975) . . . . .	59
6.2	Valor mínimo aceitável para a validade de conteúdo ( $\alpha = 0,05$ ) . . . . .	59
8.1	Índice de Validade de Conteúdo para os critérios de Jogabilidade . . . . .	81
8.2	Índice de validade de conteúdo para os critérios de Experiência de Usuário . . . . .	82
8.3	Índice de Validade de Conteúdo para os critérios de Motivação ao Aprendizado . . . . .	82
8.4	Índice de validade de conteúdo para cada área de avaliação . . . . .	83
8.5	Conjunto de itens cujo consenso foi de que o item é essencial . . . . .	83
8.6	Dados estatísticos descritivos das respostas ao formulário de avaliação da reação a jogos educacionais digitais. . . . .	85
8.7	Alfas de Cronbach caso determinados itens fossem descartados. . . . .	86
8.8	Resultados de confiabilidade do instrumento de medição desta pesquisa. . . . .	86
8.9	Métrica Kaiser-Meyer-Olkin dos itens do formulário de avaliação. . . . .	87
8.10	Resultado da análise fatorial confirmatória. . . . .	88
8.11	Métricas de adequabilidade dos dados ao modelo teórico ( <i>goodness-of-fit</i> ). . . . .	88
8.12	Teste qui-quadrado das estimativas da análise fatorial confirmatória. . . . .	89
8.13	Métricas de adequabilidade do modelo de medição. . . . .	89
8.14	Métricas de adequabilidade dos dados ao modelo estrutural ( <i>structural model fit</i> ). (Graus de liberdade = 64) . . . . .	90
8.15	Resultados da análise paralela para a seleção de Fatores . . . . .	91
8.16	Matriz fatorial da pesquisa no cenário de três fatores selecionados. . . . .	91
8.17	Matriz fatorial da pesquisa no cenário de dois fatores selecionados. . . . .	92

8.18	Cargas fatoriais significativas para os fatores selecionados pela análise paralela na segunda rodada, no cenário no qual a primeira análise teve três fatores selecionados. . . . .	95
8.19	Matriz fatorial escolhida como a ideal nesta pesquisa, rotulada e interpretada. . . . .	95
8.21	Identificação de cargas fatoriais estatisticamente relevantes por tamanho de amostra (DIXON, 1992) (HAIR <i>et al.</i> , 2009) . . . . .	103
A.1	Composição do Critério de Diversão . . . . .	124
A.2	Composição do Critério de Comunicação . . . . .	124
A.3	Composição do Critério de Interações com o Sistema . . . . .	125
A.4	Composição do Critério de Escolha . . . . .	125
A.5	Composição do Critério de Fluxo . . . . .	125
A.6	Composição do Critério de Presença . . . . .	126
A.7	Composição do Critério de Usabilidade . . . . .	126
A.8	Composição do Critério de Acessibilidade . . . . .	127
A.9	Composição do Critério de Proteção de Erros . . . . .	127
A.10	Composição do Critério de Atenção . . . . .	127
A.11	Composição do Critério de Relevância . . . . .	128
A.12	Composição do Critério de Confiança . . . . .	128
A.13	Composição do Critério de Satisfação . . . . .	128
A.14	Composição do Critério de Curiosidade . . . . .	129
A.15	Composição do Critério de Desafio . . . . .	129
C.1	Itens da Área de Avaliação de Jogabilidade . . . . .	146
C.2	Itens de Avaliação da Experiência do Usuário . . . . .	147
C.3	Itens de Avaliação de Motivação ao Aprendizado . . . . .	147
D.1	Grau de certeza médio do consenso sobre a ambiguidade ou prolixidade na área de jogabilidade . . . . .	163
D.2	Distância ao consenso sobre a validade de conteúdo na área de jogabilidade . . . . .	163
D.3	Grau de certeza médio do consenso sobre a ambiguidade ou prolixidade na área de experiência de usuário . . . . .	164
D.4	Distância ao consenso sobre a validade de conteúdo na área de experiência de usuário . . . . .	164
D.5	Grau de certeza médio do consenso sobre a ambiguidade ou prolixidade na área de motivação ao aprendizado. . . . .	165
D.6	Distância ao consenso sobre a validade de conteúdo na área de motivação ao aprendizado . . . . .	166

E.1	Resultado do teste Kruskal-Wallis para o dado de Área de Formação (graus de liberdade = 1) . . . . .	170
E.2	Resultado do teste Kruskal-Wallis para o dado de Faixa Etária (graus de liberdade = 2) . . . . .	170
E.3	Resultado do teste Kruskal-Wallis para o dado de Gênero (graus de liberdade = 1) . . . . .	171
E.4	Resultado do teste Kruskal-Wallis para o dado de maior titulação acadêmica (graus de liberdade = 3) . . . . .	171
E.5	Resultado do teste Kruskal-Wallis para o dado de raça/cor (graus de liberdade = 1) . . . . .	171
F.1	Resultado do teste Kruskal-Wallis para a pergunta sobre jogos educacionais serem divertidos (graus de liberdade = 2) . . . . .	175
F.2	Resultado do teste Kruskal-Wallis para a pergunta se o respondente já jogou um jogo educacional (graus de liberdade = 1) . . . . .	176
G.1	Matriz fatorial da pesquisa no cenário de dois fatores selecionados. . . . .	177
G.2	Matriz fatorial da pesquisa no cenário de três fatores selecionados. . . . .	177
G.3	Matriz fatorial da pesquisa no cenário de dois fatores selecionados, decorrente de uma segunda rodada de análise fatorial exploratória no qual a primeira também teve dois fatores selecionados. . . . .	178
G.4	Matriz fatorial da pesquisa no cenário de dois fatores selecionados, decorrente de uma segunda rodada de análise fatorial exploratória no qual a primeira rodada teve três fatores selecionados. . . . .	178
G.5	Cargas fatoriais significativas para os fatores selecionados pela análise paralela na segunda rodada, no cenário no qual a primeira análise teve três fatores selecionados. . . . .	178
G.6	Matriz fatorial escolhida como a ideal nesta pesquisa, rotulada e interpretada. . . . .	179
G.7	Matriz fatorial da pesquisa em um cenário de dois fatores selecionados, no qual uma matriz de dois fatores foi selecionado também na rodada inicial de análise fatorial. . . . .	179
G.8	Matriz fatorial da pesquisa em um cenário de dois fatores selecionados, no qual uma matriz de três fatores foi selecionada na rodada inicial de análise fatorial. . . . .	179
G.9	Matriz fatorial decorrente de uma quarta rodada de análise fatorial. . . . .	181



# Lista de Símbolos

$E_i$	Valores esperados para cada observação no cálculo da estatística qui-quadrado,.....	24
$M_1$	Matriz de consenso fuzzy que representa a completa ambivalência .....	23
$O_i$	Valores observados no cálculo da estatística do qui-quadrado.	24
$\Sigma$	Notação de somatório .....	31
$\alpha$	Alpha de Cronbach.....	31
$\alpha$	Erro estatístico do Tipo 1 .....	23
$\beta$	Erro estatístico do Tipo 2 .....	23
$\chi^2$	Teste qui-quadrado, é um teste de hipótese estatística para determinar se duas variáveis categóricas independem uma à outra.....	27
$a_i$	Total de especialistas que escolheram a alternativa $i$ .....	62
$n$	Número de alternativas.....	62
$n_e$	Número de especialistas que consideram o item essencial à escala.....	33
$r_{i,j}$	Proporção de pessoas que preferem a alternativa $j$ em relação às outras .....	62
$s_x$	Estimativa do erro padrão.....	26
C(R)	Grau de certeza médio .....	22
F(R)	Grau fuzzy médio .....	22

$m(R)$	Métrica de distância ao consenso .....	23
$R$	Matriz de confusão da relação recíproca na qual está sendo calculado o grau de consenso .....	22
$V_i$	Variância do item $i$ .....	31
$V_t$	Variância total da escala .....	31

# Lista de Abreviaturas

BPMN	Business Process Model Notation.....	51
CFA	análise fatorial confirmatória.....	68
DSR	Design Science Research.....	45, 52
EFA	análise fatorial exploratória.....	69
GQM	Goal-Question-Metric.....	53
ISO	International Organization for Standardization.....	9
IVC	Índice de Validade de Conteúdo.....	33
JED	Jogo educacional digital.....	1
KMO	Kaiser-Meyer-Olkin.....	67
MSA	measurement of sampling adequacy.....	67
SEM	Structural Equation Modeling, modelagem de equações estruturais.....	40
SQuaRE	Systems and software Quality Requirements and Evaluation	9

# Capítulo 1

## Introdução

### 1.1 Motivação

Nos últimos anos, observou-se um crescimento significativo na preferência por ferramentas de educação à distância em relação aos métodos presenciais e tradicionais. Em 2022, apenas 46,7% dos alunos universitários americanos estavam matriculados exclusivamente no modelo presencial (CARLTON, 2024). Em contraste, dez anos antes, um total de 74,5% dos alunos não estavam inscritos em cursos online em 2012 (CARLTON, 2024). Durante este período, a pandemia de COVID-19 levou à suspensão das aulas presenciais em diversas regiões (UNESCO, 2022). A solução encontrada, em muitos casos, foi a adoção de formatos de ensino à distância ou híbridos.

Jogos educacionais digitais (JEDs) estão entre os materiais instrucionais adequados ao aprendizado em ambientes virtuais (TOKARIEVA *et al.*, 2019). A introdução de JEDs na sala de aula é acompanhada de uma diversa gama de benefícios. Jogos possuem aspectos intrinsecamente motivacionais (MALONE, 1981; SAVI e ULBRICHT, 2008) e, quando aplicados em contextos educacionais, facilitam o aprendizado e o desenvolvimento de habilidades dos alunos ao exigir tomadas de decisão (SAVI e ULBRICHT, 2008), entre outras qualidades.

GEE (2003) argumenta que jogos educacionais digitais de boa qualidade não apenas tem sua capacidade de aprendizado confirmada por estudos de ciência cognitiva, como possuem ferramentas que aprimoram essa geração de conhecimento. Entre os conceitos que tornam jogos boas ferramentas instrucionais estão: a facilidade em prover informação *on-demand* e dentro do contexto, a dificuldade crescente à medida que o jogador aprimora suas habilidades, e ciclos de resolução de problemas com variações que forcem o jogador a repensar suas estratégias (GEE, 2003) .

No entanto, nem todos os produtos categorizados como *jogos educacionais* têm êxito em cumprir as suas funções de oferecer valor pedagógico. O material pode

falhar como jogo educacional caso não consiga manter o aluno imerso na sua experiência com o jogo ou não ofereça algum desafio a ser solucionado pelo jogador (VITIELLO, 2022). Simplesmente agregar fatos e informações relacionados ao conteúdo sendo ensinado a um *jogo de trilha*, por exemplo, não é o suficiente para motivar um aluno a aprender satisfatoriamente a partir desse formato (VITIELLO, 2022).

Por outro lado, a escolha por materiais como *jogos de trilha* ou *jogos de memória* podem ser bons candidatos para uma aprendizagem via repetição e memorização, desde que, seguindo a taxonomia de aprendizado revisada por KRATHWOHL (2002), o instrutor opte por limitar o aprendizado à obtenção de conhecimento factual através do processo cognitivo de *lembrar*. O aprendizado pode exigir uma profundidade de conhecimento mais elevada, como um conhecimento conceitual ou procedural, ou processos cognitivos mais complexos, como compreender ou aplicar (BLOOM *et al.*, 1956). Estas formas mais complexas de instrução também são possíveis mediante a aplicação de jogos educacionais (WITECK *et al.*, 2021).

Jogos digitais são excelentes materiais educacionais por permitirem a alternância entre aprendizado via texto ou imagem em cenários complexos (MAYER, 1997). Os ensinamentos que mesclam componentes espaço-visuais com elementos verbais-auditivos são categorizados com *aprendizado multimídia*, e segundo MAYER (1997), esse tipo de instrução é efetivo em guiar o processo cognitivo do aluno para o aprendizado e é mais eficaz comparado a materiais puramente verbais.

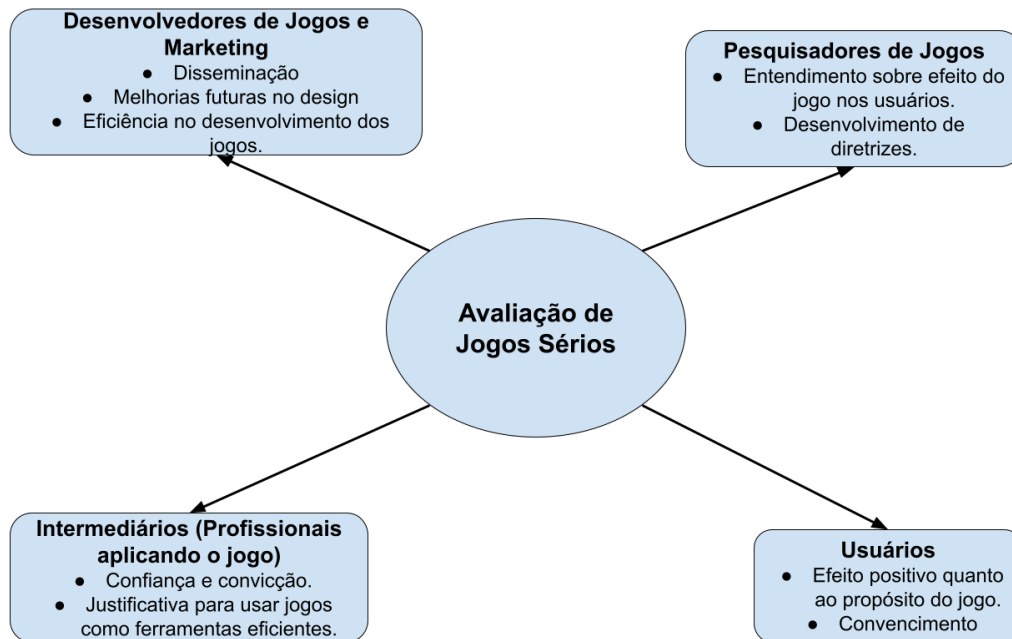
Embora jogos educacionais de boa qualidade sejam excelentes ferramentas didáticas, é importante avaliá-los para atestar a sua efetividade em cumprir seus propósitos (EMMERICH e BOCKHOLT, 2016). Os benefícios da avaliação de jogos sérios para diferentes partes interessadas, segundo EMMERICH e BOCKHOLT (2016), está disposta na Figura 1.1.

Os termos **jogos sérios** e **jogos educacionais** às vezes são confundidos, mas definem categorias diferentes de jogos; o conceito de **jogos sérios** é mais abrangente.

Jogos sérios são aqueles que possuem algum propósito além do entretenimento. O objetivo adicional que torna o jogo sério pode ser pedagógico ou educacional. Portanto, jogos educacionais digitais são um subconjunto de jogos sérios (BREUER e BENTE, 2010).

Por conseguinte, os benefícios expostos na Figura 1.1 aplicam-se também à avaliação de jogos educacionais. Estudos pautados em instrumentos de medição que apresentem resultados confiáveis e válidos, podem ser usados para convencer as partes interessadas, incluindo, mas não se limitando ao próprio educador que está aplicando o jogo, do valor do JED (EMMERICH e BOCKHOLT, 2016) e nortear futuras decisões de design (EMMERICH e BOCKHOLT, 2016), seja do próprio JED ou de produtos futuros.

Figura 1.1: Benefícios da avaliação de jogos sérios para diferentes stakeholders, diagrama adaptado de EMMERICH e BOCKHOLT (2016).



Porém, o aspecto motivacional de um material didático é um pré-requisito de sua capacidade de gerar aprendizado, pois a motivação do aluno influencia no seu desempenho escolar independente do domínio do curso (STEINMAYR e SPINATH, 2009). Outros aspectos como a satisfação do aluno frente à sua experiência com o jogo e a qualidade de componentes como as mecânicas ou interface gráfica, também entram no escopo de avaliação de reação do jogo (EMMERICH e BOCKHOLT, 2016). Diante disso, esta pesquisa elabora um modelo cujo intuito é avaliar a motivação do aluno ao aprendizado mediante a medição destes e outros critérios, expostos no Capítulo 7.1.

## 1.2 Como avaliar um Jogo Educacional Digital?

Jogos digitais são sistemas de software; de fato, estudos de engenharia de software aplicado no contexto de desenvolvimento de jogos abrangem desde comparações do desenvolvimento de jogos digitais com o desenvolvimento de software tradicional (KASURINEN, 2016) até a proposta de métodos, que incluem técnicas de engenharia de software, para o ciclo de desenvolvimento de jogos digitais (ASLAN e BALCI, 2015). A avaliação da qualidade de sistemas de software conta com padrões internacionais, como a família ISO 25000 (ISO/IEC 25000).

Critérios tradicionais da experiência de usuário para sistemas de software em geral, como o de usabilidade, serão contemplados no modelo desenvolvido por esta

pesquisa. Esses padrões indicam como determinar tanto a qualidade do produto quanto a sua qualidade em uso, porém jogos digitais estão entre os casos excepcionais que demandam um conjunto de outros critérios (SÁNCHEZ *et al.*, 2012). Ao avaliar jogos digitais, as divergências de características e objetivos quando comparado com outros tipos de software têm que ser considerados (KORHONEN *et al.*, 2009).

As características inerentes à experiência de usuário em jogos são definidas por especialistas como *playability* (SÁNCHEZ *et al.*, 2012); ou jogabilidade. A jogabilidade se diferencia da UX tradicional, pois jogos digitais são de natureza recreativa, enquanto a maioria dos outros tipos de sistema são projetados para cumprir alguma tarefa desejada pelo usuário.

A natureza dos sistemas de software tradicionais implica na qualidade de sua UX ter como peça central a sua usabilidade, enquanto a medição da jogabilidade é complementada pelos efeitos causados pelas interações com o sistema que gerem diversão e entretenimento. Software de produtividade e jogos digitais tem qualidades contraditórias em certos aspectos, por exemplo, um software de produtividade almeja facilitar a execução de alguma tarefa enquanto o jogo digital propositalmente impõe obstáculos e dificulta a experiência do usuário para agregar à experiência do jogar (KORHONEN *et al.*, 2009).

A jogabilidade é interpretada neste estudo como uma área fundamental para a avaliação de jogos digitais e, conseqüentemente, foi escolhida como uma das áreas de avaliação do modelo. A inclusão de conceitos de experiência de usuário e jogabilidade no modelo se mostra suficiente para estabelecer métricas para a avaliação de um jogo digital, contudo ainda não seriam estabelecidas as métricas para a avaliação do aspecto educacional do jogo.

A jogabilidade pode ser avaliada por diversos métodos e ferramentas. Um dos métodos é o *playtesting*, no qual as mecânicas do jogo são validadas ao jogar um protótipo de um jogo. Assim, são identificadas carências do jogo que podem ser solucionadas com alterações nos componentes ou nas regras do jogo. Entretanto, a avaliação via *playtesting* é direcionada para a perspectiva do sistema de jogo, e não mede as estéticas (HUNICKE *et al.*, 2004) ou efeitos (WINN, 2009) sentidas pelo jogador. As estéticas, ou efeitos, de um jogo são as reações instintivas ou emocionais provocadas pelas dinâmicas do jogo.

O caráter subjetivo e comportamental das reações do jogador se adéqua ao uso de um formulário, que através de seus itens mede indiretamente fenômenos comportamentais dos respondentes. O formulário, ou escala, foi escolhido como instrumento de medição desta pesquisa, mas entrevistas ou a observação dos jogadores durante as partidas também são utilizados para avaliar as reações a jogos educacionais (DE OLIVEIRA, 2022).

## 1.3 Problema

Jogos educacionais digitais possuem características que, quando aplicados corretamente, tornam-se ferramentas didáticas de alta qualidade (GEE, 2003; MAYER, 1997; SAVI e ULBRICHT, 2008; TOKARIEVA *et al.*, 2019). A avaliação de jogos é importante para verificar se estão cumprindo seus propósitos (EMMERICH e BOCKHOLT, 2016). O impacto no aprendizado do aluno é a função primordial, mas é importante avaliar a qualidade de sua experiência com o jogo. A avaliação da reação de alunos aos jogos educacionais digitais conta com alguns modelos consolidados; entre eles, destacam-se o MEEGA+ e o EGameFlow, frequentemente mencionados na literatura.

Entretanto, os formulários desses modelos possuem, respectivamente, 35 e 42 itens, o que pode desestimular respondentes e dificultar a avaliação (KITCHENHAM e PFLEEGER, 2002b). A redução na taxa de resposta implica em um tamanho amostral menor para o estudo. Amostras diminutas oneram a avaliação feita pelo formulário, pois diversas etapas da análise de dados têm um tamanho amostral mínimo aceitável; por exemplo, a análise fatorial (HAIR *et al.*, 2009; MACCALLUM *et al.*, 1999), métricas de confiabilidade (BUJANG *et al.*, 2018).

Esta pesquisa aplica a metodologia de Design Science Research, cujo estágio inicial é a definição de um problema e motivação. Ao longo desta dissertação, houve duas iterações de design e rigor do DSR, cada uma com a intenção de solucionar o problema desta pesquisa mediante a solução de diferentes questões. Deste modo, a questão-problema que esta pesquisa visa solucionar é: há como avaliar jogos educacionais digitais de forma mais concisa do que via os formulários atualmente à disposição?

Os objetivos e questões que buscam solucionar este problema mediante os ciclos DSR estão delineados na seção a seguir.

## 1.4 Objetivos

Esta dissertação usa a metodologia DSR, seguindo as orientações de PEFFERS *et al.* (2007), para o desenvolvimento de um modelo de avaliação de jogos educacionais digitais. Esta seção aborda a segunda atividade citada por PEFFERS *et al.* (2007), de definição dos objetivos para uma solução. Uma explicação aprofundada no capítulo de Método Proposto (Capítulo 6).

Na primeira iteração dos ciclos de design e rigor desta pesquisa, o objetivo é desenvolver um modelo teórico holístico da reação de alunos a jogos educacionais digitais. A premissa é encontrar conceitos fundamentais à avaliação de jogos para identificar critérios que possivelmente não foram cobertos por modelos de avalia-



ção elaborados em trabalhos relacionados. O resultado esperado com o cumprimento deste objetivo é um modelo robusto, cuja estrutura segue a abordagem Goal-Question-Metric (GQM) (SOLINGEN e BERGHOUT, 1999).

O resultado dessa iteração não atenderia à questão-problema desta dissertação, mas fornece à segunda iteração da pesquisa uma base teórica sólida para fundamentar o artefato final desta dissertação. Além disso, a aplicação da abordagem GQM faz com que o modelo não se limite ao campo conceitual e tenha métricas quantificáveis voltadas ao objetivo desta pesquisa.

O objetivo da segunda iteração do ciclo de design e rigor é desenvolver um modelo de avaliação conciso e válido para avaliar jogos educacionais digitais. O modelo de avaliação deve ser acompanhado de um formulário, para medir os fenômenos delimitados pelo modelo.

Para solucionar o problema desta dissertação, é necessário que o formulário de avaliação seja mais diminuto que os formulários que acompanham os modelos MEGA+ e EGameFlow, mas demonstre níveis satisfatórios de confiabilidade e validade enquanto instrumento de medição.

## 1.5 Organização Trabalho

Essa dissertação se inicia com a descrição de conceitos encontrados na literatura. Inicialmente, são apresentadas várias definições sobre o que são jogos. Em seguida, é listada uma gama de modelos de qualidade e de avaliação, cobrindo desde modelos voltados à avaliação de políticas públicas até padrões internacionais para a medição da qualidade de software.

Posteriormente, no capítulo de conceitos fundamentais (Capítulo 3) são definidos atributos abordados pelo modelo de avaliação desta pesquisa. O capítulo possui seções para as diferentes áreas de avaliação que o autor incluiu no modelo desenvolvido nesta pesquisa. Além disso, são discutidos os fundamentos matemáticos abordados na pesquisa e as técnicas de análise de dados feita para validar os artefatos resultantes da pesquisa.

Em sequência, são definidos os métodos e técnicas para o desenvolvimento do artefato e do conhecimento teórico gerado por esta pesquisa, e também para a estruturação do modelo e do formulário de avaliação.

Após a revisão dos conceitos fundamentais, há uma comparação da pesquisa feita nesta dissertação com trabalhos relacionados. Os trabalhos mencionados tratam da avaliação de jogos educacionais digitais e há um foco em trabalhos que também utilizaram instrumentos psicométricos para a avaliação dos jogos. No final da seção, os trabalhos são justapostos em uma tabela que cobre alguns atributos de cada pesquisa.

Em seguida, o método proposto para esta dissertação é detalhado. É descrita a aplicação da metodologia DSR e sua relação com o andamento desta pesquisa. O método cobre as seguintes etapas:

- Desenvolvimento do modelo de avaliação.
- Desenvolvimento do formulário de avaliação.
- Análise de dados do desenvolvimento do formulário de avaliação.
- Aplicação e distribuição do formulário.
- Análise de dados de validação do modelo de avaliação e formulário.

O método proposto por essa pesquisa pode ser visualizado em diagramas BPMN dispostos no Apêndice B.

Após o detalhamento do método proposto, inicia-se a apresentação dos resultados desta pesquisa. O capítulo *Artefato* (Capítulo 7.1) exhibe o modelo e se aprofunda sobre a definição de seus componentes. Uma explicação mais detalhada dos critérios do modelo encontra-se no Apêndice A. Esse capítulo exhibe também a evolução do modelo de avaliação; o modelo inicial é resultado de conceitos encontrados na literatura, enquanto a versão final foi moldada por uma análise exploratória fundamentada nas respostas obtidas por esta pesquisa.

O capítulo em sequência continua a apresentação de resultados. Desta vez, os resultados são aqueles decorrentes da coleta e análise de dados usando as técnicas explicitadas no capítulo de método proposto. Entre os resultados estão métricas de avaliação de validade de conteúdo e de construtos, confiabilidade e de comparação entre grupos categóricos.

Os resultados são examinados na seção *Discussão*. Nesta seção, os resultados impróprios e insatisfatórios são relatados. Nestes casos, nos quais é apresentada uma desconformidade entre os dados observados e aqueles esperados, são feitas críticas que avaliam o impacto dos resultados negativos.

Finalmente, o capítulo de conclusão aborda as contribuições do trabalho, suas limitações e possíveis trabalhos futuros.

# Capítulo 2

## Avaliação de Jogos

Esse capítulo apresenta conceitos importantes encontrados na revisão da literatura. Primeiro, são abordadas definições de jogos provenientes de diversos especialistas de ludologia. Depois, são delineados conceitos de qualidade de software e qualidade da experiência de jogos educacionais. Em seguida, há uma seção que discorre sobre modelos de avaliação, com destaque ao modelo escolhido para esta pesquisa: o modelo Kirkpatrick.

Uma das características inerentes a jogos bem-sucedidos, segundo KOSTER (2013), é uma pluralidade de habilidades necessárias para superar os obstáculos. O domínio dessas habilidades permite variações estratégicas ao chegar em um alto nível (KOSTER, 2013). Outra característica de jogos é os que jogadores estão em busca da identificação de algum padrão mediante os dados oferecidos em cada permutação do jogo (KOSTER, 2013). Estes novos padrões encontram-se sempre no contexto familiar que são as regras fixas do jogo (KOSTER, 2013).

Seguindo essa linha de pensamento, KOSTER (2013) define a diversão como aprender algum padrão diferente em um contexto familiar. Ainda segundo o autor, após a descoberta de alguma estratégia que resulte em uma vitória garantida, um jogo fica entediante e é prontamente abandonado. Jogos que aparentam ser impossíveis de dominar também são descartados por jogadores devido à frustração. Uma dificuldade moderada, na qual o jogador consegue decifrar progressivamente os padrões dos obstáculos do jogo, mantém o jogador engajado por um tempo mais prolongado ao não tender demasiadamente ao tédio ou à frustração.

Outras definições sobre o que é um jogo são propostas por Caillois e Huizinga. Um ponto em comum é a delimitação de um jogo perante o seu ambiente; a restrição de jogos impostos por seu ambiente espaço-temporal e por suas regras é abordado nas definições de jogos de Huizinga (HUIZINGA, 2019) e Caillois (CAILLOIS, 2001).

O ambiente, ou mundo, do jogo é identificado por Huizinga como um mundo temporário no qual as regras do jogo são imprescindíveis (HUIZINGA, 2019). Há também a imersão do jogador na atividade, que o leva a um rompimento da sua

percepção espaço-temporal (HUIZINGA, 2019), como se a sessão do jogo fosse um interlúdio de sua “vida real”. Tal sensação é comparável à distorção na percepção da passagem do tempo e a sensação de envolvimento pertencentes à sensação de fluxo.

Huizinga e Caillois também concordam que jogos são atividades com propósitos intrínsecos, autotélicos. A definição de jogos de CAILLOIS (2001) argumenta que jogos são atividades que não produzem bens ou produtos, mesmo que os participantes possuam algum interesse material relacionado ao jogo. Por exemplo, em jogos de azar ou na profissionalização dos jogadores. Tal definição se difere da definição de Huizinga, que argumenta que jogos não são jogados com interesses materiais ou por lucro.

Jogos digitais, em específico, são também sistemas de software. O notável e reiteradamente crescente poder computacional dos computadores e consoles de jogos possibilita que jogos consigam processar dados em alta quantidade e velocidade. Os jogadores então podem ter alta interatividade com os jogos através dos *feedback loops*. O *feedback* resultante das ações do jogador são usados para a identificação de padrões que melhoram o desempenho no jogo.

## 2.1 Modelos de Qualidade em Geral e Qualidade de Software

Jogos digitais são produtos de software, que por mais que não possuam necessariamente as características de uso tradicional, devem ser de alta qualidade quanto ao produto em si e ao seu uso. A *International Organization for Standardization*, ISO, provê uma série de documentos sobre frameworks, modelos e definições referentes à qualidade, à medição de qualidade e à aplicação de tais conceitos à avaliação da qualidade de sistemas de software.

A família de padrões ISO/IEC 25000 (ISO/IEC 25000) chamada de *Systems and software Quality Requirements and Evaluation* (SQuaRE), apresenta modelos de medição e avaliação de qualidade em uso e a qualidade de produto de um sistema de software. As métricas de qualidade de produto são diretamente relacionadas ao software sob análise. Enquanto as métricas de qualidade em uso dependem das interações do software com o seu usuário e os contextos de utilização do software.

O modelo de qualidade de software utilizado no SQuaRe é originalmente fruto de outro documento ISO, agora abandonado, o ISO/IEC 9126, e hoje é alinhado com a família 25000. Nele, são definidos seis características de qualidade de produto interna e externa a um software. Entre as características, a de maior interesse ao modelo desta dissertação é a de usabilidade. Em relação às quatro características de qualidade em uso mapeadas pelo modelo ISO 25000, destaca-se para esta pesquisa

a de satisfação do usuário Uma grande porção das características de usabilidade e de satisfação, e as sub características que as compõem, foram reaproveitadas para o modelo aqui sendo desenvolvido. O impacto do modelo ISO 25000 no modelo desta dissertação encontra-se de forma mais aprofundada no capítulo *Artefato* (Capítulo 7.1) deste texto.

Outro modelo de qualidade de software, focado na experiência das interações com usuário, é o Qualinet. O Qualinet define a Qualidade de Experiência, no qual o termo *experiência* significa o fluxo de percepções e interpretações de uma sequência de um ou mais eventos (BRUNNSTRÖM *et al.*, 2013). Nesta definição não é necessário haver algum julgamento de sua qualidade nem um vínculo a eventos passados (BRUNNSTRÖM *et al.*, 2013). A experiência no contexto desta pesquisa engloba as interações humano-computador, com o sistema do jogo, e aquelas feitas com outros jogadores. A qualidade de experiência é o prazer ou aborrecimento do usuário ao interagir com algum serviço. e está relacionado com o cumprimento de suas expectativas de utilidade ou satisfação, considerando a personalidade e estado atual da pessoa (BRUNNSTRÖM *et al.*, 2013).

A qualidade de experiência para aprendizado multimídia tem como pilares o fator cognitivo, de percepção do aprendizado do usuário, e afetivos, como o engajamento e a satisfação do usuário durante a sua experiência (BRUNNSTRÖM *et al.*, 2013). Além disso, há de ser oferecida uma experiência inclusiva e acessível para os usuários.

Assim, medidas devem ser tomadas para assegurar que pessoas com deficiência tenham uma experiência, é com esse intuito que REIS (2020) elaborou uma série de recomendações para o *design* de jogos educacionais multiplataforma e inclusivos. Esta série de recomendações, por sua vez, baseia-se nas recomendações de FONTOURA JUNIOR (2018), que elaborou uma recomendação para o *design* de jogos educacionais inclusivos, mas que não contemplavam o advento de jogos em dispositivos móveis.

O modelo inicialmente elaborado nesta pesquisa contempla diversos fatores cognitivos para a motivação ao aprendizado, como desafio, curiosidade e atenção. O modelo também aborda fatores afetivos, abrangendo as estéticas na avaliação de jogabilidade e critérios como satisfação na avaliação de experiência de usuário.

## 2.2 Modelos de Avaliação

Os modelos acima definem o conceito de qualidade, mas por outra perspectiva, deve ser medido o impacto do jogo digital sendo utilizado como ferramenta educacional. Existem diversos modelos de avaliação à disposição, variando nos processos, objetivos e área de avaliação. Nesta seção, estão descritos alguns desses modelos de avaliação. Os modelos são comparados e o modelo escolhido para esta pesquisa,

o Modelo de Kirkpatrick, é descrito de forma mais aprofundada. Alguns modelos de avaliação destacados são voltados para a avaliação de políticas públicas. Outros, têm sua utilidade voltada ao mundo corporativo, abrangendo a avaliação do desempenho da empresa e de programas de treinamento. Visto que esta pesquisa é para a avaliação de materiais educacionais, é feita uma ponderação em relação à utilização de um modelo originalmente criado para programas de treinamento para um contexto acadêmico.

O *Outcome-Based Evaluation* descreve um método de avaliação para políticas (programas) nas áreas de educação, saúde e serviço social (SCHALOCK, 2001). Nele, avaliação é feita individual e organizacionalmente. A avaliação das políticas é feita com as finalidades de refinamento das políticas e de *accountability*<sup>1</sup>, ressaltando que para que um programa disponibilizado por um órgão seja transparente e responsabilizável, ele deve ser avaliável. A avaliação dos programas é feita na ótica de efetividade, impacto e enquanto política pública. A efetividade corresponde aos resultados obtidos comparado com os resultados esperados para determinado indicador, enquanto a medição de impacto é a quantificação da diferença entre a implementação ou não do programa sob análise. A avaliação política usa indicadores de: equidade, medindo se cidadãos em situações parecidas são tratados igualmente; eficiência, medindo o custo-benefício da política; efetividade, descrita anteriormente.

A análise de políticas públicas é um campo um tanto quanto polivalente, permitindo diversas abordagens quantitativas ou qualitativas (DUNN, 2015). Por exemplo, o Public Policy Analysis, uma alternativa ao Outcome-Based Evaluation para políticas públicas, utiliza uma abordagem multi-método para a avaliação (DUNN, 2015).

Os métodos de avaliação de políticas públicas podem seguir um modelo descritivo, comum nas ciências sociais, para descrever as causas e consequências das políticas. Em contraste, existem modelos de uma vertente normativa. Nestes modelos, é comparada a situação observada com a situação considerada ideal para avaliar atributos como eficiência, equidade ou segurança, originários de áreas como economia e filosofia (DUNN, 2015). Os modelos de avaliação de políticas públicas divergem daqueles que avaliam organizações do mundo corporativo devido à diferença nos objetivos almejados e características. Portanto, são especificados dois modelos cuja finalidade original está na avaliação de corporações ou programas corporativos.

Estratégias como programas de treinamento são recomendadas quando a capacitação exigida para os empregados perante a realidade atual da organização é relativamente baixa. A aplicação de programas de treinamento deve ser acompanhada de uma avaliação em relação a sua efetividade. Com esta finalidade, diversos

---

<sup>1</sup>Essa palavra vai além do conceito responsabilidade, a qual é a tradução mais próxima em português, por isso decidimos manter o termo em inglês.

modelos de avaliação de programas de treinamento foram desenvolvidos. Entre eles, destacam-se o Success Case Method e o Modelo Kirkpatrick.

O *Success Case Method* é um modelo de avaliação voltado a programas de treinamento (BRINKERHOFF, 2005). No cenário competitivo de negócios, há a necessidade de investir no treinamento dos funcionários para obter melhores resultados para a empresa. As melhorias nos programas de treinamento não devem ser feitas para torná-las experiências melhores, o foco deve estar no ganho de produtividade dos participantes.

O processo se inicia com a identificação de casos de sucesso interno a uma organização na adoção de uma ferramenta ou método. Os casos de sucesso são identificados por intervenção de um questionário ou relatórios de dados de uso da ferramenta digital. Em seguida, é feita uma entrevista com potenciais casos de sucesso, inicialmente verificando se de fato foi um caso exitoso, para, em seguida, documentar e apurar o que levou ao sucesso.

Esse processo de avaliação foca nos extremos, não é do interesse deste método medidas de tendência central, como um caso *mediano* de algum colaborador (BRINKERHOFF, 2005). Histórias detalhadas dos casos de sucesso e a identificação de fatores vantajosos e onerosos do treinamento nos resultados almejados são os produtos do *success case method* (BRINKERHOFF, 2005).

O SCM surge na tentativa de corrigir defeitos nos modelos estabelecidos para a avaliação de programas de treinamento. Uma das críticas é direcionada ao Modelo Kirkpatrick, ao qual é levantado que a avaliação fica limitada ao treinamento e não a um cenário organizacional mais amplo (BRINKERHOFF, 2005). O Modelo Kirkpatrick foi o selecionado para a avaliação de instrumentos educacionais nesta pesquisa. Segue uma descrição do modelo e na utilização em contextos educacionais.

O modelo elaborado por KIRKPATRICK e CRAIG (1970) separa a avaliação em quatro níveis sequenciais, propondo maneiras de avaliá-los e fornecendo dicas e conselhos para a realização do programa em si. É importante notar que cada nível tem como pré-requisito os níveis anteriores. Segue uma breve descrição de cada nível:

- **Reação:** algo fundamental em qualquer conteúdo didático, é que os alunos participantes tenham uma reação positiva ao material. Quando o participante não se interessa pelo material didático é mais difícil de conseguir e principalmente manter um nível de atenção suficiente para ensiná-lo. Uma reação positiva não garante sucesso no aprendizado, mas demonstra que o participante está mais disposto a aprender.
- **Aprendizado:** no nível de aprendizado, as definições de Kirkpatrick demonstram alguma semelhança com os tipos de aprendizado cognitivos, psicomotores

e afetivos descritos por BLOOM *et al.* (1956). A adaptação de Kirkpatrick é de medir uma potencial melhora no conhecimento (cognitivo), habilidade (psicomotor) ou de atitude (afetivo). Um exemplo de uma melhora de atitude seria de ensinar aos funcionários a importância da diversidade no ambiente de trabalho.

- **Comportamento:** a aplicação e repetição da aprendizagem no ambiente de trabalho ao longo de um período para mudar um comportamento antigo. Além do aprendizado, é importante que o funcionário encontre um ambiente favorável a mudança proposta pelo programa de treinamento e que haja um desejo intrínseco de mudar KIRKPATRICK e KIRKPATRICK (2006).
- **Resultado:** são os impactos a longo-prazo, alinhados com os interesses da organização. No contexto original do modelo Kirkpatrick, voltado a programas de treinamento, seriam resultados de negócio para a empresa aplicando o treinamento. Por exemplo, um aumento na produtividade dos profissionais que passaram pelo treinamento.

Esta pesquisa faz uma avaliação de um instrumento educacional apenas o primeiro nível de avaliação, de reação. Entretanto, a conclusão desta pesquisa aborda possíveis trabalhos futuros para a avaliação dos demais níveis.

Há de ser feita uma ressalva no que se refere ao Modelo Kirkpatrick. A mídia na qual programas de treinamento são oferecidos mudou significativamente desde que o modelo foi proposto em KIRKPATRICK e CRAIG (1970). Os avanços tecnológicos permitiram que as atividades migrassem para um formato digital. A transição para o formato digital reduz os custos para a empresa e torna os programas mais padronizados, convenientes e aplicados em um ritmo mais adaptável a rotina dos empregados STROTHER (2002). Além disso, essa migração faz com que algumas das sugestões quanto ao formato de programas de treinamento sugeridos em *Evaluation of Training* sejam repensados para se adequarem ao padrão de conteúdos online. Entretanto, a estrutura da avaliação dos programas segue válida para programas de treinamento digitais, inclusive no formato específico de jogos digitais que interessa a essa dissertação.

Apesar de ter sido designado para a avaliação de programas de treinamento, o modelo pode ser aplicado em outros contextos. A aplicação do modelo Kirkpatrick não é unânime, HOLTON III (1996) cita que os níveis hierárquicos do modelo são uma taxonomia da avaliação de programas de treinamento, mas que o desenvolvimento de uma taxonomia é um estágio inicial de pesquisa e que estudos experimentais e correlacionais devem ser feitos para comprovar as relações entre as classificações do modelo.



Entretanto, diversos estudos encontrados na literatura fazem um paralelo entre a efetividade de um programa de treinamento e a efetividade de um material educacional em um contexto acadêmico (ARTHUR JR *et al.*, 2003; COLLINS *et al.*, 2006; PRASLOVA, 2010). Por exemplo, PRASLOVA (2010) elabora uma versão adaptada do modelo para a avaliação de materiais educacionais de ensino superior; a adaptação do modelo é acompanhada de um estudo de caso o aplicando. Os impactos em cada nível hierárquico do modelo são mencionados ao descrever as diferenças para o contexto acadêmico. Ao passo que os níveis seguintes possuem mudanças mais substanciais, o nível de reação é mudado para ter como objetivo a avaliação dos alunos sobre o material educacional e autoavaliações sobre os ganhos na aprendizagem que acreditam ter obtido mediante o material (PRASLOVA, 2010).

# Capítulo 3

## Conceitos Fundamentais

Esse capítulo continua a apresentação de conceitos fundamentais ao modelo de avaliação de jogos. Desta vez, há inicialmente uma explicação sobre a base teórica para os critérios incluídos no modelo de avaliação desta pesquisa. Assim, esta seção discute teorias sobre a motivação de alunos ao aprendizado e sobre a qualidade da experiência de usuário no contexto de jogos.

Depois, são explicados fundamentos matemáticos aplicados na análise de dados feita nesta dissertação. O capítulo se encerra com descrições das técnicas usadas na análise de dados, englobando a avaliação da confiabilidade e validade do modelo de avaliação e a comparação entre grupos categóricos de respondentes.

### 3.1 Avaliação de Jogos Educacionais Digitais

Em complemento à experiência de usuário, abordada ao discutir os modelos de qualidade em geral e aplicada a software, a avaliação de jogos educacionais digitais deve passar também por motivação ao aprendizado e experiência do usuário.

#### 3.1.1 Motivação ao Aprendizado

As ferramentas de aprendizado devem ser intrinsecamente motivadoras, estimulando o jogador a jogar múltiplas vezes e, assim, reforçar o aprendizado. O framework de instrução intrinsecamente motivacional foi desenvolvido após uma pesquisa que encontrou fatores relevantes para a motivação intrínseca de jogadores após experimentos utilizando diversos jogos (MALONE, 1981). MALONE (1981) afirma, com um devido embasamento empírico, que curiosidade, desafio e fantasia são conceitos importantes para a motivação via jogos educacionais. A partir das conjecturas estabelecidas, o autor faz uma série de experimentos com jogos diferentes, como *Breakout* e *Darts*, para extrair conhecimento que aprofunde as características motivacionais de curiosidade, desafio e fantasia em materiais instrucionais.

MALONE (1981) diferencia entre a fantasia intrínseca, na qual a temática e as mecânicas do jogo são interdependentes, e a fantasia extrínseca, ocorrente quando a fantasia é apenas uma temática aplicada em cima das mecânicas, que independem do tema. Em contrapartida, fantasias extrínsecas são caracterizadas por jogos cuja camada temática tem uma fraca ligação com as mecânicas do jogo, por exemplo, um jogo de xadrez no qual às peças são personagens de algum universo cinematográfico. Uma hipótese interessante do autor é que a fantasia intrínseca tem valor instrucional maior do que a extrínseca, ocorrente quando a fantasia é apenas uma temática aplicada em cima das mecânicas, que independem do tema.

Em relação à curiosidade, é feita a distinção entre a curiosidade sensorial e a curiosidade cognitiva. A curiosidade sensorial é provocada por estímulos audiovisuais que podem exercer diferentes funções em um jogo. Enquanto a curiosidade cognitiva ocorre quando é apresentado que o conhecimento do aluno sobre algum conceito está incompleto ou inconsistente e mostra ferramentas para retificar tal conhecimento.

Ao discutir o uso de desafio como ferramenta motivacional, o autor faz sugestões para características dos objetivos que guiam os jogos, destaca a importância de ter resultados incertos, lista diferentes estratégias para obter essa natureza incerta dos resultados e aponta que o jogo deve ter uma dificuldade variável, pois uma dificuldade demasiadamente alta reduz a autoestima e, por consequência, a motivação do aluno em jogar.

Outro modelo de motivação ao aprendizado é o modelo ARCS, desenvolvido por KELLER (1987). O ARCS é fundamentado no conhecimento do autor sobre design instrucional e na teoria de macro-motivação.

O propósito do modelo é facilitar o ciclo de design de criação, desenvolvimento e refinamento de materiais educacionais que motivem seus alunos ao aprendizado. O modelo denota uma série de estratégias para potencializar cada um dos quatro componentes: atenção, relevância, confiança e satisfação. O modelo é válido para os mais diversos tipos de material instrucional e muitas das suas estratégias para motivar o aluno são altamente compatíveis com jogos digitais. Entre as estratégias compatíveis com jogos estão:

- **participação:** KELLER (1987) lista jogos como uma das alternativas que exigem participação do jogador. Jogos por natureza exigem haver pelo menos interações entre o jogador e o sistema e muitas vezes entre os jogadores.
- **escolha:** o aluno deve ter controle sobre a maneira na qual as tarefas são feitas.
- **recompensas inesperadas** para as diversas tarefas do material instrucional  
**dificuldade crescente:** para que o aluno realisticamente consiga completar as tarefas a medida que aprimore as suas habilidades no jogo.

Outra teoria relevante de motivação ao aprendizado é a de Zona de Desenvolvimento Proximal, de Vygostky, e a de Scaffolding, comumente atribuída ao mesmo autor, mas havia sido definida por WOOD *et al.* (1976). A Zona de Desenvolvimento Proximal, proposta em VYGOTSKY (1978), rompe com a noção de que o desenvolvimento cognitivo de uma criança deve ser avaliado somente conforme a sua capacidade de resolver problemas de forma independente. Alternativamente, o conceito de ZDP sugere que a avaliação deve ser feita em relação ao quão capaz o aluno é de resolver problemas quando auxiliado por instrutores ou alunos mais capazes. É avaliado o potencial de aprender conceitos próximos do nível de conhecimento do aluno, não apenas um retrato da capacidade do aluno naquele momento.

O conceito de auxiliar o aluno no aprendizado para maximizar o seu potencial dentro da Zona de Desenvolvimento Proximal é conhecido como Scaffolding. OBIKWELU (2017) descreveu diversas maneiras que jogos educacionais podem usar *Scaffolding*, para auxiliar o aprendizado do aluno. Estas formas de auxílio no aprendizado foram validadas através do estudo de um nove jogos sérios.

As maneiras identificadas por OBIKWELU (2017) foram:

- **Explicação/Deixas:** uso de informações não-personalizadas comunicadas verbalmente. Jogos costumam ter deixas na forma de descrições de componentes de forma isolada ou agregados em manuais e glossários.
- **Orientação:** no contexto de jogos é o uso de dicas, alertas e explicações para apoiar na resolução de problemas. Orientações podem ter caráter retrospectivo, oferecendo um *feedback* positivo ou negativo ao jogador.
- **Colaboração:** em jogos educacionais colaborativos, os jogadores apoiam uns aos outros na resolução de problemas. A colaboração tem efeitos motivacionais e incentivam a comunicação entre os jogadores.
- **Reflexão:** reuniões entre o instrutor e os alunos para refletir e discutir a experiência do jogo. As reflexões podem ser feitas após alguma etapa ou do jogo em si. No lugar de reuniões, jogos digitais possuem ferramentas para a captura de dados durante o jogo, e os dados podem ser processados e comunicados ao usuário para prover informações sobre o seu desempenho. Isso convida-o a refletir sobre a sua experiência do jogo sem a necessidade de uma intervenção manual por parte de um instrutor.

A colaboração, incluída como elemento de apoio ao aprendizado no modelo acima, possui efeitos também na motivação ao aprendizado, pois jogos cooperativos tem características particulares quanto aos seus efeitos motivacionais. O papel desempenhado por cada jogador no trabalho em equipe faz com que os indivíduos

tenham uma melhora na autoestima devido à *aceitação* (ORLICK, 1981). A sensação de *pertencimento* também é evocada por conta dos objetivos e procedimentos compartilhados com outros (ORLICK, 1981). É importante ressaltar que cooperação e competição não são inteiramente mutualmente exclusivas em jogos. Jogos cooperativos demandam um grau de competitividade, por serem configurados no formato *jogadores vs. sistema*. Outra exceção são jogos competitivos disputados em formato *equipe vs. equipe*. Nesses jogos, há uma **cooperação** interna à equipe e o fator soma-zero, uma equipe ganha e a outra perde, inerente a jogos **competitivos**.

Jogos educacionais digitais devem ser ferramentas capazes de motivar o aluno a aprender sobre o assunto em questão. Em GARRIS *et al.* (2002), é proposto um modelo de utilização de jogos digitais como ferramentas motivacionais de aprendizado. A atividade instrucional deve ser intrinsecamente instigante, mas também gerar interesse por conta dos resultados no aprendizado.

Seguindo o modelo definido em GARRIS *et al.* (2002), o material instrucional deve ter características ou atributos inerentes a jogos. Em seguida, há a etapa de processo no qual o jogo desenvolvido é fruto de uma interação cíclica com os alunos. O ciclo de jogo detalhado pelo modelo passa pelos: efeitos que o aluno sente em relação aos acontecimentos do jogo, comportamentos do jogador consequentes aos efeitos, e o feedback do sistema do jogo aos comportamentos do jogador. Em resumo, este processo cíclico enaltece como jogos podem ser materiais instrucionais interativos e iterativos.

Jogos podem ser jogados múltiplas vezes, e o conceito do desejo de um jogador jogá-lo múltiplas vezes é conhecido como *replayability* (ADETUNJI e ADE-IBIJOLA, 2024). A variação no comportamento de diferentes componentes do jogo, introduzindo uma certa aleatoriedade ou ruído, está entre os fatores que levam um jogo a ser jogado múltiplas vezes (ADETUNJI e ADE-IBIJOLA, 2024; KOSTER, 2013).

Outro aspecto de suma importância dos jogos digitais, por serem sistemas de software, é que proporcionam uma boa experiência de usuário. Jogos digitais diferem-se de outros sistemas de software devido ao seu uso recreativo (SÁNCHEZ *et al.*, 2012). Outra diferença interessante é que a dificuldade para a resolução de tarefas é encarada como algo positivo no contexto de jogos (FU *et al.*, 2009; MALONE, 1981; SÁNCHEZ *et al.*, 2012). Por outro lado, a frustração proveniente de uma dificuldade em completar tarefas é considerada negativa para sistemas de software com fins *tradicionais* (SAARILUOMA e JOKINEN, 2014). Consequentemente, pode ser argumentado que modelos de qualidade de software focados em usos gerais não englobam todos os requisitos para uma boa qualidade de experiência de jogos digitais; de fato, já existem modelos voltados à avaliação da experiência de usuário voltado ao contexto específico de jogos; por exemplo, o modelo proposto em SÁNCHEZ *et al.*

(2012) complementa a experiência de usuário com características específicas a jogos, denominando esta experiência como a jogabilidade do produto.

### 3.1.2 Jogabilidade

O termo *playability*, comumente traduzido como jogabilidade, se refere à qualidade da experiência do usuário nas suas interações com um jogo digital. A jogabilidade abrange componentes do jogo como as suas regras, mecânicas, objetivos e design (SÁNCHEZ *et al.*, 2012). O conceito de jogabilidade é definido por SÁNCHEZ *et al.* (2012) como a *Player Experience (PX)*, ou experiência do jogador e não rompe com noções gerais de experiência de usuário, mas sim aplica critérios de UX como efetividade, eficiência e satisfação ao contexto e aos objetivos de jogos (SÁNCHEZ *et al.*, 2012).

Algo que diferencia a jogabilidade do conceito mais geral de UX é que jogos digitais tem fins recreativos. Isso os diferencia de sistemas de software *tradicionais*, feitos para cumprirem alguma tarefa desejada pelo usuário e comumente medidos por sua usabilidade (SÁNCHEZ *et al.*, 2012). Por exemplo, no contexto específico a jogos, as sensações de frustração e de dificuldade para completar uma tarefa são bem-vindas em um nível moderado, apesar de serem indesejadas em sistemas digitais em geral.

O modelo de jogabilidade de SÁNCHEZ *et al.* (2012) organiza os critérios de jogabilidade por atributos intrínsecos ao produto de software e seus processos e os de interação com o usuário. Estas atribuições dos atributos traçam um paralelo com SQuaRe, que especifica critérios de qualidade interna e externa ao produto e qualidade em uso do mesmo. Apesar das semelhanças gerais, critérios como diversão, conhecimento sobre o jogo, entre outros são exclusivos ao contexto de experiência de jogador ou sofrem adaptações significativas em suas definições comparadas a critérios de experiência de usuário.

A jogabilidade também diferencia-se de experiência de usuário tradicional por abordar o aspecto social de jogos. Medem-se os diferentes tipos de interação entre os jogadores, a percepção do jogador sobre seu papel no jogo e, em casos de jogos colaborativos, o reconhecimento de pertencimento a uma equipe, entre outros critérios (SÁNCHEZ *et al.*, 2012).

A avaliação de jogos digitais, portanto deve ser acompanhada de critérios além daqueles propostos por modelos projetados para sistemas de software *tradicionais*. Assim, há como medir de forma mais completa a experiência do jogador.

A área de avaliação de jogabilidade está diretamente relacionada aos efeitos (WINN, 2009) ou as estéticas (HUNICKE *et al.*, 2004). O termo efeito será usado nesta pesquisa para abranger a definição de estética, de forma que fique pa-

dronizado.

Os efeitos são evocados pela experiência com o jogo, nas interações com os componentes e comportamentos do seu sistema. Esses efeitos podem ser modelados através do design de jogos. Portanto, conceitos presentes nos seguintes modelos de *Game Design* foram incorporados ao modelo sendo proposto: o *GameFlow* (SWETSER e WYETH, 2005) e o *6-11 Framework* (DILLON, 2010).

Antes de adentrar no modelo GameFlow, é importante definir *flow*, do qual o modelo adapta para o contexto de jogos. Ao definir a sensação de fluxo, CSIKSZENTMIHALYI (2008) encontrou um grupo de características em comum nas mais diversas atividades que levam a uma sensação de plenitude e satisfação. Essas atividades abrangem desde leitura à prática de esportes, independente da região onde estão sendo praticadas e outros fatores externos à atividade (CSIKSZENTMIHALYI, 2008).

Segundo o autor, as atividades que proporcionam experiências de fluxo devem atender os seguintes critérios (CSIKSZENTMIHALYI, 2008):

1. As tarefas confrontadas são aquelas com ambições realísticas.
2. As atividades são exercidas em ambientes que permitam a concentração dos envolvidos.
3. As tarefas tem objetivos claros.
4. As tarefas tem um retorno imediato às ações das pessoas envolvidas.
5. As tarefas proporcionam às pessoas uma sensação de envolvimento, ao ponto de proporcionar uma sensação de distanciamento temporário dos problemas cotidianos.
6. A pessoa exerce controle sobre a tarefa.
7. A pessoa perde o senso de autoidentificação durante a atividade, mas ao seu término tem este sentido aguçado.
8. A tarefa proporciona uma distorção na percepção da passagem do tempo.

Há uma congruência relevante entre as características de atividades que gerem experiências ótimas, de fluxo, e as características inerentes a jogos. A adaptação do fluxo para o contexto de jogos é modelada pelo GameFlow. Além dos requisitos listados acima, deve ser destacado que as atividades devem ser demarcadas por regras, devem demandar alguma habilidade e serem autotélicas, tais qualidades também são intrínsecas a jogos.

O papel do *GameFlow* nessa pesquisa foi de fundamentar a medição de requisitos de fluxo e do componente de interação social adicionado ao GameFlow por ser fundamental a experiências em jogos.

A sensação de presença é um pré-requisito de fluxo, os fenômenos compartilham características teóricas (TAKATALO *et al.*, 2010) e existem resultados científicos que provam a relação de pré-requisito entre estes conceitos (NOVAK *et al.*, 2000). Uma vertente da sensação de presença em interações humano-computador e outras mídias é a de *realismo*, no qual se distingue entre a realidade perceptível, construída via representações visuais ou sonoras, e a realidade social, construída a partir das interações entre agentes do jogo (LOMBARD e DITTON, 1997). A presença relacionada ao realismo ocorre quando algum evento ou objeto dão a impressão em algum grau de que são *reais* (LOMBARD e DITTON, 1997). Outro tipo de presença é a *presença como transporte*, no qual a pessoa sente que (LOMBARD e DITTON, 1997):

1. está no ambiente da mídia com a qual está interagindo
2. que os eventos e objetos da mídia estão no seu ambiente *real*
3. ou que está compartilhando o mesmo espaço com as outras pessoas interagindo com o sistema virtual

Outro modelo de design de jogos é o 6-11 framework(DILLON, 2010). O nome do modelo faz referência às seis emoções e onze instintos que o compõem, definidos como uma taxonomia de possíveis ganhos estéticos ou reações de jogadores. Enquanto o GameFlow descreve com profundidade a sensação de fluxo e sua relação a jogos, a contribuição do 6-11 Framework está na sua taxonomia cujas emoções serviriam para auxiliar na classificação do critério de diversão. Além disso, os instintos de competitividade, comunicação, apreciação por cores e curiosidade reforçaram o embasamento de critérios relacionados aos mesmos no modelo.

## 3.2 Fundamentos Matemáticos Abordados

Antes de especificar as técnicas e métodos aplicados nessa pesquisa, é importante repassar alguns dos fundamentos matemáticos utilizados pelas técnicas de avaliação do modelo de avaliação e do formulário. Um deles, a lógica Fuzzy, mais especificamente em sua aplicação para análise de consenso e tomadas de decisão, foi utilizada para um método originado nesta pesquisa para a etapa de validação do conteúdo do formulário.



## Lógica Fuzzy

A lógica clássica, aristotélica, tem como características valores booleanos, verdadeiro (1) ou falso (0) e declarações linguísticas com conjuntos verdadeiros e conjuntos falsos. Entretanto, a imprecisão de determinadas proposições, como as de interpretação da linguagem natural, não se encaixam devido à limitação dos valores lógicos a apenas o verdadeiro ou falso absoluto.

Pessoas fazem decisões racionais em cenários imprecisos e conseguem inferir mesmo com graus de inexatidão ou incongruência de informação. A lógica fuzzy, diferente daquela baseada em *crisp sets*, permite a inferência em cenários imprecisos ao representar uma preposição lexicalmente imprecisa como uma restrição elástica numa variável e o resultado como um desdobramento das restrições elásticas (ZADEH, 1988). O objetivo da lógica Fuzzy é formar uma fundamentação teórica para tais contextos com proposições imprecisas (ROSS, 2009).

O cálculo Fuzzy de grau de consenso é uma das aplicações de tomadas de decisão Fuzzy-Bayesiana.

$$F(R) = \frac{tr(R^2)}{\frac{n(n-1)}{2}} \quad (3.1)$$

A função  $tr()$  é o traço da matriz é a soma algébrica dos valores da diagonal principal da matriz. Expressa como:

$$tr(R) = \sum_{i=1}^n r_{ii} \quad (3.2)$$

$$C(R) = \frac{tr(RR^T)}{\frac{n(n-1)}{2}} \quad (3.3)$$

O grau *fuzzy* médio é proporcional à confusão ou incerteza no fenômeno que está sendo medido, o valor é maximizado quando não há consenso algum entre as alternativas e  $r_{ij} = r_{ji} = 0,5$ . De forma recíproca, o grau de certeza médio é proporcional ao grau de consenso que uma alternativa tem preferência sobre a outra, com o valor maximizado em cenários no qual  $r_{ij} = 1$  e  $r_{ji} = 0$ . Os valores são dependentes e ficam num intervalo  $[0, 1]$ , no qual, segundo BEZDEK *et al.* (1978):

$$F(R) + C(R) = 1 \quad (3.4)$$

As métricas de preferência são úteis para o cálculo de distância ao consenso, a distância ao consenso tem definido dois tipos de consenso que podem ser atingidos:

- **Tipo I:** representa o caso no qual a alternativa preferencial é clara, mas as

demais alternativas são igualmente escolhidas. O consenso do Tipo I tem uma distância ao consenso  $m(R) = 1 - \frac{2}{n}^{\frac{1}{2}}$ .

- **Tipo II:** ocorre quando a alternativa preferencial,  $i$ , é evidente e as alternativas secundárias têm preferências certas sobre as demais ( $r_{kj} = 1; i \neq k$ ).

Para atingir o Tipo II de consenso não é necessário haver uma ordenação definida entre as opções secundárias.

A distância ao consenso é calculada a partir do grau de consenso, sendo definida por ROSS (2009) como:

$$m(R) = 1 - (2C(R) - 1)^{\frac{1}{2}} \quad (3.5)$$

Cada um dos tipos de consenso definido é acompanhado de uma métrica para definir se houve consenso ou não. O Tipo I tem como definição de consenso o seguinte valor de distância ao consenso, em um cenário com  $n$  alternativas:

$$m(R) = 1 - \left(\frac{2}{n}\right)^{\frac{1}{2}} \quad (3.6)$$

Enquanto o Tipo II é mais restritivo e exige que a distância ao consenso seja equivalente a 0.

A antítese do consenso é representada pela matriz  $M_1$ , na qual todos os itens fora da diagonal principal possuem um valor de 0,5, indicando uma preferência equivalente entre alternativas  $i$  e  $j$ .

## Estatística

Ao lidar com variáveis estatísticas é necessário identificar previamente qual é o nível de erro aceitável e fazer a medição dos erros. Medições estatísticas têm dois tipos de erro diferente:

- **Tipo 1 ( $\alpha$ )** : a probabilidade de rejeitar a hipótese nula incorretamente, a chance de ocorrer um *falso positivo*.
- **Tipo 2 ( $\beta$ )** : a probabilidade de aceitar a hipótese nula erroneamente, a chance de ocorrer um *falso negativo*.

Outro conceito fundamental ao testar hipóteses no campo da estatística é o de poder estatístico. O poder estatístico é a probabilidade que uma hipótese nula será rejeitada, levando à comprovação de um fenômeno de interesse. A averiguação do poder estatístico deve ser feita, pois mesmo que o fenômeno esteja presente, a hipótese nula da pesquisa só pode ser rejeitada caso o experimento possua significância estatística (HAIR *et al.*, 2009). Consequentemente, deve ser feito um planejamento

antes mesmo da condução do experimento para evitar uma execução que leve a resultados ambíguos. O poder estatístico possui três dependências (COHEN, 2013):

1. a confiabilidade das medições
2. o critério de significância
3. o efeito

A confiabilidade das medições da amostra é a medição do quão próximo é esperado que os resultados da amostra sejam dos valores populacionais (COHEN, 2013). Há sempre uma dependência desta métrica ao tamanho da amostra (COHEN, 2013). O tamanho do efeito costuma ser a diferença de uma métrica de interesse em uma amostra na qual um fenômeno está presente e uma amostra de controle.

Existem quatro tipos de análise de poder estatístico, um para cada parâmetro da função de inferência estatística: o valor do poder em si, o nível de significância ( $\alpha$ ), o tamanho da amostra e o tamanho do efeito. Cada um desses parâmetros é funcionalmente dependente dos outros três. Assim, é possível fixar três dos parâmetros para obter o resultado necessário para o quarto valor ausente.

A definição do nível de poder estatístico é feita antecedendo a execução do experimento, avaliando os efeitos sobre os erros do Tipo 1 e Tipo 2. Também compara-se a gravidade que ambos os erros terão caso ocorram na prática. Tornar a tolerância de erros do Tipo 1 menor implica na redução do poder estatístico e aumenta o risco de um falso negativo, cabe ao pesquisador pesar a proporção aceitável de ambos os tipos ao definir o nível estatístico.

O teste qui-quadrado, elaborado inicialmente em PEARSON (1900), é usado para testar hipóteses sobre a relação de duas variáveis categóricas. A hipótese nula do teste qui-quadrado é que as amostras das variáveis sob comparação não possuem uma diferença estatisticamente significativa. Isto é, o teste é usado para provar a similaridade ou disparidade entre um par de variáveis, com um grau de rigor determinado pela estatística  $\alpha$ .

O valor da estatística qui-quadrado é obtido através da comparação agregada entre os valores observados ( $O_i$ ) e os esperados para cada observação ( $E_i$ ). Segue a fórmula do cálculo do valor de qui-quadrado:

$$\chi^2 = \frac{\sum (O_i - E_i)^2}{E_i}$$

A estatística qui-quadrado gerada por um experimento é comparada à distribuição qui-quadrado para analisar, via o teste exato de FISHER (1925). O teste exato de Fisher calcula o valor p, que representa a probabilidade de um valor ser tão ou mais extremo segundo a distribuição. A hipótese nula é rejeitada caso o valor da

estatística qui-quadrado gerada for superior ao valor na distribuição referente aos graus de liberdade e valor-p do experimento.

Essa pesquisa usou o teste qui-quadrado em diversos contextos. Por exemplo, foi feita uma análise usando um teste Kruskal-Wallis, cuja distribuição se aproxima de uma distribuição qui-quadrado para amostras maiores que 5, para identificar diferenças significativas entre grupos sociodemográficos. Outra aplicação do teste foi na verificação de adequabilidade dos dados, na qual o objetivo era aceitar uma hipótese nula que afirma que a matriz de cargas fatoriais confirmatórias possuem valores esperados dada a matriz fatorial teórica. Essa segunda aplicação não usa o teste qui-quadrado tradicional. De fato, o teste em questão é o teste da razão de verossimilhança, que por sua vez utiliza a distribuição qui-quadrado.

O teste qui-quadrado da razão de verossimilhança tem uma distribuição assintótica à distribuição qui-quadrado assumindo que a hipótese nula é verdadeira. A estimativa de máxima verossimilhança, inicialmente teorizada por FISHER (1922), possui uma variação para o contexto de matrizes de covariância.

O valor do teste estatístico da razão de verossimilhança é obtido ao calcular a diferença da máxima verossimilhança de: uma matriz de covariância seguindo um modelo irrestrito e uma matriz de covariância com as restrições impostas por um modelo teórico. Tais restrições impostas pelo modelo teórico fixam itens da matriz como zero. Os graus de liberdade do teste estatístico de máxima verossimilhança são calculados a partir da diferença na quantidade de parâmetros do modelo irrestrito comparado ao restrito.

Por se tratar de um teste estatístico assintótico à distribuição qui-quadrado, é possível compará-lo ao valor de tal distribuição dado os graus de liberdade e nível de significância estatística escolhidos para o experimento. Assim, é esperado que o valor do teste estatístico seja inferior àquele da distribuição qui-quadrado e, conseqüentemente, a hipótese nula seja aceita.

O teste t, definido por Student em STUDENT (1908), é outro teste estatístico usado para aceitar ou rejeitar hipóteses para variáveis que seguem distribuições normais. Nem sempre dados populacionais estão disponíveis em uma pesquisa, impedindo que métricas como a média populacional sejam usadas para análises estatísticas. Alternativamente, são usadas estimativas dos dados populacionais a partir dos dados amostrais.

A estimativa do erro padrão é calculada usando os dados amostrais de desvio padrão da amostra e a quantidade de observações que compõem a amostra. O erro padrão é a quantificação da diferença esperada entre uma média amostral e a média populacional. Em situações na qual dados de toda a população estão disponíveis, esta diferença esperada é o desvio padrão. A fórmula da estimativa do erro padrão é:

$$s_x = \frac{s}{\sqrt{n}} \quad (3.7)$$

Na qual  $s_x$  é a estimativa do erro padrão,  $s$  é o desvio padrão da amostra e  $n$  é a quantidade de observações da amostra.

O teste t (Fórmula 3.8) é usado para testar a hipótese que uma média amostral é igual à média populacional em situações nas quais os dados populacionais são desconhecidos. O valor  $\mu_0$  representa um valor fixo escolhido como a hipótese nula que está sendo testada.

$$t = \frac{M - \mu_0}{s_x} = 0 \quad (3.8)$$

O teste de t possui  $n - 1$  graus de liberdade. Além disso, sua distribuição t é uma aproximação da distribuição normal. Portanto, é assumido que os dados sendo testados seguem uma distribuição normal. Outro pré-requisito do teste é que as observações sejam independentes.

O valor da estatística t é então comparado à distribuição t e, em conjunto com os graus de liberdade, usado para identificar o valor-p correspondente. A hipótese nula é aceita caso o valor-p seja superior à estatística t e rejeitada caso contrário.

### Comparações entre Grupos Categóricos (Análise de Variância)

Fatores de influência humana, como o perfil demográfico ou socioeconômico, constituição física ou mental e o estado emocional da pessoa afetam a qualidade de experiência em jogos. Por exemplo, estudos passados já apresentaram diferenças entre os gêneros referente a padrões neurológicos enquanto jogam (HOEFT *et al.*, 2008) e relativos a vício em jogos eletrônicos (HSU *et al.*, 2009). A análise de variância (ANOVA) é uma técnica apropriada para a resolução deste problema, pois averigua se respondentes de determinados grupos possuem padrões de resposta significativamente diferentes para algum item.

Entretanto, a análise de variância exige que os grupos tenham variâncias equivalentes e os valores dos grupos sigam uma distribuição normal. Comparações entre grupos categóricos que não atendem a tais requisitos podem ser feitas com o teste Kruskal-Wallis.

O teste Kruskal-Wallis é outro método de análise de variância usado para a identificação de diferenças estatisticamente relevantes entre grupos independentes sobre uma variável de interesse (KRUSKAL e WALLIS, 1952). Este teste difere-se da análise de variância por se tratar de um método não paramétrico. Além disso, o método não usa métricas de tendência central que assumem valores contínuos. Portanto, dados ordinais discretos podem ser utilizados no teste Kruskal-Wallis.

Enquanto o ANOVA utiliza a média de grupos categóricos para testar a igualdade, o Kruskal-Wallis utiliza um sistema de ranqueamento para seu teste.

Os dados de todos os conjuntos independentes são ranqueados de maneira ascendente, e caso haja um empate o valor é substituído pela média dos *rankings* que os dados empatados receberiam. O valor dos rankings é então usado na seguinte fórmula para calcular a estatística H do teste:

$$H = (N - 1) \frac{\sum_{i=1}^g n_i (\bar{r}_i - \bar{r})^2}{\sum_{i=1}^g \sum_{j=1}^{n_i} (\bar{r}_{ij} - \bar{r})^2} \quad (3.9)$$

No qual:  $N$  é o total de observações,  $g$  é a quantidade de grupos,  $n_i$  são as observações do grupo  $i$ ,  $r$  são os rankings e  $\bar{r}$  é a média dos rankings. A estatística de teste H é comparada a um valor crítico da distribuição qui-quadrado, usando o nível de significância desejado e  $g - 1$  graus de liberdade.

É recomendável realizar o teste Kruskal-Wallis para amostras cujo total é superior a cinco, por ser nesses casos que o teste estatístico se aproxima da métrica  $\chi^2$  (KRUSKAL e WALLIS, 1952) (OF VIRGINIA LIBRARY, Year the webpage was last updated or accessed).

### 3.3 Desenvolvimento de Escalas Psicométricas

DEVELLIS e THORPE (2021) propuseram um guia de desenvolvimento de escalas psicométricas, adaptado para o contexto desta pesquisa. A versão original possui os seguintes passos:

1. Determinar com clareza o que deve ser medido.
2. Gerar um conjunto de itens candidatos a entrar na escala.
3. Determinar o formato de medição.
4. Ter o conjunto de itens candidatos a entrar na escala inicial revisado por especialistas.
5. Considerar a inclusão de itens de validação.
6. Administrar os itens para uma amostra de treinamento.
7. Avaliar os itens.
8. Otimizar o tamanho da escala.

Esta pesquisa segue estes passos, fazendo algumas adaptações. A diferença se encontra na ausência das etapas de incluir itens de validação e de administrar os itens para uma amostra de treinamento.

O terceiro passo do guia exige a determinação de um formato de medição para o instrumento psicométrico. A próxima seção detalha o formato de medição utilizado nessa pesquisa.

### 3.3.1 Escalas de Medição

Existem diversos tipos de medição encontrados na literatura. Três tipos de escalas são discutidos a seguir: a escala de Guttman, a escala Stapel e a escala Likert.

A escala de Guttman (GUTTMAN, 1944) é usada para derivar um fenômeno unidimensional a partir de perguntas dicotômicas. GUTTMAN (1944) se baseou no conceito de função simples; isto é,  $y$  é uma função de  $x$  apenas se para cada valor de  $x$  existir apenas um valor de  $y$  correspondente.

Os itens da escala são hierarquizados de forma que, se um respondente responde negativamente a um item, o mesmo comportamento deve ser observado nos posteriores. Esse comportamento se deve à ordenação de afirmações gradualmente mais extremas no decorrer do formulário.

Entre as vantagens do uso desse método de medição estão: a capacidade de quantificar o nível ou ranking de um respondente dado o caráter hierárquico da escala (DEVELLIS e THORPE, 2021), e a simplicidade de responder perguntas dicotômicas. Em contrapartida, escalas Guttman não são indicadas para avaliar fenômenos abstratos na qual não há uma hierarquia clara entre os itens (DEVELLIS e THORPE, 2021), também não é indicada caso o fenômeno sob medição seja multidimensional.

A escala Stapel, introduzida por (CRESPI, 1961), é um método de medição unipolar e de 10 pontos. A escala apresenta um adjetivo correspondente a uma afirmação e pede ao respondente que avalie em uma escala que varia de -5 a -1 e 1 a 5 o quanto o adjetivo descreve tal afirmação. Portanto, a escala mede a compatibilidade de um adjetivo com alguma afirmação e captura também a sua intensidade. Neste tipo de escala não é permitido uma resposta neutra.

A escala Likert, apresentada por LIKERT (1932), é um formato de resposta bipolar, permitindo que o respondente avalie algo de forma positiva ou negativa. A escolha de uma escala Likert permite uma variação na polaridade e na intensidade (e.g. discordo moderadamente vs. discordo fortemente). Outra característica é que a escala é de natureza ordinal; suas opções não se equiparam a uma quantia real, apenas oferecem uma ordenação de opções.

Escalas Likert com um número par de alternativas têm a ausência de uma opção neutra. Nestes casos, é exigido que o respondente afirme, mesmo que ligeiramente, para qual lado ele está mais inclinado. Entretanto, o respondente verdadeiramente tem uma opinião neutra quanto ao item e, portanto, muitos estudos têm optado por

escalas com quantidades ímpares de alternativas (COLMAN *et al.*, 1997).

Por lidar com categorias não quantificáveis, não podemos assumir valores unitários e uniformes entre as opções de uma escala Likert. Por exemplo, em uma escala que possua “não concordo, nem discordo”, “discordo moderadamente” e “discordo fortemente” entre as suas alternativas de resposta não se pode assumir que a opção do meio é equidistante das outras duas. Nestas situações, pode-se assumir apenas que as opções estão ordenadas numa escala Likert de satisfação do respondente.

Seguindo esse raciocínio, caso a escala dessa pesquisa seja aplicada para avaliar dois jogos educacionais digitais distintos e tivermos uma média de respostas mais favoráveis para um dos jogos, podemos afirmar que um determinado jogo foi melhor avaliado que outro. Entretanto, a diferença das avaliações não poderia ser quantificada conforme os resultados da escala (HAIR *et al.*, 2009).

## 3.4 Avaliação do Instrumento de Medição

### 3.4.1 Psicometria

A informação psicométrica na distribuição de valores de uma escala se resume nos seguintes fatores (NUNNALLY, 1978):

1. média
2. desvio padrão
3. formato da curva de distribuição
4. confiabilidade
5. composição fatorial

Deste modo, um instrumento de psicometria deve otimizar essas métricas ou a relação entre elas para ser de alta qualidade. Os objetivos são uma distribuição próxima da normal, alta confiabilidade e uma composição fatorial que descreva o construto medido pela escala psicométrica (NUNNALLY, 1978). O objetivo dos testes psicométricos é obter diferenças significativas, individuais e confiáveis para ser possível identificar correlações entre os fatores e o construto (NUNNALLY, 1978).

Um instrumento com conteúdo considerado confiável, mas inválido, para os construtos em análise, pode ser válido para uma pesquisa observando um conjunto de variáveis diferente (HAYNES *et al.*, 1995). Por isso, é importante avaliar tanto a confiabilidade quanto a validade e refinar o conteúdo do instrumento de medição a partir dos resultados.



A validade determina se o instrumento realmente mede aquilo que se propõe a medir; a confiabilidade mede o quanto essa medição do instrumento é reproduzível. A validade de um instrumento de medição depende de sua confiabilidade. Não há como assegurar que o instrumento está de fato medindo algum fenômeno de interesse se o mesmo instrumento não é consistente nos seus resultados. A inconsistência nos resultados impõe que o instrumento não pode ser válido para a medição do fenômeno de interesse, por sequer estar medindo algum fenômeno consistentemente. Entretanto, uma observação confiável não é necessariamente válida (HOSS e TEN CATEN, 2010).

## Confiabilidade

Ao desenvolver uma escala psicométrica, é fundamental assegurar sua confiabilidade para atestar a sua veracidade científica (STRAUB *et al.*, 2004). A afirmação da validade de um instrumento de medição é infundada se antes não for atestado que o instrumento está seguramente medindo algum construto.

No contexto de desenvolvimento de escalas psicométricas, a confiabilidade indica o quanto a variável latente influencia os itens da escala. Variância nas medições que não são provenientes da variável latente são consideradas erro, ou ruído. É desejável que uma proporção significativa da variância da escala seja conjunta, compartilhada entre todos os itens, pois isso indica haver uma consistência interna alta, e, por consequência, que a medição da escala é confiável.

A variância conjunta é explicada por mudanças no fenômeno observado. Por exemplo, uma escala confiável que meça o nível de ansiedade de algum paciente deveria variar paralelamente consoante à variável latente, o nível real de ansiedade. A variância de um item consigo mesmo, a sua covariância, é identificada como erro. A covariância interna a um item indica o nível de dispersão nas suas respostas. Níveis elevados de covariância interna significam que tal item não apresenta uma consistência interna suficiente para medir um fenômeno.

É desejado que haja uma covariância de alta magnitude entre os itens servindo como *proxy* para o mesmo fator latente; itens que representem o mesmo construto. Artefatos com uma consistência interna alta indicam uma correlação alta entre os itens. Assumindo os pressupostos estabelecidos pelo modelo de testes paralelos, uma boa consistência interna também significa uma forte correlação entre os itens em questão e o fenômeno sob observação. O modelo de testes paralelos assume que cada variável representa tão bem o fenômeno observado quanto as outras variáveis sendo medidas. Isto é, a variável latente tem um efeito igual sobre cada item e o nível de erro também é igual (DEVELLIS e THORPE, 2021).

Assume-se que a variância conjunta dos itens varia conforme a variável latente ali representada, enquanto a variância individual é atribuída ao erro. A variância

dos itens consigo mesmo e com os outros itens do instrumento pode ser visualizada em matrizes de covariância. No exemplo abaixo,  $\sigma_{i,j}$  representa a covariância entre o par de itens  $i$  e  $j$  e  $\sigma_i^2$  representa a variância do item  $i$  consigo mesmo.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \dots \\ \vdots & \ddots & \\ \sigma_{k,1} & & \sigma_k^2 \end{bmatrix}$$

Os dados agregados de variância total,  $V_T = \sigma_y^2$ , equivalente à soma de todos os elementos na matriz, e erro total, equivalente à soma de todos os elementos na diagonal principal da matriz,  $\sum_i^n V_i$  são fundamentais no cálculo de consistência interna.

O Alfa de Cronbach, proposto em CRONBACH (1951), é a métrica de referência na área de psicometria para a medição da consistência interna de instrumentos de medição. Esta métrica, na verdade, é um piso para a medição da consistência interna do instrumento TEN BERGE e ZEGERS (1978). Outras métricas, como a confiabilidade *Split-Half* e a correção Spearman-Brown também são popularmente usadas para medir a confiabilidade de um instrumento.

O valor do alfa de Cronbach (CRONBACH, 1951), calculado a partir da fórmula disponível em seguida, é relativo à proporção variância comunal. Esta relação é evidenciada na expressão,  $(1 - \frac{\sum_i^n V_i}{V_t})$ , que se limita ao intervalo  $0 \leq \alpha \leq \frac{n^2-n}{n^2}$ . Para mantê-la no intervalo desejado,  $0 \leq \alpha \leq 1.0$  o resultado há de ser multiplicado pela recíproca,  $\frac{n^2}{n^2-n} = \frac{n}{n-1}$ .

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_i^n V_i}{V_t}\right) \quad (3.10)$$

No qual,  $i$  representa o item,  $V_i$  representa a variância daquele determinado item,  $V_t$  é a variância total dos resultados e  $n$  a quantidade de itens na escala.

O Alfa de Cronbach é amplamente usado na avaliação de instrumentos de medição, entretanto, é possível avaliar a consistência interna a partir de outras métricas. Uma delas é a de confiabilidade *Split-Half*, ou de confiabilidade por Duas Metades, cujo valor é derivado de uma técnica pertencente à classe de testes paralelos. Em testes paralelos de confiabilidade, é assumido que múltiplos instrumentos de medição sejam causados pela mesma variável latente e a correlação entre estes instrumentos seja equivalente à confiabilidade deles (DEVELLIS e THORPE, 2021).

O cenário descrito referente ao diagrama acima representa dois instrumentos de medição completos. No caso do *Split-Half*, a métrica de confiabilidade resultante da variância entre as metades corresponde à confiabilidade justamente para estes segmentos do formulário (DEVELLIS e THORPE, 2021). Entretanto, a confiabilidade

de interesse é a do formulário na totalidade. Consequentemente, o valor de confiabilidade de *Split-Half* é usado para identificar a média de correlação entre os itens e recalculá-lo o valor de confiabilidade, desta vez assumindo que a quantidade de itens é a do formulário completo. Este valor corrigido é alcançado mediante a fórmula de confiabilidade de Spearman-Brown (BROWN, 1910; SPEARMAN, 1910):

$$\alpha = \frac{nr}{1 + (n - 1)r} \quad (3.11)$$

Esta fórmula evidencia que a confiabilidade é proporcional ao número de itens no formulário; fixando as demais variáveis, um formulário mais longo é mais confiável que uma alternativa mais curta (DEVELLIS e THORPE, 2021). Portanto, o valor anterior à correção do valor de confiabilidade *Split-Half* é uma subestimativa da confiabilidade real do formulário.

O processo do cálculo do *Split-Half* se resume à divisão de um formulário em dois conjuntos e calcular a correlação entre as partes. Deve ser estabelecido algum critério que reduza a presença de erros sistemáticos, por exemplo, uma simples divisão entre os primeiros itens e os últimos. Nesse cenário, especialmente se o formulário for excessivamente longo, o respondente pode estar mais cansado ou mais preparado para responder as perguntas finais. Isso implicaria em um erro mais elevado e uma variância menor decorrente da variável latente. A divisão entre itens pares e ímpares é um critério comum que evita essa fonte de erro.

## Validade

Instrumentos de medição que não passam por uma validação do instrumento abrem margem para que a sua metodologia e seus resultados sejam questionados (HOSS e TEN CATEN, 2010). Mesmo com a fundamentação em premissas e conjecturas válidas, o formulário em si e os itens que o compõem precisam ser validados. A inclusão dos fatores e critérios tal qual a maneira que estão arranjados na versão inicial do modelo são conjecturas teóricas. Os critérios serviram como hipóteses, que foram aceitas ou refutadas após a etapa de validação desta pesquisa.

Uma das validações a qual um instrumento psicométrico deve passar é o de conteúdo. Este tipo de validação tem o intuito de assegurar que o conteúdo textual de um item é pertinente ao fenômeno sendo medido e que a inclusão do item é necessária para tal medição.

O conteúdo dos itens de formulários psicométricos devem capturar bem a definição conceitual do fenômeno sendo pesquisado. O conteúdo do item deve ser formulado para o escopo adequado. Uma especificidade muito alta pode causar uma *sub representação* do conceito, negligenciando outros aspectos relevantes ao fenômeno. O oposto também é prejudicial, um escopo de demasiada abrangência resulta

em *variância irrelevante ao constructo*. Nestes casos, o instrumento de medição fica altamente sensível a aplicações em populações ou contextos diferentes.

É necessário que o conteúdo do instrumento de medição seja válido para que a variável agregada, composta pelos itens do formulário, seja representativa do construto agregado em questão (HAYNES *et al.*, 1995). Segundo HAYNES *et al.* (1995), o conteúdo de um formulário torna-se inválido quando:

- os itens dos componentes relevantes ao construto não fizeram parte do formulário
- itens observando componentes que não pertencem ao construto foram incluídos na avaliação
- Um peso desproporcional for dado a algum componente do construto na sua métrica de avaliação

Uma das métricas usadas para a medição de validade de conteúdo é o *Índice de Validade de Conteúdo* (IVC). O IVC exige o questionamento a especialistas quanto a essencialidade de um item perante um determinado objetivo de medição. O formato dos itens segue um formato de resposta categórico e ordinal, com as opções, “essencial”, “útil, mas não essencial” ou “desnecessário”. Após a coleta de dados, o IVC é computado seguindo a fórmula elaborada por LAWSHE (1975):

$$IVC = \frac{(n_e - \frac{N}{2})}{\frac{N}{2}} \quad (3.12)$$

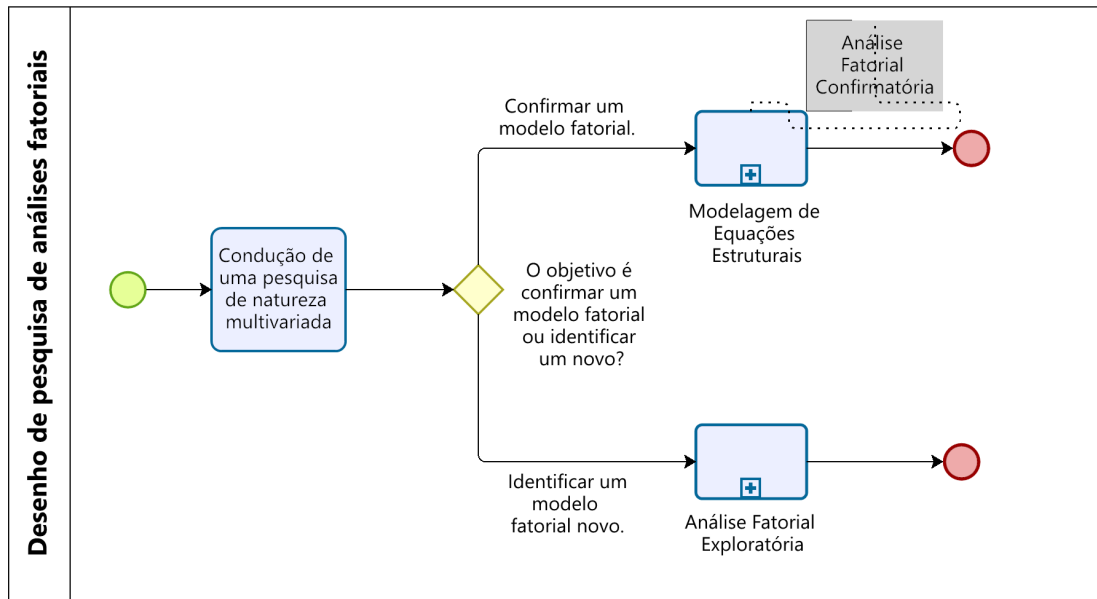
A validade de construtos é, em sua essência, a averiguação de que um conjunto de itens supostamente medidores de um fenômeno de fato medem o constructo substancialmente. Diferente da validade do conteúdo, a validade de construto não considera o texto dos itens e sim a relação entre os dados do conjunto de itens e os constructos englobados pela pesquisa.

Uma das técnicas comumente usadas para a validação dos constructos é a análise fatorial. A análise fatorial é um método estatístico que identifica e valida as relações entre variáveis, conforme as correlações entre tais variáveis da pesquisa.

A análise fatorial é um conjunto de técnicas que utilizam a correlação entre variáveis para descrever relações entre as mesmas, seja em uma perspectiva confirmatória ou exploratória. As relações entre as variáveis são atribuídas a fenômenos que não podem ser diretamente observados, mas são evidenciados pela variância conjunta das variáveis observadas.

Algumas decisões de desenho de experimento devem ser tomadas para pesquisas que englobam análises fatoriais. Um desenho de pesquisa de análises fatoriais elaborado por HAIR *et al.* (2009), foi adaptado e ilustrado nas Figuras 3.1, B.2 e B.3.

Figura 3.1: Diagrama geral de pesquisa de análises fatoriais, adaptado de HAIR *et al.* (2009)



Powered by  
bizagi  
Modeler

Estes diagramas do desenho de pesquisa seguem o padrão de modelagem de processos BPMN. Segue uma breve explicação sobre cada uma das escolhas e a alternativa escolhida no contexto desta pesquisa.

A primeira escolha que deve ser feita é referente aos objetivos da análise. Busca-se confirmar uma estrutura e hierarquia teórica pré-estabelecida? Então deve ser conduzida uma análise fatorial confirmatória, fundamentada pela modelagem de equações estruturais. É do interesse da pesquisa identificar um novo conjunto de relações entre as variáveis e fenômenos que as expliquem? Nesse caso, a análise fatorial exploratória é a recomendada.

Em adição à escolha referente à natureza confirmatória ou exploratória da análise, deve ser decidido se o objetivo da análise é a identificação da estrutura ou uma simples redução das variáveis. A identificação de uma estrutura por meio de sumarização dos dados é o uso da correlação entre as diferentes variáveis para encontrar fenômenos que expliquem as variáveis. A redução das variáveis ou redução dos dados é a identificação de um conjunto mais diminuto de variáveis que corresponda suficientemente bem à variância presente no conjunto original.

Outra decisão de natureza categórica na análise fatorial baseia-se no que a análise agrupará: as variáveis ou os respondentes? Análises fatoriais do tipo R analisam o conjunto de variáveis para identificar as variáveis latentes as quais elas explicam

(HAIR *et al.*, 2009). Já as análises fatoriais do tipo Q, que podem ser substituídas por uma análise de *cluster*, são realizadas para o agrupamento de respondentes em grupos significativamente distintos.

A análise fatorial assume que o grupo de respondentes é homogêneo, que não existem diferenças entre inerentes a subgrupos dos respondentes que impliquem diferenças significativas nas respostas (HAIR *et al.*, 2009). Por exemplo, se a nacionalidade dos respondentes for um fator determinante nas respostas, não pode ser conduzida apenas uma análise fatorial para a totalidade dos respondentes. A solução nesse caso seria a realização de uma análise fatorial para cada grupo representado.

Um dos problemas na execução de análises fatoriais é a indeterminação fatorial. Isto é evidenciado em cenários nos quais a carga fatorial de uma variável é significativa para mais de um fator. Esse problema pode acontecer mesmo após o tratamento dos dados por uma rotação dos fatores. A indeterminação fatorial implica na possibilidade de múltiplas soluções possíveis, porém não é comum que a indeterminação fatorial cause diferenças substanciais (HAIR *et al.*, 2009).

A análise fatorial difere-se de técnicas de dependência ou preditivas, como a análise multivariada de variância (ANOVA) e regressão linear, no sentido que consiste em uma análise que não distingue entre variáveis dependentes e independentes. Em essência, ela pertence às técnicas de interdependência.

Existem duas vertentes de análise fatorial: o tipo R e o tipo Q. O tipo R, utilizado nesta pesquisa, analisa o conjunto de variáveis para identificar a variável latente. Já o tipo Q é derivado do conjunto de respostas de um indivíduo, detectando padrões nas respostas para formar grupos dentre a população total de respondentes.

Outra classificação de análises fatoriais está na natureza confirmatória ou exploratória da análise. A confirmatória visa a validação de uma estrutura fatorial fundamentada na literatura e hipóteses do autor. Enquanto isso, a exploratória se fundamenta nos resultados da aplicação do experimento para o desenvolvimento de uma nova estrutura fatorial dos itens.

Antes da condução de uma análise fatorial, é necessário validar se os dados obtidos pela pesquisa estão aptos para uma análise fatorial. Neste processo de verificação de adequabilidade, a correlação entre as variáveis mensuradas são testadas para verificar se é possível serem correspondentes a fatores em comum. Os métodos de verificação de adequabilidade dos dados para análises fatoriais usados nessa pesquisa estão detalhados em seguida.

A adequabilidade do conjunto de dados coletados em um experimento a análises fatoriais deve ser assegurada antes da condução da análise. Os métodos de verificação de adequabilidade testam a existência de correlações significativas entre as variáveis. Assim, evitam-se análises fatoriais com resultados errôneos que poderiam resultar a partir de conjuntos de dados cujas variáveis não são correlatas. As duas métricas que

se sobressaem na literatura para verificar a adequabilidade perante análises fatoriais são o teste de esfericidade de Bartlett e a métrica Kaiser-Meyer-Olkin.

O teste de esfericidade de Bartlett mede a significância estatística a partir dos valores da matriz de correlação. O teste calcula a probabilidade que existam correlações estatisticamente significantes entre as variáveis da escala, conforme os valores da matriz de correlação (BARTLETT, 1954). O teste averigua se a diferença da matriz de correlação para uma matriz de identidade com as mesmas dimensões é estatisticamente relevante. Uma ressalva a este teste é o aumento na sensibilidade para a detecção de correlações significativas conforme o aumento do tamanho de amostra que gerou a matriz de correlação (HAIR *et al.*, 2009).

A fórmula para a estatística qui-quadrado do teste de Bartlett é definida como:

$$\chi^2 = -(n - 1 - (\frac{2p + 5}{6})) * \ln(\det(R)) \quad (3.13)$$

Na qual  $n$  é o tamanho da amostra,  $p$  é a quantidade de variáveis e  $R$  é a matriz de correlação.

A métrica Kaiser-Meyer-Olkin, KMO, mede a proporção de variância que pode ser atribuída a uma variância comunal. Uma alta proporção de variância comunal indica uma boa aptidão dos dados a análises fatoriais.

$$KMO = \frac{\sum \sum_{j \neq k} r_{jk}^2}{\sum \sum_{j \neq k} r_{jk}^2 + \sum \sum_{j \neq k} q_{jk}^2} \quad (3.14)$$

No qual  $r_{jk}$  é a correlação entre dois itens diferentes e  $q_{jk}$  é a correlação parcial entre os itens. A correlação parcial é a correlação entre duas variáveis após a eliminação da influência de outras variáveis nesta correlação; isto é, para testar se de fato há alguma correlação entre o par de variáveis.

KAISER e RICE (1974) define uma classificação para todos os intervalos possíveis da métrica Kaiser-Meyer-Olkin. Há de se destacar que valores abaixo de 0,5 são considerados inaceitáveis. Além disso, valores acima de 0,8 demonstram uma boa aptidão dos dados a uma análise fatorial.

Tabela 3.1: Classificação dos valores da métrica Kaiser-Meyer-Olkin de acordo com KAISER e RICE (1974).

Intervalo	Classificação
Acima de 0,90	Maravilhoso
Entre 0,80 e 0,90	Digno
Entre 0,70 e 0,80	Razoável
Entre 0,60 e 0,70	Medíocre
Entre 0,50 e 0,60	Miserável

Os dados estão aptos para uma análise fatorial em situações nas quais as métricas Kaiser-Meyer-Olkin e o teste de esfericidade de Bartlett tenham resultados satisfatórios. Uma das finalidades da análise fatorial é identificar um conjunto de fatores inobserváveis que representem as variáveis medidas em um experimento. Nessas situações, é conduzida uma análise fatorial do tipo exploratória.

O objetivo da análise fatorial exploratória é agrupar um conjunto de variáveis que tenham uma correlação alta entre si. A análise fatorial é um modelo reflexivo, pois é assumido que as variáveis medem conjuntamente um fator inobservável. Na análise fatorial, os fatores explicam a variância de um grupo de itens, sendo determinados pelas correlações observadas.

A relação entre um par de fatores pode ser oblíqua ou ortogonal. Nas relações ortogonais, o valor de um fator é estatisticamente independente do outro, enquanto relações oblíquas indicam alguma correlação entre eles. A relação teórica entre os fatores deve ser contemplada na rotação dos fatores, pois a correlação entre fatores é ignorada caso se opte por uma rotação ortogonal.

A carga fatorial dos itens é usada para identificar qual é o fator que mais exerce influência sobre cada item. O ideal é que haja uma discrepância significativa entre a influência do fator primário comparado aos dos fatores secundários. As cargas fatoriais de fatores secundários próximas a zero indicam uma relação ortogonal entre o par de fatores, o quanto mais próximo a zero a carga fatorial for, melhor (DEVELLIS e THORPE, 2021).

A estrutura fatorial é comumente simplificada mediante a rotação dos fatores, facilitando a interpretabilidade dos fatores (DEVELLIS e THORPE, 2021). THURSTONE (1947) definiu um conjunto de critérios para determinar que uma estrutura fatorial é *simples*:

1. Cada linha da matriz fatorial deve conter pelo menos um zero.
2. Para cada coluna da matriz fatorial  $V$  deve haver um conjunto de cargas fatoriais de tamanho igual à quantidade de colunas cujo valor para cada carga fatorial,  $v_{ij}$ , é igual a zero.
3. Para cada par de fatores, deve haver muitos itens cuja carga fatorial é nula para um fator e tenha algum valor para outro.
4. Para cada par de fatores, uma porção significativa deve ser igual a zero para ambos os fatores.
5. Para cada par de fatores, apenas uma pequena porção deve ter uma carga fatorial alta.



Cenários nos quais a relação entre os fatores extraídos pela análise fatorial são ortogonais são aptos para a rotação de fatores usando o método Varimax (KAISER, 1958). A rotação Varimax maximiza as cargas fatoriais com valores significantes e minimiza aqueles com valores mais baixos. Este processo aproxima a carga fatorial de cada variável para o eixo do fator que explica a maior proporção de sua variância.

A fórmula do Varimax é a seguinte:

$$v = \sum \frac{[n \sum_j (\frac{a_j^2}{h_j^2})^2 - [\sum_j (\frac{a_j^2}{h_j^2})^2]^2]}{n^2} \quad (3.15)$$

Após a execução da rotação de fatores usando Varimax, os itens da escala são agrupados em *clusters*. A proximidade entre os itens no gráfico, dentro de cada *cluster*, corresponde a uma correlação alta. Um *cluster* formado por um conjunto de itens indica que tais itens representam o mesmo fator.

A rotação de fatores resulta em uma estrutura fatorial mais *simples*, isso é fundamental para definir qual é o fator que corresponde às variáveis de um experimento. Entretanto, é necessário identificar quais são os fatores significativos, que de fato representam um conjunto de variáveis, antes de interpretar a matriz fatorial de uma pesquisa.

A análise fatorial exploratória resulta na explicação das variáveis do experimento por várias cargas fatoriais, atribuídas a fatores inicialmente não-rotulados. Segue que o conjunto de fatores teóricos é menor que o das variáveis de um experimento (CATTELL, 1966). Esta etapa, de seleção de fatores, deve ser conduzida somente após uma rotação fatorial, para evitar que a variância comunal decorrente de erro impacte nas decisões (CATTELL, 1966).

A decisão da quantidade de fatores é crucial, uma definição abrangente ou restrita demais no tamanho do modelo fatorial pode levar a resultados inesperados HAYTON *et al.* (2004). Um critério restrito demais faz com que variáveis tenham cargas falsamente atribuídas aos fatores retidos ou ao menos com cargas distorcidas em tais fatores (HAYTON *et al.*, 2004). A junção de múltiplos fatores comunais às variáveis em um torna a estrutura fatorial complexa e difícil de interpretar (FABRIGAR *et al.*, 1999).

O cenário no qual o critério retém fatores demais tende a produzir menos erro na computação das cargas fatoriais do que modelos fatoriais resultantes de critérios excessivamente rigorosos (FABRIGAR *et al.*, 1999). Entretanto, a retenção de fatores a mais também deve ser evitada por tornar o modelo fatorial menos replicável e adiciona constructos de pouco valor teórico.

O critério de Kaiser, também conhecido como critério de raiz latente, seleciona apenas os fatores que tenham uma carga superior ao peso de uma variável. Ao aplicar o critério, apenas os fatores com autovalores acima de 1,0 são retidos KAISER

(1960). No artigo que deu nome ao critério, KAISER (1960) discute alguns limites e condições provados por GUTTMAN (1954) e conclui que o critério de retenção de variáveis com autovalores acima de um é o mais adequado. A sugestão de Kaiser fundamenta-se na necessidade deste limite para atingir níveis aceitáveis de confiabilidade elaborados em KUDER e RICHARDSON (1937). Além disso, Kaiser relata que em muitos estudos observados, o uso do critério resultou em um número de fatores interpretável por especialistas (KAISER, 1960).

Ao propor o teste de Scree, outra técnica para a seleção de fatores, CATTELL (1966) faz críticas ao critério de Kaiser. É afirmado que o critério é selecionado frequentemente devido a sua simplicidade, em detrimento de sua fundamentação. Além disso, é constatado que o uso do critério de Kaiser como único critério pode levar a resultado errôneos e que a regra de selecionar apenas fatores com carga superior a 1,0 deveria se aplicar a nível populacional, não amostral.

O teste de Scree também baseia-se nos autovalores gerados pela análise fatorial. No entanto, a seleção é feita a partir de uma interpretação do gráfico de autovalores.

Ao gerar um gráfico da relação entre os autovalores e os fatores, ordenados de forma decrescente, torna-se visível um padrão no qual há inicialmente uma curva íngreme seguida de um declive praticamente linear, com pouca variação entre os autovalores dos fatores. A porção menos íngreme do gráfico chama-se *scree*, em inglês (CATTELL, 1966). Foi identificado por CATTELL (1966), que somente os fatores até o índice  $i$ , no qual tal fator é o início do *scree*, devem ser retidos para o modelo fatorial.

Um terceiro método para determinar o conjunto de fatores relevantes à medição do fenômeno é a análise paralela de HORN (1965). A análise paralela se baseia na comparação entre os valores da matriz de correlação do experimento com uma matriz de correlação das mesmas dimensões decorrente da geração de dados pseudoaleatórios.

Cada variável da matriz autogerada tem uma amostra de tamanho considerável de valores selecionados pseudo aleatoriamente a partir de uma distribuição normal. Segundo (HORN, 1965), a matriz de coeficientes de correlação, gerada seguindo as especificações listadas acima, é uma aproximação de uma matriz de identidade.

A análise *paralela* compara os autovalores da aproximação da matriz de identidade com aqueles da matriz de correlação observada e, por conseguinte, são retidos apenas os autovalores que superam os autovalores homólogos. Os itens cujo autovalor real for menor que o autovalor de mesmo índice no conjunto paralelo são descartados (DINNO, 2009).

Após a escolha dos fatores, deve ser identificada e destacada a maior carga fatorial para cada variável. Em pesquisas com amostras de tamanho inferior a 100, apenas cargas fatoriais acima de 0,30 devem ser consideradas significativas (HAIR *et al.*,

2009).

A identificação da carga fatorial predominante de uma variável nem sempre é trivial. É possível que os fatores com maior carga tenham um nível considerável, porém moderado. Nesses casos, a variável não pode ser desconsiderada para nenhum dos fatores com cargas mais elevadas nesta etapa de interpretação fatorial, mesmo que um dos fatores não seja necessariamente aquele com a carga fatorial de maior valor (HAIR *et al.*, 2009).

É comum que a estrutura fatorial não seja simples, e que não haja apenas uma carga fatorial elevada para cada variável. Entretanto, é visado que cada variável tenha apenas um fator significativamente correspondente, variáveis que não atendem este critério são fortes candidatas a exclusão de novas iterações do modelo HAIR *et al.* (2009).

Após a interpretação das cargas fatoriais das variáveis resultar em uma solução fatorial, os fatores devem ser rotulados pelos especialistas conforme o conjunto de variáveis que cada fator expressa (HAIR *et al.*, 2009). As variáveis com cargas mais elevadas para cada fator devem ser dadas um peso maior na atribuição do nome ao fator (HAIR *et al.*, 2009).

A interpretação bem-sucedida dos fatores selecionados pela análise fatorial conclui uma iteração de uma análise fatorial exploratória. Análises fatoriais também podem cumprir o propósito de validação de uma interpretação fatorial pré-definida, nesses casos é feita uma análise fatorial do tipo confirmatória.

A análise fatorial confirmatória busca uma validação empírica para o modelo fatorial teórico proposto pelos pesquisadores em relação ao conjunto de variáveis escolhidas para representar os fatores. As variáveis observadas pelo instrumento de medição são indicativas de uma variável latente que corresponde a um fator. A validação de construtos feita pela análise engloba a validade discriminante e a validade convergente.

A validade discriminante corresponde à disparidade entre os fatores representados pelos itens que o indicam. A validade convergente é o quão coeso os itens são na medição do seu fator correspondente.

A validade divergente testa a insignificância estatística das variáveis aos outros fatores. Enquanto a validade convergente testa a significância estatística das variáveis indicadores aos fatores.

A análise fatorial confirmatória é uma das aplicações da técnica de modelagem de equações estruturais (SEM) . Essa técnica é identificada pelas características de estimativa de relações de dependências inter-relacionadas entre variáveis (HAIR *et al.*, 2009). Além disso, técnicas SEM se destacam pela capacidade de representar constructos latentes nas ditas relações, considerar os erros de medição nas estimativas e verificar a adequabilidade dos dados obtidos frente ao modelo teórico sendo

testado (HAIR *et al.*, 2009).

Em contraste às análises fatoriais exploratórias, as análises confirmatórias delimitam a especificação do modelo fatorial. São especificadas as relações teóricas entre variáveis do instrumento de medição e seus fatores. Após a execução da análise, é possível avaliar o grau de adequação (*goodness-of-fit*) do modelo teórico perante os dados analisados.

Análises fatoriais confirmatórias costumam ser conduzidas usando uma estimativa de máxima verossimilhança (GERBING e HAMILTON, 1996). Este método de estimativa é válido para pesquisas cujo tamanho de amostra é superior a 50, mas é recomendável que a amostra seja de 100 a 200 indivíduos (DING *et al.*, 1995).

Resultados satisfatórios de análises fatoriais confirmatórias indicam uma boa compatibilidade entre o modelo fatorial teórico e os resultados práticos. Contudo, níveis indesejados em métricas como o índice de reajuste, resultantes da análise, assinalam que a especificação há de ser refinada. Nesses casos, os resultados da análise confirmatória apontam para a necessidade de mudanças nos itens do instrumento de medição para obter resultados que se adequem mais ao modelo teórico (GERBING e HAMILTON, 1996).

Métodos de análise fatorial confirmatória não visam afirmar que o modelo teórico sob análise é o único adequado para o conjunto de variáveis de uma pesquisa. A afirmação testada por tais análises é que o modelo especificado é um dos possíveis modelos fatoriais aceito entre os modelos alternativos adequados HAIR *et al.* (2009).

É recomendável que o nível de variáveis indicadoras de um fator seja alto. Uma quantidade maior resulta em uma probabilidade menor da análise fatorial confirmatória não convergir (DING *et al.*, 1995). Em adição a outras relações estabelecidas como um tamanho de amostra e carga fatorial mais elevados reduzirem a probabilidade de uma solução divergente, foi encontrado por DING *et al.* (1995) que a quantidade e variáveis indicadores por fator tem um efeito significativo em estudos com um tamanho amostral inferior a 100. A análise fatorial confirmatória não converge na maioria dos cenários com menos de 2 variáveis por fator e uma amostra inferior a 100 (DING *et al.*, 1995). Sugere-se então uma proporção de ao menos três variáveis por fator.

As análises fatoriais confirmatórias devem ser avaliadas para verificar se os dados estão adequados para o modelo teórico. A adequabilidade do modelo é verificada primeiro mediante uma análise para a identificação de estimativas impróprias, três tipos de estimativas impróprias segundo HAIR *et al.* (2009) são:

1. Variância do erro negativa ou insignificante para algum fenômeno.
2. Coeficientes de correlação padronizados superiores a 1,0.

3. Erro padrão excessivamente alto associado com qualquer coeficiente de correlação.

Em situações nas quais nenhum dos critérios possui estimativas impróprias, há de se verificar quanto os dados se adequam ao modelo teórico (*goodness-of-fit*). A adequabilidade do modelo deve ser testada em termos gerais, bem como na adequação do modelo de medição e do modelo estrutural. Em relação ao modelo de medição, são avaliados a unidimensionalidade e a confiabilidade dos constructos (HAIR *et al.*, 2009). A verificação de adequabilidade do modelo estrutural é referente a avaliação da significância estatística dos coeficientes de estimativa (HAIR *et al.*, 2009).

A multicolinearidade, que ocorre quando variáveis independentes possuem uma correlação muito alta, pode afetar resultados de análises fatoriais confirmatórias. Não há um consenso sobre o limite do grau de correlação nesse contexto, mas correlações acima de 0,8 são alarmantes e as acima de 0,9 devem ser investigadas (HAIR *et al.*, 2009).

Uma das medidas de adequabilidade é a razão de verossimilhança usando a estatística qui-quadrado. Nesses casos, a relação com a estatística qui-quadrado é invertida, por ser desejado que a hipótese nula tradicional, de que a matriz esperada e a observada possuem diferenças estatisticamente insignificantes (FORNELL, 1983). Um poder estatístico de 0,05 é o limite mínimo usado, mas valores de 0,1 ou 0,2 devem ser usados para assegurar a insignificância estatística das diferenças entre os dados e o modelo.

O índice *goodness-of-fit* é usado para validar a adequação dos dados ao modelo. Os valores do índice variam num intervalo de 0 a 1, no qual 1 significa uma adequação perfeita ao modelo teórico. Este índice não considera os graus de liberdade do modelo e não possui um limite mínimo estabelecido para determinar se os dados se adequam. A partir desta métrica, foi criado o índice Adjusted Goodness-of-Fit (AGFI), que ajusta o índice para os graus de liberdade da análise. Um valor aceitável para a métrica de AGFI é de pelo menos 0,9 (HAIR *et al.*, 2009).

A confiabilidade composta dos constructos é calculada para cada fator a partir da carga fatorial padronizada das variáveis indicadoras, seguindo a fórmula:

$$cc = \frac{(\sum cfp)^2}{(\sum cfp)^2 + \sum e_i} \quad (3.16)$$

Na qual, *cc* se refere a confiabilidade do construto, *cfp* à carga fatorial padronizada e *e<sub>i</sub>* ao erro de medição da variável *i*, indicadora do constructo. O erro de medição é equivalente à diferença de 1,0 e a confiabilidade da variável indicadora, ou  $1,0 - (cfp)^2$  (HAIR *et al.*, 2009). A variância média extraída segue uma fórmula parecida, a diferença está no momento da elevação ao quadrado das cargas fatoriais, feita desta vez antes de somá-las. Segue a fórmula da variância extraída média:

$$cc = \frac{(\sum_i (cfp^2))}{\sum_i (cfp^2) + \sum_i e_i} \quad (3.17)$$

Um terceiro tipo de verificação de adequabilidade do modelo é o da estrutura do modelo. Nesse caso, os coeficientes de estimativa de cada variável indicadora em relação ao respectivo fator modelo teórico são testados. É feito um teste t dos coeficientes de estimativa usando seus valores e os valores de erro padrão correspondentes, o teste t em questão é referente à hipótese de que os coeficientes são estatisticamente relevantes.

# Capítulo 4

## Técnicas e Métodos Utilizados

Este capítulo discorre sobre as ferramentas utilizadas no decorrer da pesquisa. Primeiro, é discutida a metodologia DSR, usada para a geração de conhecimento científico e desenvolvimento de um artefato; ainda em relação ao desenvolvimento do modelo, há uma explicação sobre a abordagem Goal-Question-Metric, que foi usada para estruturar o modelo de forma que seus critérios sejam mensuráveis.

Em seguida, os modelos, padrões e conceitos encontrados em artigos científicos para o embasamento teórico do conteúdo do modelo são discutidos. A partir de uma perspectiva ampla, os conceitos ali elaborados encaixam-se nas áreas de avaliação do modelo original desta pesquisa: jogabilidade, experiência de usuário e motivação ao aprendizado.

Posteriormente, técnicas de desenvolvimento e validação de escalas psicométricas são apresentadas, mantendo uma ordem similar àquela na qual são aplicadas no método proposto desta pesquisa.

Esta dissertação tem em seu escopo o desenvolvimento de um modelo e um instrumento de medição voltados à avaliação de um jogo educacional digital. Um conjunto de métodos e técnicas foi utilizado para desenvolver cada um destes artefatos e em sua avaliação. Os métodos e técnicas utilizados serão discutidos nesta seção, acompanhados dos fundamentos matemáticos utilizados nas técnicas escolhidas para a análise de dados.

### 4.1 Desenvolvimento do Modelo

A parte inicial dessa pesquisa focou no desenvolvimento de um modelo de avaliação de jogos educacionais digitais. O modelo de avaliação seguiu a metodologia proposta por PEFFERS *et al.* (2007). O processo DSRM pode ser visualizado na Figura A.4. A maneira como o DSRM foi aplicado na metodologia deste trabalho se encontra no capítulo de *Técnicas e Métodos Utilizados* (Capítulo 4).

### 4.1.1 Design Science Research (DSR)

O modelo utilizou a metodologia de pesquisa *Design Science Research* (DSR), proposta por HEVNER *et al.* (2004) para a criação de um artefato a partir de uma pesquisa centrada em design. O DSR é um sistema de pesquisa frequentemente utilizado na área de Sistemas de Informação para fundamentar cientificamente, mediante conjecturas teóricas, artefatos abstratos que resolvam problemas em algum contexto específico. O tipo de conhecimento gerado por artefatos de DSR é prescritivo (GREGOR e HEVNER, 2013), os artefatos descrevem e delimitam como executar um processo. O artefato pode ter como base algo material e existente, por exemplo, no caso de Sistemas de Informação pode ser um software em operação, mas deve incluir aspectos que tornem o artefato abstrato ao ponto de generalizar para outras situações (GREGOR e HEVNER, 2013).

, O DSR tem como uma de suas funções a criação de uma solução prática. Porém, para avaliações psicométricas é essencial também a medição estruturada de sua confiabilidade e validade (DEVELLIS e THORPE, 2021). No contexto desta pesquisa, o Design Science Research (DSR) e o Goal-Question-Metric (GQM) desempenham papéis complementares, pois o GQM visa justamente a criação de artefatos operacionais e quantificáveis (CALDIERA e ROMBACH, 1994).

### 4.1.2 Goal-Question-Metric (GQM)

Tradicionalmente usada para a avaliação de software, a abordagem Goal-Question-Metric (GQM) para o desenvolvimento de modelos de medição segue uma perspectiva *top-down* (CALDIERA e ROMBACH, 1994). Isto é, o processo de desenvolvimento inicia-se com a definição de um objetivo de medição (SOLINGEN e BERGHOUT, 1999), o objetivo serve como um guia para os níveis menos abstratos do modelo de medição.

Após a definição do nível conceitual, mediante o objetivo de medição, são elaboradas as questões que compõem o modelo. As questões encontram-se no nível operacional e decompõem o objetivo em uma série de características ou perspectivas de como avaliá-lo (CALDIERA e ROMBACH, 1994). Tais questões têm como finalidade a redução do nível de abstração inerente a um objetivo de medição, o que torna o cumprimento do objetivo em algo mais interpretável. Passa a ser determinado que o objetivo foi alcançado quando o fenômeno, cujas características estão sendo medidas, tenha as respostas esperadas às questões elaboradas.

Finalmente, após a definição e refinamento das questões, são escolhidas as métricas que tornarão as respostas das questões em algo tangível e quantitativo. As métricas devem conter informação suficiente para responder às questões de forma satisfatória e também devem ser elencados fatores que potencialmente influenciam



os valores das respostas.

# Capítulo 5

## Trabalhos Relacionados

### 5.1 Trabalhos Relacionados

Esta seção discute trabalhos relacionados ao desta pesquisa. Os trabalhos relacionados foram encontrados durante uma revisão da literatura. Neste capítulo, primeiro há uma breve apresentação de abordagens diferentes daquela utilizada nesta pesquisa, que utilizou um instrumento psicométrico para avaliar a reação de alunos. Posteriormente, são apresentados dois modelos de avaliação de jogos educacionais que também dependeram de formulários psicométricos para a avaliação; são elencados pontos em comum e divergências entre estes trabalhos. A seção termina com uma tabela que dispõe uma série de características dos trabalhos analisados e o desta dissertação.

A revisão foi feita durante a primeira iteração do ciclo de design e rigor do DSR, com o intuito de acrescentar ao conhecimento científico sobre a avaliação de jogos educacionais digitais. BAUMEISTER e LEARY (1997) coloca a amplificação do conhecimento teórico sobre um fenômeno como um dos objetivos principais a serem atingidos por uma revisão da literatura.

A revisão foi feita com o intuito de encontrar diferentes maneiras de avaliar jogos educacionais digitais. Após uma revisão inicial, optou-se por uma avaliação mediante um instrumento psicométrico. A partir disso foram buscados conceitos importantes à avaliação de JEDs que pudessem ser avaliados mediante uma autoavaliação da reação do aluno. Os conceitos fundamentais ao modelo encontram-se no capítulo anterior (Capítulo 3 ).

Este capítulo dedica-se principalmente à comparação com outros modelos de avaliação de JEDs cuja avaliação foi feita por formulários psicométricos, mas antes será brevemente discutido outras maneiras de avaliar os jogos encontrados durante a revisão.

Um dos métodos encontrados foi de uma avaliação interna ao jogo incluída na

avaliação holística proposta por OLIVEIRA e ROCHA (2021). Nesta abordagem, dados de pontuação e registros de interações do jogador com o sistema são usados para avaliar o desempenho do jogador (OLIVEIRA e ROCHA, 2021). Além disso, por se tratar de uma avaliação de jogos sérios, registros de atuação de erros e acertos relacionados ao conteúdo sendo ensinado também são avaliados (OLIVEIRA e ROCHA, 2021).

Outra abordagem interessante encontra-se em LEE e HAO (2015), no qual uma avaliação da reação do aluno ao jogo é medido usando escalas, desenvolvidas em outros estudos, para a motivação de aprendizado, seguindo o modelo ARCS de KELLER (1987), e a diversão. Além disso, uma avaliação do aprendizado dos alunos foi feita usando avaliações pré-teste e pós-teste para medir a efetividade do jogo no aprendizado do aluno (LEE e HAO, 2015). Para avaliar o efeito no aprendizado, os alunos foram separados em um grupo experimental, que aprendeu mediante um JED, e um grupo de controle, cuja atividade didática foi um vídeo sobre o mesmo assunto.

Dois modelos de avaliação de jogos educacionais se sobressaem atualmente na literatura: o MEEGA/MEEGA+ SAVI *et al.* (2011) PETRI *et al.* (2019) , o EGameFlow FU *et al.* (2009) . O objetivo comum de avaliar jogos educacionais leva a uma semelhança conceitual em diversos pontos. Ambos os modelos contemplam critérios relacionados a: desafio, aprendizagem percebida, interação social e atenção.

Outra semelhança é que desenvolvem um formulário para avaliar a validade e confiabilidade do instrumento. O processo de desenvolvimento, medição e avaliação dos formulários que acompanham estes modelos tem como referência (DEVELLIS e THORPE, 2021). O guia sobre desenvolvimento de escalas em questão será discutido na próxima seção, sobre as técnicas aplicadas nesta pesquisa.

Ambas as pesquisas também utilizaram métodos de validade de conteúdo para assegurar que os itens de seus respectivos formulários refletissem o objetivo de medição e os fenômenos que o descrevem. Entretanto, diferentes métodos foram seguidos; os dois modelos desfrutaram de uma revisão de especialistas para validar o conteúdo, mas apenas o EGameFlow usou um pré-teste para validar o conteúdo do formulário.

As divergências entre os modelos podem ser explicadas tanto pelo contexto quanto pelas fontes que os fundamentam. O MEEGA+ possui critérios provenientes de modelos de motivação de design motivacional, e.g. ARCS (KELLER, 1987), e de experiência de usuário, e.g. usabilidade, provenientes do ISO/IEC 25000. Em contrapartida, o EGameFlow é amplamente baseado no modelo GameFlow (SWETSER e WYETH, 2005), evidente pela semelhança no nome dos modelos. Outros critérios, como clareza do objetivo, autonomia, imersão, entre outros que se originam no conceito de fluxo (CSIKSZENTMIHALYI, 2008) fazem-se presentes no EGameFlow.

Um resumo comparativo entre os frameworks do EGameFlow, MEEGA+, e o modelo desenvolvido nesta pesquisa encontra-se na tabela abaixo (5.1).

Tabela 5.1: Comparação entre esta pesquisa e trabalhos relacionados

Área de Avaliação	Modelo Educacional de Jogos Educacionais Digitais	MEEGA+	EGameFlow
Objetivo	Avaliação da motivação de alunos ao aprendizado mediante jogos educacionais digitais.	Avaliação da qualidade, percebida por alunos, de jogos educacionais voltados ao ensino superior de computação.	Avaliação da diversão/satisfação de alunos em relação a jogos educacionais.
Escopo	Critérios de motivação ao aprendizado delimitados nas teorias de KELLER (1987); MALONE (1981) com o acréscimo de critérios de diversão e capacidade de aprendizado.	O escopo engloba diversas dimensões de jogos educacionais digitais. São cobertos critérios de: motivação ao aprendizado, provenientes de KELLER (1987); usabilidade, seguindo o padrão ISO/IEC 25010; além de critérios adicionais de diversão, interação social e aprendizagem percebida.	Avaliação de jogos digitais pela perspectiva da sensação de fluxo do jogador e de incremento percebido no conhecimento.
Validade de Conteúdo	Revisão de especialistas sobre a essencialidade dos itens referente à avaliação de jogos educacionais digitais, usando lógica Fuzzy para determinar se houve consenso. Há também uma avaliação binária sobre a concisão e ambiguidade dos itens e espaço para comentários sobre inclusão, remoção ou modificação de itens.	Revisão de especialistas sobre a qualidade dos itens seguindo os critérios de: clareza, relevância, consistência e completude.	Revisão e modificação da escala por especialistas; uso de dados obtidos em um pré-teste para identificar itens que devem ser removidos e modificados.

Método de Avaliação	Formulário com formato de resposta de escala Likert de cinco pontos.	Formulário com formato de resposta de escala Likert de cinco pontos	Formulário com formato de resposta de escala Likert de sete pontos.
Validação do Instrumento de Medição	A confiabilidade do instrumento de medição foi determinada usando três métricas de consistência interna: o Alfa de Cronbach, a confiabilidade por Duas Metades e fórmula de profecia de Spearman-Brown. A validade de constructos é feita por meio de análises fatoriais confirmatórias ou exploratórias.	A confiabilidade dos itens do formulário foram calculados usando o Alfa de Cronbach para a consistência interna. A validade de construto do instrumento foi averiguada mediante uma análise fatorial exploratória.	A confiabilidade foi atestada com a métrica Alfa de Cronbach para consistência interna e de teste-reteste, que usou o coeficiente de correlação intraclasse como indicador. Foram medidos as validades de: construto, divergente, convergente e relacionada a critérios.
Comparação entre grupos de respondentes	Uso do teste Kruskal-Wallis para identificar diferenças na distribuição de respostas do formulário para diferentes grupos socio-demográficos.		Uso de análise de variância simples para identificar diferenças no grau de diversão/satisfação conforme o nível de experiência em diversos jogos.

# Capítulo 6

## Método Proposto

Este capítulo detalha o processo seguido e o método proposto para esta dissertação. O processo engloba dois ciclos de rigor e de conhecimento do DSR. A metodologia segue mais especificamente o processo delineado por PEFFERS *et al.* (2007), que pode ser visualizado na Figura A.4.

Em uma iteração inicial seguindo o DSRM, foi identificado o problema e a motivação da pesquisa, expostos no capítulo de introdução. Em seguida, esta iteração seguiu com a identificação de objetivos iniciais para solucionar a questão-problema da pesquisa. Então, houve uma etapa de desenvolvimento do artefato e, finalmente, comunicação sobre seus resultados.

A segunda iteração seguindo o processo DSRM, se iniciou com a identificação de novos objetivos para solucionar a questão-problema. Em seguida, o artefato foi refinado para manter apenas critérios considerados relevantes. A etapa de *demonstração* do DSRM equivale à de *aplicação do formulário* nesta pesquisa, por ser nela que o artefato é aplicado em seu devido contexto. A etapa de *avaliação* do DSRM também se faz presente nessa segunda iteração, pois a efetividade do formulário é avaliada mediante a sua confiabilidade e validade na Seção 6.5.

Neste capítulo, é delineado o processo de desenvolvimento do modelo teórico de avaliação de jogos educacionais digitais. Em seguida, o processo de desenvolvimento do formulário de avaliação é apresentado. A aplicação do formulário de avaliação também é discutida pelo capítulo. Ademais, duas instâncias de análise de dados são aprofundadas: uma para validar o conteúdo do formulário, anterior à sua aplicação, e outra posterior à aplicação, que mede a sua confiabilidade e a validade de construtos.

Ao longo do capítulo, são apresentados diagramas no padrão BPMN 2.0, o *Business Process Model Notation* (BPMN). O BPMN é um modelo de anotação que ilustra processos e subprocessos de forma clara e padronizada às partes interessadas. O objetivo do BPMN é detalhar processos necessários para auxiliar na análise, implementação, implantação e monitoramento de uma atividade (CHINOSI e TROMBETTA, 2012).

No contexto desta pesquisa, o processo descrito é o de desenvolvimento do modelo de avaliação de jogos digitais educacionais e do formulário que o acompanha (Figura B.4). O processo se inicia com o desenvolvimento do modelo teórico de avaliação de jogos digitais e resulta em uma versão aprimorada do modelo e de um instrumento de medição baseado nele.

O modelo e o instrumento de medição devem ter a sua validade empiricamente atestada. No caso do instrumento de medição, há de ser evidenciado que o mesmo mede algo de forma confiável. Essas questões foram abordadas nas análises de dados do experimento. Uma delas foi conduzida antes da aplicação do formulário para validar o conteúdo dos itens que o compõem. A outra análise fez as demais revisões de validade e de confiabilidade fundamentais a esta pesquisa.

## 6.1 Desenvolvimento do Modelo

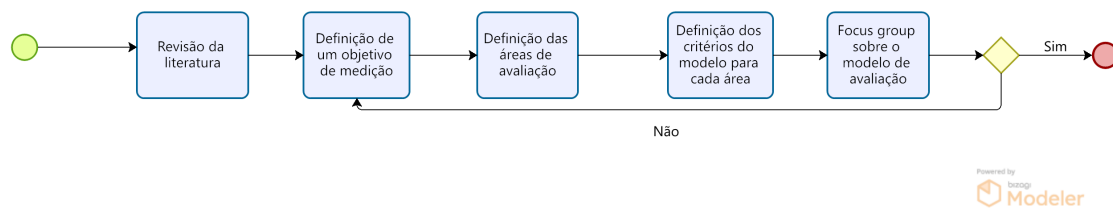
Esta seção elabora sobre o desenvolvimento do modelo teórico de avaliação de jogos educacionais digitais. O resultado desta etapa foi a versão inicial do modelo, anterior às etapas de validação.

O desenvolvimento do modelo corresponde ao primeiro ciclo de conhecimento de design e de rigor do DSR. Essa dissertação segue o *Design Science Research* (DSR), proposta por HEVNER *et al.* (2004) com a criação de um artefato a partir de uma pesquisa centrada em design. O modelo desenvolvido nesta pesquisa é fruto de duas iterações de ciclos de design e de conhecimento (Figura 6.1).

O desenvolvimento do modelo iniciou-se com uma revisão da literatura. A revisão da literatura foi voltada em parte a trabalhos relacionados, outros modelos de avaliação de jogos ou de materiais instrucionais. Em outra parte, foram revisados fenômenos de interesse, modelos e componentes voltados à qualidade da experiência de jogadores, experiência de usuário de sistemas de informação e à motivação de alunos perante materiais educacionais, em geral. Após a etapa de revisão da literatura, iniciou-se o desenvolvimento do modelo de reação a jogos educacionais digitais.

O modelo foi desenvolvido de forma top-down inspirado na etapa de Desen-

Figura 6.1: Processo iterativo do desenvolvimento do modelo de avaliação de jogos educacionais digitais.



volvimento do método GQM (Goal-Question-Metric) (SOLINGEN e BERGHOUT, 1999). O processo de desenvolvimento do GQM, idealizado para a medição e avaliação de software em ambientes corporativos, foi aplicado nesta pesquisa da seguinte forma:

1. O objetivo da medição foi definido como a reação de alunos a jogos educacionais digitais.
2. Foi estabelecida a classificação das questões em dois níveis hierárquicos: áreas de avaliação e critérios.
3. Para o nível hierárquico menos abstrato, componentes que receberam a denominação de *subcritérios* no contexto desta pesquisa serviram como métricas. Os subcritérios foram atribuídos definições concisas e claras.

As áreas de avaliação abordam de forma mais abrangente os aspectos relevantes, porém díspares, que devem ser considerados ao avaliar a experiência de um jogo educacional digital. A ideia é que estas áreas de avaliação representem fatores ortogonais na avaliação de jogos educacionais digitais. Teoricamente, não deve haver uma correlação significativa nas respostas aos itens que representem estes fatores ortogonais.

Cada uma das áreas de avaliação é composta por diversos critérios que definem pontos importantes para a avaliação de tal área de avaliação. Por último, os subcritérios são responsáveis pela quantificação dos conceitos abstratos e gerais definidos pelos critérios e áreas de avaliação escolhidos para o modelo. Esta quantificação do subcritério ocorre mediante um item que represente o seu conteúdo no formulário, instrumento de medição desta pesquisa.

A construção do modelo de avaliação usou definições de qualidade de software, estabelecidas pelo ISO 9126 (ISO/IEC 9126) e de um modelo de medição e avaliação de qualidade de software definido em ISO 25000 (ISO/IEC 25000).

Com a conclusão da versão original do modelo de avaliação, foi publicado o artigo VAZ e XEXÉO (2022). Na metodologia DSRM de PEFFERS *et al.* (2007), esta etapa corresponde à *comunicação*, mediante uma publicação acadêmica. Depois, o processo foi reiterado desde a etapa de definir objetivos para solucionar o problema de pesquisa.

O desenvolvimento do modelo dessa pesquisa delimita conceitos aplicáveis a avaliação de JEDs, englobando as áreas de motivação do aprendiz, experiência de usuário e jogabilidade. Estas áreas são entendidas pelo autor como importantes para a medição da reação a jogos educacionais digitais. Entretanto, para possibilitar uma avaliação da reação do aluno ao jogo educacional digital, há de ser definido



algum instrumento apropriado para a medição direta dos conceitos encapsulados pelo modelo.

A aplicação de um instrumento que meça os conceitos explorados pelo modelo também é necessária para a sua validação. Portanto, a elaboração do formulário e a verificação se o item do formulário é representativo do subcritério do modelo são descritas na próxima seção.

## 6.2 Desenvolvimento do Formulário

Esta seção discorre sobre o processo de elaboração do formulário de avaliação, baseado no modelo de avaliação desenvolvido na seção anterior (6.1).

O ciclo inicial de design e rigor desta pesquisa resultou em um artefato com uma fundamentação teórica sólida. O modelo se baseia em conjecturas validadas por outras pesquisas e elabora definições particulares ao modelo proposto. No entanto, o modelo por si só não consegue realizar a medição dos construtos que o compõem.

O modelo há de ser acompanhado de algum instrumento que possibilite a medição de seus fenômenos. No contexto desta pesquisa, o instrumento de medição deve ser elaborado de tal maneira que os itens indiquem agregadamente a reação do aluno ao jogo. O desenvolvimento de escalas psicométricas é um processo frequentemente usado para a medição de fenômenos sociais e psicológicos que normalmente se originam de conjecturas teóricas (DEVELLIS e THORPE, 2021).

Dada a natureza da pesquisa e do modelo, o instrumento de medição mais apropriado para a avaliação da reação dos alunos é o desenvolvimento e operação de um formulário psicométrico. O processo de desenvolvimento do formulário está exposto na Figura B.6. O instrumento elaborado por essa pesquisa mede variáveis que não podem ser observadas diretamente, como as emoções e as expectativas dos jogadores. Conseqüentemente, a medição dessas variáveis latentes demanda um questionário ou algum outro procedimento de pesquisa cognitiva.

O questionário desenvolvido na escala é transversal, por medir uma situação de um subconjunto de uma população em um momento específico (KITCHENHAM e PFLEEGER, 2002a). O formulário foi administrado aos respondentes imediatamente após o término da atividade didática, cujo jogo educacional digital será a peça central, para medir a reação do aluno.

Uma das preocupações quanto à escala desenvolvida nessa pesquisa é do seu tamanho, uma vez que o modelo original possui um total de 48 itens. Posto que há um número expressivo de itens, mesmo que cada subcritério possua apenas um item correspondente na escala, o número de respostas esperado de um aluno é acentuado. Isto é um problema porque, segundo KITCHENHAM e PFLEEGER (2002a), formulários excessivamente extensos podem levar a uma taxa de desistência mais

elevada.

Para solucionar este problema, os itens candidatos a entrarem no formulário de avaliação passaram por uma rodada de validação de conteúdo. Essa validação de conteúdo filtrou os itens após a identificação por especialistas em educação ou no desenvolvimento de jogos de quais itens são essenciais para a avaliação de jogos educacionais.

Além de reduzir a quantidade de itens pertencentes ao formulário de avaliação, a validação do conteúdo é necessária para assegurar que o formulário cubra os critérios necessários para medir o fenômeno desejado. O formulário elaborado nesta pesquisa faz medições subjetivas, dependendo da opinião do respondente para a obtenção dos dados. Portanto, é esperado que o formulário seja validado enquanto instrumento de medição e só assim é possível afirmar que o formulário de fato cumpre o seu objetivo.

Foi incluída uma pergunta em formato de escala Likert (ALBAUM, 1997) correspondente a cada item do modelo de avaliação de reação dos jogadores. O método de medição de escala Likert foi escolhido para essa pesquisa por ser um modelo simples de se aplicar e de conseguir uma boa quantidade de respondentes e de conseguir estimativas confiáveis quanto às opiniões, atitudes ou sentimentos dos respondentes (NEMOTO e BEGLAR, 2013). Outra vantagem é a familiaridade dos respondentes com este formato de medição, visto que é o tipo de escala psicométrica mais usado em questionários de autoavaliação (WAKITA *et al.*, 2012).

Não há um consenso quanto ao número correto de alternativas para as respostas às escalas Likert (WAKITA *et al.*, 2012). Entretanto, MILLER (1956) argumenta haver uma limitação na capacidade humana de processar informações em aproximadamente sete pedaços de uma vez. Aplicado ao desenvolvimento de escalas psicométricas, essa lógica sugere não haver necessidade em ultrapassar o limite de sete opções (TAHERDOOST, 2019).

A escolha do autor para esta pesquisa foi por uma escala Likert usada com cinco alternativas. Assim, foram oferecidos dois graus de intensidade para respostas positivas e negativas e uma opção neutra aos respondentes.

As etapas de inclusão de itens de validação e de administração dos itens para uma amostra de desenvolvimento do guia de desenvolvimento de escalas encontrado em DEVELLIS e THORPE (2021) não entraram na versão adaptada a esta pesquisa. As demais etapas foram contempladas e detalhadas neste capítulo de método proposto. Excepcionalmente, a última etapa, de otimizar o tamanho da escala, foi incluída como uma proposta de trabalho futuro.

## 6.2.1 Validade

### Validade de Conteúdo

A avaliação das inferências feitas pelo estudo em relação ao problema de pesquisa só pode ser considerada válida se antes mostrar-se adequada ao passar por uma validação interna. No paradigma qualitativo de pesquisa, é a experimentação que identifica as variáveis, ao explorar o fenômeno sob análise (GOMES e GOMES, 2020). Então, devem ser estabelecidos critérios e métodos para questionar a relevância dos itens, além de sua concisão e integridade. Estes questionamentos e o refinamento e desenvolvimento do formulário a partir disso resultam em uma avaliação da reação a jogos educacionais digitais mais próxima da ideal.

O conteúdo dos itens da escala deve refletir o mais próximo possível do conteúdo definido para o constructo no escopo e contexto da escala. Uma forma de garantir a validade do conteúdo é pedir que outros especialistas analisem os itens e os comparem aos conceitos medidos pela escala.

Os critérios presentes nesse modelo foram devidamente revisados por especialistas, em um *focus group*, e comparados aos itens candidatos a entrar na escala. *Focus groups* contribuem para a validade de instrumentos (KITCHENHAM e PFLEEGER, 2002a) e podem identificar maneiras de medir o construto de interesse que não foram inicialmente previstas pelo autor. Outros benefícios de *focus groups* incluem identificar componentes ambíguos, faltantes ou que aparentam ser inválidos para a escala em questão.

Ao final do focus group, foi feita uma pergunta aos especialistas para identificarem critérios importantes para a avaliação de jogos educacionais digitais, mas que não haviam sido incluídos originalmente pelo autor desta pesquisa.

Uma carta de apresentação foi incluída nos formulários de validade de conteúdo. A carta de apresentação seguiu as recomendações incluídas em GRANT e DAVIS (1997), cumprindo algumas finalidades como:

- a descrição do objetivo do formulário
- a razão e os critérios pelos quais o especialista foi escolhido
- a importância que a contribuição do especialista terá na pesquisa ao responder

Além disso, o conceito geral do domínio do modelo foi descrito na carta de apresentação, enquanto descrições sobre os critérios foram situadas na mesma seção que os seus itens correspondentes.

A seção principal do formulário de validade de conteúdo foi disposta após a carta de apresentação. O formulário de validade de conteúdo seguiu um formato de pergunta elaborado por (LAWSHE, 1975) e exemplificado na Figura 6.2. Os

Figura 6.2: Exemplo do formulário de validade de conteúdo, usado para calcular o IVC)

Validade de Conteúdo dos Itens			
O quão relevante você considera os itens abaixo para a medição de motivação ao aprendizado? *			
	Essencial	Útil, mas não essencial	Desnecessário
Houveram momentos de descontração durante o jogo. (Critério: Humor)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Senti compatibilidade entre as demandas do jogo e as minhas habilidades para superá-las. (Critério: Demanda de Habilidade)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores. (Critério: Experiência)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
O jogo não me causou nenhum momento em que fiquei fisicamente desconfortável. (Critério: Conforto)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo. (Critério: Curiosidade Cognitiva)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

especialistas também foram indagados quanto a ambiguidade e concisão dos itens, seguindo o formato da Figura 6.3.

A análise de dados referente à validade de conteúdo foi detalhada na Seção 6.3. Após a validação do conteúdo, os itens considerados essenciais pelos especialistas foram incluídos no formulário de avaliação de jogos educacionais digitais, cuja estrutura será detalhada a seguir.

### **6.2.2 Estrutura do Formulário de Avaliação**

O formulário começou com uma breve introdução seguida de um Termo de Consentimento Livre e Esclarecido. O papel da introdução foi motivacional, encorajando o aluno a participar da pesquisa. Respondentes sentem-se motivados a participarem de pesquisas quando entendem a sua utilidade (KITCHENHAM e PFLEEGER, 2002a). Portanto, a introdução da carta de apresentação abordou alguns pontos destacados por KITCHENHAM e PFLEEGER (2002a) como:

- o propósito do estudo
- a relevância do estudo para os respondentes
- a importância que cada participação individual tem na pesquisa
- os critérios por trás da seleção dos participantes

Logo após o Termo de Consentimento Livre e Esclarecido foram disponibilizadas perguntas para identificar o perfil sociodemográfico do respondente e os seus hábitos relacionados a jogos. Estas perguntas podem providenciar percepções sobre a maneira como diferentes fatores ligados ao perfil ou hábitos das pessoas influenciam a sua reação a um jogo educacional digital. Tais relações entre as questões sociodemográficas e de hábitos de jogo foram exploradas na análise de variância, descrita na seção de análise de dados (Seção 6.5.5). A inclusão dessas seções de perguntas aumenta o tamanho do questionário, mas são importantes fontes de conhecimento para esta pesquisa, apesar de não contribuírem ao seu objetivo cardinal.

Por último, há a seção principal, com os itens de medição da avaliação do jogo educacional digital cujos alunos jogaram em uma atividade anterior. Todas as perguntas desta seção foram obrigatórias e seguiram o mesmo formato de resposta.

## **6.3 Análise de Dados do Desenvolvimento do Formulário**

Esta seção tem como escopo a análise de dados anterior à aplicação do formulário de avaliação de jogos educacionais digitais, cujo processo percorrido está exposto na

Figura B.7. A análise anterior à aplicação do formulário resume-se à validação de conteúdo do formulário.

A validade de conteúdo assegura que o formulário é composto de itens relevantes ao objetivo de medição. A validade de conteúdo de formulários psicométricos é popularmente avaliada usando o método criado por LAWSHE (1975), que tem como métrica o índice de validade de conteúdo (IVC).

O IVC tem um valor positivo quando mais de metade dos especialistas avaliam um determinado item como “essencial”, indicando que o item tem algum grau de validade de conteúdo (LAWSHE, 1975). LAWSHE (1975), que propôs a métrica do IVC originalmente, divulgou no mesmo artigo uma tabela com o IVC mínimo com significância estatística dado que  $\alpha = 0,05$ . É interessante notar que pelo menos 8 especialistas devem opinar sobre os itens para que o mínimo não seja de 0,99 (Tabela 6.1), um valor que não permite nenhum nível de discordância entre os especialistas sobre o quão essencial é um item em questão.

A tabela de valores críticos para a aceitação de um item como válido foi questionada posteriormente por não explicar suficientemente a sua fundamentação e apresentar um valor interpretado por WILSON *et al.* (2012) como inconsistente. Conforme a tabela, um item revisado por 9 especialistas exige um IVC satisfatório mínimo superior àquele exigido quando apenas 8 respostas foram obtidas. Por conseguinte, os valores de IVC satisfatório mínimo foi recalculado por WILSON *et al.* (2012) (TABELA 6.2). Itens que não cumprem o mínimo estipulado pela tabela são desconsiderados da versão final da escala por não terem a validade de conteúdo almejada.

Tabela 6.1: Valor mínimo aceitável para a validade de conteúdo ( $\alpha = 0.05$ ). Tabela original de LAWSHE (1975)

Número de especialistas	IVC Satisfatório Mínimo	Proporção de Especialistas Satisfatória Mínima
5	0,99	5
6	0,99	6
7	0,99	7
8	0,75	7
9	0,78	8
10	0,62	9

Tabela 6.2: Valor mínimo aceitável para a validade de conteúdo ( $\alpha = 0,05$ )

Número de especialistas	IVC Satisfatório Mínimo	Proporção de Especialistas Satisfatória Mínima
5	0,877	5
6	0,800	5
7	0,741	6
8	0,693	6
9	0,653	6
10	0,620	7

Além da medição da validade de conteúdo, representadas pelas métricas de IVC

por item e do valor agregado para o instrumento em sua totalidade, os especialistas foram encorajados a responder com comentários em relação a:

- a relevância dos itens
- a identificação de itens ambíguos ou prolixos
- sugestões de itens que agregariam à medição do fenômeno

Essas observações qualitativas são sugeridas na etapa 4, *Have Initial Item Pool Reviewed by Experts*, do guia de desenvolvimento de escalas de DEVELLIS e THORPE (2021).

A validação de conteúdo de escalas usadas em pesquisas da área de Sistemas de Informação é fortemente recomendada (STRAUB *et al.*, 2004). A medição da validade de conteúdo foi feita mediante um formulário que perguntou sobre a relevância do item para a medição do constructo sendo avaliado. A resposta a esta pergunta foi usada para o cálculo do IVC. Em seguida, foi apresentado um espaço para comentários sobre o item e uma pergunta com resposta em formato binário para identificar se o item é conciso e claro.

Em paralelo à avaliação de validade de conteúdo usando as métricas tradicionais de IVC, foi feita uma avaliação da validade de conteúdo dos itens do formulário por meio de um método de determinação de consenso usando lógica Fuzzy. Esta avaliação, usando métodos Fuzzy, utilizou-se também das respostas às perguntas feitas para o cálculo do IVC, que seguem o formato de múltipla escolha entre as opções: *essencial, útil, mas não essencial e desnecessário*.

Foi determinado se houve um consenso entre os especialistas sobre a relevância do item perante o constructo medido pelo formulário. Itens cujo consenso foi de essencialidade foram incluídos no formulário desta pesquisa; os demais foram descartados. Os resultados do método Fuzzy foram mais satisfatórios para a continuidade da pesquisa. Este método foi o considerado para a validação de conteúdo do formulário avaliado nessa pesquisa.

O cálculo do consenso para a lógica Fuzzy se inicia com a construção de uma matriz de confusão que indica a preferência de uma alternativa em relação às outras. Os valores da matriz variam de 0 a 1, no qual 1 indica que houve um consenso absoluto na preferência da alternativa na linha X em relação à alternativa da coluna Y. A diagonal principal da matriz é composta por zeros, pois não há como determinar uma preferência entre a mesma alternativa consigo mesma.

No universo de alternativas  $X = x_1, x_2, \dots, x_n$  os valores da matriz de confusão, de relação recíproca, tem as propriedades (BEZDEK *et al.*, 1978):

$$r_{ii} = 0; 1 \leq i \leq n \quad (6.1)$$

Figura 6.3: Exemplo da avaliação de concisão e clareza dos itens.

Você considera os seguintes itens ambíguos ou prolixos? \*

	Sim	Não
Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	<input type="radio"/>	<input type="radio"/>
O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre o jogo.	<input type="radio"/>	<input type="radio"/>
Senti que era possível ter um desempenho legal no jogo mediante o meu esforço.	<input type="radio"/>	<input type="radio"/>
Ao completar tarefas interessantes, o jogo me recompensava inesperadamente, variando de alguma forma na recompensa que era dada.	<input type="radio"/>	<input type="radio"/>
Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados para avaliar o meu desempenho.	<input type="radio"/>	<input type="radio"/>
A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	<input type="radio"/>	<input type="radio"/>
As tarefas que considerei entediantes eram recompensadas de forma evidente e previsível.	<input type="radio"/>	<input type="radio"/>



$$r_{ij} + r_{ji} = 1; 1 \leq i \neq j \leq n \quad (6.2)$$

No contexto desta pesquisa, os valores da matriz usada para calcular o grau de consenso são a proporção do número de especialistas que escolheram a alternativa  $i$ , perante o total:

$$r_{ij} = \frac{a_i}{\sum_{j=1}^n a_j} \quad (6.3)$$

As métricas de grau de certeza médio e de grau *fuzzy* médio são calculados com esses valores de preferência. A definição de consenso escolhida para validar o conteúdo dos formulários foi a do Tipo II, na qual há uma preferência clara entre a alternativa preferencial, mas não entre as alternativas secundárias. Assim, foram selecionados os itens cuja distância ao consenso foi inferior ou igual àquele definido pela métrica de consenso  $M_1^*$ .

$$m(R) = 1 - \left(\frac{2}{n}\right)^{\frac{1}{2}} \quad (6.4)$$

$$m(R) = 1 - \left(\frac{2}{3}\right)^{\frac{1}{2}} \quad (6.5)$$

$$m(R) = 0,184 \quad (6.6)$$

O consenso observado entre os especialistas foi testado mediante um teste de permutação, também conhecido como teste de aleatorização. A hipótese nula no teste é que os especialistas não chegariam a um consenso. Concluir-se-ia então que haveria uma completa ambivalência, resultando em uma distância ao consenso de 1,0.

Antes de sua realização, o teste de permutação em questão dependia de uma análise de potência. A análise de potência foi conduzida para identificar o tamanho de amostra mínimo para determinar que a magnitude do efeito desejado tenha poder estatístico suficiente. O módulo escolhido para a análise de potência desta pesquisa foi o TTestIndPower do "statsmodels", que calcula o poder estatístico para um teste t com duas amostras independentes (SEABOLD e PERKTOLD, 2010).

A análise de potência foi feita assumindo um poder estatístico desejado de 0,9 e um nível de significância,  $\alpha$ , de 0,05. Algumas referências sugerem um limite mais abrangente, aceitando um poder estatístico que atinja 0,8 (HAIR *et al.*, 2009). O tamanho do efeito para cada item é a diferença entre o valor observado e a hipótese nula.

A manutenção de um nível adequado de especificidade requer uma amostra maior

caso a tolerância para erros do Tipo 1 diminua (HAIR *et al.*, 2009). Entretanto, deve-se ter o cuidado de não permitir que a amostra fique demasiadamente grande, pois uma amostra grande implica em um poder estatístico que leva a uma sensibilidade a resultados que podem não ter significância prática (HAIR *et al.*, 2009).

Após a análise de potência, a hipótese nula foi testada com o teste de permutação, no qual o valor observado do grau de consenso para cada item foi comparado à distribuição nula. A distribuição nula foi composta de matrizes recíprocas de consenso com dados gerados de forma aleatória. A partir dessas matrizes, o grau de certeza e a distância ao consenso foram calculados e o resultado da distância ao consenso foi comparado ao valor observado. A proporção de valores tanto quanto ou mais extremos que o valor observado é o valor de  $p$ . A aceitação da hipótese alternativa, a afirmação de que o grau de consenso é estatisticamente relevante, acontece caso o valor de  $p$  for menor do que o nível de significância do experimento, 0,05.

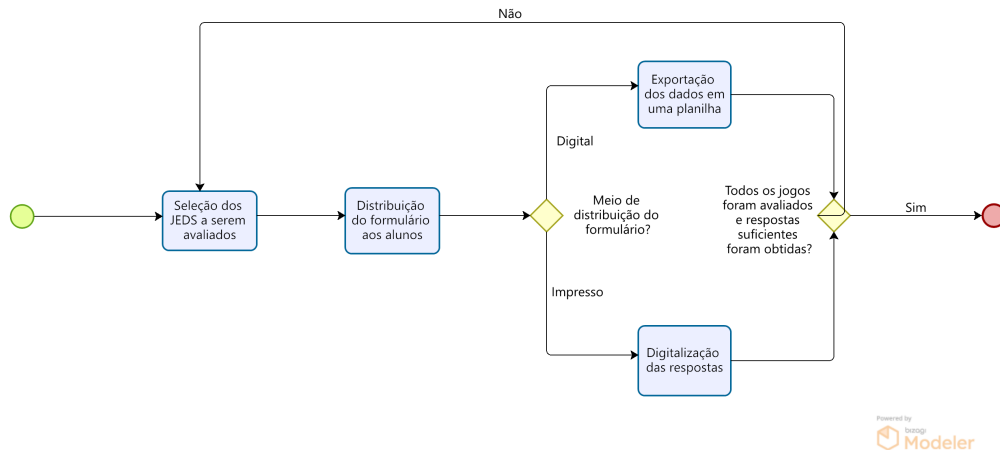
## 6.4 Aplicação do Formulário

Após assegurar que o formulário estava composto apenas de itens essenciais à avaliação de JEDs, o formulário foi distribuído e os dados devidamente coletados, seguindo o processo descrito na Figura 6.4. Um dos principais desafios na hora da aplicação de um questionário é recrutar participantes. Há a necessidade de um conjunto suficientemente numeroso de pessoas aptas a participar na pesquisa para que se tenha alguma validade estatística que não seja comprometida por uma taxa de desistência (EMMERICH e BOCKHOLT, 2016), visto que a participação no estudo é voluntária. Foi definido que a quantidade de respondentes deveria ser de pelo menos 5 vezes a quantidade de itens na escala, seguindo as orientações de HAIR *et al.* (2009). Outro limite importante a ser alcançado é de ter pelo menos 50 respostas (HAIR *et al.*, 2009). Entretanto, o limite de uma proporção de 5 vezes mais respondentes que itens é mais restritivo nesse caso, visto que a escala possui 13 itens, resultando em pelo menos 65 respondentes.

O questionário de validação de conteúdo foi distribuído somente por meios digitais. Já o formulário de avaliação do jogo foi providenciado aos respondentes de forma impressa e digital. Formulários digitais tendem a ter taxas de respostas inferiores àqueles distribuídos presencialmente (LEFEVER *et al.*, 2007). Possivelmente, por serem preenchidos de forma assíncrona e à distância, os respondentes sentem-se menos pressionados ou motivados a respondê-lo. Um benefício da aplicação de formulários digitais foi de ordenar os itens de forma aleatória, assim evita-se que a ordem aos quais os itens foram apresentados influencie o padrão de respostas, atuando como uma fonte de erro.

O conhecimento e as experiências que o respondente tem em relação ao contexto

Figura 6.4: Processo de aplicação do formulário de avaliação de jogos educacionais digitais.



da pesquisa pode influenciar os resultados e deve ser considerado (EMMERICH e BOCKHOLT, 2016). Deste modo, foi incluída uma seção de *Hábitos de Jogar* no formulário distribuído aos alunos para identificar se há padrões de resposta diferentes entre pessoas com uma considerável experiência em jogos digitais e àquelas que não tem este *hobby*.

Esta pesquisa contempla a avaliação de múltiplos jogos desenvolvidos pelo laboratório LUDES. Foi distribuído um formulário por jogo, com especificidades àquele jogo e avaliação na carta de apresentação. O formato das questões de dados demográficos, hábitos de jogo e avaliação do jogo foi homogêneo. O pré-requisito, explícito no convite para a participação e na carta de apresentação, é que o jogador tenha jogado o objeto de avaliação imediatamente antes do preenchimento do formulário.

## 6.5 Análise de Dados Após a Aplicação do Formulário

Após o término da coleta de respostas do formulário de avaliação, houve uma nova rodada de análise de dados. Esta análise de dados é a etapa final do ciclo de design e de rigor desta pesquisa. A análise engloba avaliações de confiabilidade do formulário enquanto instrumento de medição e a validade de seus construtos. O processo geral desta análise de dados está ilustrado na Figura B.9.

O processo se iniciou com a carga e tratamento dos dados. Assim, os dados ficaram prontos para prepará-los à análise de diferentes módulos desta pesquisa.

### 6.5.1 Carregamento dos Dados

O primeiro passo da análise de dados feita nessa pesquisa é a carga delas aos devidos repositórios. No caso da avaliação feita de forma impressa, deve ser feito o preenchimento manual de um arquivo que siga um formato similar ao digital.

Após a carga nos repositórios, foi feito um tratamento dos dados para padronizá-los, removendo particularidades que possam ter vindo de uma aplicação específica do formulário. Foi comum a todos os módulos o uso de arquivos CSV para carregar os dados da resposta e a maioria dos módulos foi desenvolvido em Python, recorrendo ao DataFrame do Pandas (WES MCKINNEY, 2010) como estrutura de dados responsável pelo manuseio dos dados durante a análise. A exceção é um programa escrito em R, feito para a análise paralela dos dados.

### 6.5.2 Tratamento/Pré-processamento dos Dados

Os dados da avaliação foram anonimizados, a fim de respeitar a Lei Geral de Proteção de Dados Pessoais (BRASIL, 2018). A Lei Geral de Proteção de Dados Pessoais define que dados sobre origem racial ou étnica são dados pessoais sensíveis. Em complemento, é exigido que em estudos realizados por órgãos de pesquisa, deve-se garantir, sempre que possível, a anonimização de dados pessoais sensíveis. Foram coletados dois tipos de dados que exigem tratamento, por identificarem os seus respondentes: o nome e o e-mail. Imediatamente após o carregamento dos dados em um DataFrame, foi removida a coluna de e-mail e o módulo *randomname* do Python foi usado para substituir o nome dos respondentes por nomes fictícios. Desta forma, os dados demográficos podem ser visualizados respeitando a privacidade dos participantes.

### 6.5.3 Confiabilidade

É necessário que um instrumento de medição seja confiável para afirmar que de fato está medindo algum constructo. O Alfa de Cronbach, métrica amplamente utilizada, que mede o piso da consistência interna, foi a medida selecionada para esta pesquisa. Foi calculado o alfa de Cronbach para o instrumento em sua totalidade e no cenário que um dos itens fosse excluído da análise. O prosseguimento desta pesquisa, referente à validade do instrumento e análise fatorial, depende da obtenção de um valor satisfatório, acima de 0,7, para o Alfa de Cronbach. O Alfa de Cronbach foi calculado usando o módulo `cronbach_alpha()` do pingouin (VALLAT, 2022). O módulo recebe um DataFrame dos dados de resposta como entrada e retorna o valor do Alfa de Cronbach, junto aos valores limites do intervalo de confiança da métrica. O módulo permite a configuração do intervalo de confiança, e esta pesquisa utilizou

o intervalo de confiança padrão do módulo de 95%.

Além do cálculo desta métrica de consistência interna, foi calculada a matriz de correlação dos dados de resposta do formulário. A consistência interna é resultante da covariância entre os itens, portanto, a matriz será exibida e discutida ao abordar a consistência interna. A matriz de correlação foi calculada usando a função do Pandas `corr()` (PANDAS DEVELOPMENT TEAM, 2020).

A outra métrica de confiabilidade calculada nessa pesquisa, a confiabilidade *Split-Half*, foi calculada usando o método *item\_split\_half* da biblioteca *performance* (LÜDECKE *et al.*, 2021), desenvolvida para o R. Esta função recebe os itens do formulário em sua totalidade, divide-os em um conjunto com os itens pares e outros com os itens ímpares e calcula o valor de confiabilidade *Split-Half* original e o ajustado, mediante a fórmula Spearman-Brown, para o formulário completo.

#### 6.5.4 Validade de Construtos

Esta pesquisa busca a confirmação do modelo teórico estabelecido após uma revisão bibliográfica e detalhado na próxima seção. Em paralelo, objetiva-se identificar um novo modelo que explique os itens que sobraram após a etapa de validade de conteúdo. Portanto, ambos os tipos abrangentes de análise fatorial, confirmatória e exploratória, serão conduzidos.

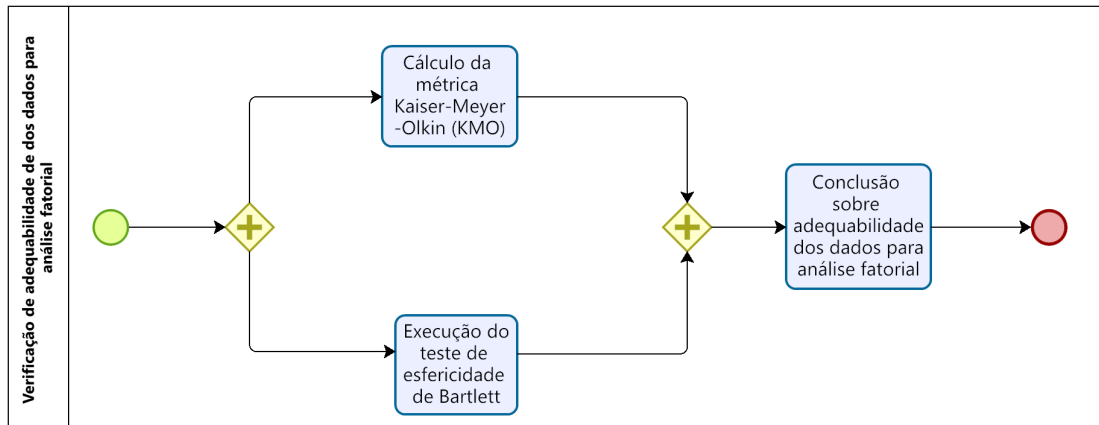
Esta pesquisa já teve o seu escopo amplamente reduzido pela etapa de validade de conteúdo. Foi alcançado um número de itens que possibilita uma medição rápida e facilmente computável pelos módulos de análise de dados atual. Dito isso, não há a necessidade da redução de dados, o objetivo é somente a identificação de uma estrutura mediante a sumarização dos dados.

O objetivo da análise fatorial nesta pesquisa é de agrupar as variáveis que variam conjuntamente. Isto é motivado pela necessidade de encontrar as variáveis latentes que expliquem os itens medidos diretamente pela pesquisa.

Nesta pesquisa, não há nenhuma característica que previa que a coleta de dados se apresentasse como divisiva nas respostas do formulário. Por conseguinte, as análises fatoriais exploratória e confirmatória feitas nessa pesquisa englobam todas as respostas.

Após essas decisões no desenho da análise fatorial, deve ser verificada a aptidão dos dados para a análise fatorial. Variáveis cujas correlações demonstram níveis baixos indicam uma baixa probabilidade da existência de fatores latentes que as expliquem, tornando inadequado a análise fatorial. Consequentemente, algumas técnicas foram aplicadas para assegurar a adequabilidade dos dados para uma análise fatorial.

Figura 6.5: Subprocesso da análise de adequabilidade dos dados da avaliação a uma análise fatorial.



Powered by  
brazagi  
Modeler

## Adequabilidade dos Dados para Análise Fatorial

As correlações entre as variáveis devem passar por algumas medidas para averiguar se estão aptas para uma análise fatorial. Esta pesquisa fará duas verificações de adequabilidade para análise fatorial: com o teste de esfericidade de Bartlett e com a métrica Kaiser-Meyer-Olkin (KMO) (Figura 6.5).

O teste de esfericidade de Bartlett foi calculado usando a função `calculate_bartlett_sphericity` do módulo `Python factor_analyzer` (BIGGS e MADNANI, 2022). Como hipótese nula, é afirmado que a matriz de correlação corresponde a uma matriz de identidade com as mesmas dimensões. Como hipótese alternativa, é afirmado que a matriz de correlação e uma matriz de identidade com as mesmas dimensões são diferentes. As métricas resultantes do teste, o valor p e o valor qui-quadrado, foram analisados na seção de resultados para verificar a validade dos dados desta pesquisa para uma análise fatorial.

Outra medida para determinar a adequabilidade dos dados para uma análise fatorial é o Kaiser-Meyer-Olkin (KMO), que dispõe de classificações para seus valores. Entre as classificações, destaca-se que um valor abaixo de 0,5 é considerado inaceitável para análise; as variáveis com valor abaixo de 0,5 devem ser excluídas da análise e uma nova rodada da técnica de *measurement of sampling adequacy* (MSA) deve ser realizada sem aquela variável. Esse processo é repetido até que apenas variáveis aptas a passar pela análise fatorial estejam presentes.

## Análise Fatorial Confirmatória

Análises fatoriais confirmatórias são feitas para validar hipóteses fundamentadas na teoria sobre a relação entre um conjunto de variáveis que expliquem determinados fatores (SUHR, 2006). Análises fatoriais confirmatórias também são conhecidas como modelos fatoriais restritos, por restringirem a 0 a carga fatorial de variáveis as quais é presumido que não são explicadas pelo mesmo fator. O modelo teórico da relação entre as variáveis dita as restrições deste tipo de análise fatorial.

A análise fatorial confirmatória nessa pesquisa visa confirmar ou reprovar as hipóteses de que os itens da escala correspondem aos respectivos fatores do modelo, presentes em um nível hierárquico de abstração mais elevado. A confirmação ou refutação da hipótese só prova a existência de uma relação entre a variável e os fatores. Entretanto, não se pode assumir uma relação de causalidade a partir disso. Idealmente, as variáveis que pertencem ao mesmo grupo deveriam ter uma correlação alta, enquanto variáveis de grupos diferentes devem ter correlações baixas, por descreverem fenômenos diferentes. O fator ao qual o nome do método se refere é a variável latente que os itens da escala, as variáveis dependentes, estão medindo.

Foi conduzida uma análise fatorial confirmatória, delineada na Figura B.11, que visa atestar que o conjunto de critérios escolhidos pelo autor e validado por especialistas representa realmente o mesmo fenômeno. Os critérios delineados pelo autor para a área de motivação ao aprendizado, por exemplo, devem ter uma correlação alta comparada à correlação das outras duas áreas.

A análise fatorial confirmatória foi feita com o pacote `factor_analyzer` (BIGGS e MADNANI, 2022) do Python. O módulo de análise fatorial confirmatória, `CFA`, recebe os dados de resposta do formulário para cada item e a estrutura fatorial hipotética. A estrutura fatorial prevê um mapa com a lista de itens que teoricamente correspondem a cada fator.

Após a análise, realizada pela função `fit()` do módulo, é retornado a carga fatorial para cada item. Outrossim, são retornadas informações da variância do erro e do erro padrão da matriz de covariância.

As estimativas geradas pela análise confirmatória foram testadas para verificar se estão adequadas ao modelo pré-estabelecido. Inicialmente, foram identificadas estimativas impróprias para o modelo teórico. Critérios cujo valor da estimativa for superior a 1 ou cujo valor de variação do erro for ínfimo ou negativo foram rotulados como impróprios. O mesmo rótulo foi dado a critérios cujo valor padrão da estimativa for excessivamente alto. Após esta classificação das estimativas, foram feitas três verificações de adequabilidade.

A verificação geral de adequabilidade ao modelo foi feita usando o módulo do Python `SEMOPY`, que possui diversas funcionalidades ligadas a modelagem de equa-

ções estruturais. O processo se inicia com a definição do modelo teórico, especificando as variáveis latentes e as variáveis observadas correspondentes. Em seguida, é fornecido um DataFrame com os dados obtidos para cada variável observada do modelo usando a função `fit()`. Finalmente, os índices de adequabilidade dos dados ao modelo podem ser extraídos usando a função `calc_stats()`. O artigo (IGOLKINA e MESHCHERYAKOV, 2020) descreve as métricas de adequabilidade calculadas pelo modelo e especifica as fórmulas usadas para obter seus valores.

Em paralelo, foram calculadas a confiabilidade composta dos construtos e a variância média extraída. Esta análise foi feita para cada um dos constructos e considerou a possibilidade de estimativas impróprias. Em caso de ocorrência de estimativas impróprias durante a análise, tais estimativas foram desconsideradas na análise para aquele determinado; assim evitam-se distorções no cálculo das métricas de adequação do modelo de medição, mas nota-se que a métrica se refere a um subconjunto das variáveis do construto e não ao construto original.

A terceira verificação foi em relação ao modelo estrutural, na qual foi avaliada a significância estatística dos valores das estimativas, com atenção a possíveis multicolinearidades do modelo. Nesta verificação de adequabilidade, foi calculada a significância estatística de cada uma das variáveis indicadoras. Os coeficientes das estimativas foram testados contra uma hipótese nula, que o coeficiente é significativamente superior a 0. Foram feitos testes usando o nível de significância tradicional,  $\alpha = 0,05$  e um nível de significância mais conservador,  $\alpha = 0,01$ .

## **Análise Fatorial Exploratória**

A validade fatorial é um dos métodos pertencentes à validade de construtos e não depende de conexões teóricas nos seus procedimentos (STRAUB *et al.*, 2004). O objetivo da validade fatorial é observar a relação entre diferentes variáveis no mesmo nível hierárquico, no sentido de que não é presumido que um componente tenha uma relação causal com outro (STRAUB *et al.*, 2004).

Esta pesquisa optou pela análise fatorial exploratória, EFA, em relação a alternativas de redução ou sumarização de dados, como a análise de componentes principais, devido ao objetivo de encontrar o modelo fatorial. Esta etapa identifica quais fenômenos inobserváveis estão sendo medidos pelos itens, e quais itens compõem cada fator.

Nessa pesquisa, uma análise fatorial foi conduzida para redefinir quais fatores realmente estão sendo medidos pelo instrumento. O processo de análise fatorial foi disposto na Figura B.12 Uma das aplicações de análises fatoriais é a validação de dados coletados por uma pesquisa, verificando se as medições correspondem aos constructos teóricos que o instrumento de medição teoricamente mede (MATSUNAGA, 2010). A análise fatorial calcula a proporção da variância explicada por distintas



variáveis inobserváveis. Este procedimento é importante para reavaliar se fatores encontrados no embasamento teórico do modelo de fato estão sendo capturados pelo instrumento.

O processo da análise fatorial começa com a hipótese de que uma variável latente explica toda a correlação dos itens. Para testar esta hipótese, é calculada uma matriz de correlações residuais. A correlação residual é o valor da diferença entre a correlação esperada com a correlação real entre um par de itens. Correlações residuais significativas indicam haver pelo menos algum outro fator que explique o restante da correlação. Esse processo é seguido iterativamente até que a correlação residual esteja abaixo de um limite aceitável.

A rotação de fatores é um procedimento amplamente utilizado na análise fatorial por simplificar a interpretação da estrutura fatorial. Em contextos no qual se assume uma relação ortogonal, de nenhuma correlação entre os fatores, as cargas fatoriais de cada item são rebalanceadas para que apenas uma pequena porção tenha uma carga fatorial alta e o restante fique zerado. Isso facilita a interpretação de quais fatores possuem mais influência sob o item.

A rotação fatorial feita por padrão pelo módulo *factor\_analyzer* (BIGGS e MADNANI, 2022), utilizado nesta pesquisa, é o Promax. O Promax é um método oblíquo, divergindo do modelo teórico desta pesquisa, que assume uma ortogonalidade entre os fatores. Em decorrência disto, será especificado entre os parâmetros que o módulo deverá realizar a rotação fatorial usando o método tradicional Varimax (KAISER, 1958).

Ao reduzir o conjunto de fatores do modelo é calculado a carga fatorial dos itens, quantificando a correlação do item com os fatores que sobraram após a análise fatorial. Como mencionado anteriormente, a validade de constructos, do qual a análise fatorial e análise de componentes fazem parte, não considera a relação teórica entre as variáveis. Portanto, o método não rotula os fatores, cabe ao pesquisador identificar qual é o termo mais apropriado para o conjunto de variáveis explicados por cada fator.

Os fatores encontrados na análise fatorial são variáveis estatísticas. Portanto, uma das preocupações do pesquisador deve ser o tamanho de amostra, buscando maximizar o número de respondentes por variável. Uma proporção de ao menos 10 vezes mais pessoas respondendo comparado ao número de variáveis em análise é desejável, enquanto uma proporção de 5 vezes mais é um mínimo aceitável (HAIR *et al.*, 2009). No contexto desta pesquisa, a proporção de 10 vezes mais respondentes por itens demandaria aproximadamente 500 respondentes, assumindo a distribuição de um formulário com um item por critério do modelo original. Isso reforça a necessidade da redução do formulário para contemplar apenas os itens tidos como essenciais pelos especialistas.

O autovalor mede a magnitude de informação capturada por um fator (DEVELLIS e THORPE, 2021). Em técnicas como a análise de componentes principais, o valor unitário de um autovalor corresponde à quantidade de informação capturada por um item. Portanto, o valor máximo do autovalor passa a ser o total de itens na escala. Fatores com um valor próximo ou igual a um explicam a correlação de apenas um pouco mais do que um item. Portanto, uma das heurísticas, conhecida como o critério de Kaiser, para a escolha de fatores é que apenas fatores com um autovalor superior a 1 são selecionados ao conjunto de fatores da escala (KAISER, 1960).

Outra heurística é o teste Scree, no qual o pesquisador traça um gráfico com os autovalores dos fatores encontrados na análise fatorial, ordenados de forma decrescente, e identifica um ponto no gráfico no qual o declive vai de principalmente vertical para uma tendência mais horizontal (CATTELL, 1966). Os fatores anteriores a essa mudança para uma tendência horizontal são mantidos, enquanto fatores com autovalores de valores baixos e relativamente semelhantes são descartados. Ambas as técnicas de seleção de fatores descritas foram conduzidas nesta pesquisa. Por último, foi feita uma análise paralela como o terceiro método para identificar o conjunto de fatores que deve ser selecionado.

A análise paralela desta pesquisa foi conduzida usando o comando *paran* da biblioteca homônima (DINNO, 2009) na linguagem R. Esta análise paralela necessita apenas da matriz de correlação e das dimensões da matriz. A matriz de correlação foi calculada anteriormente e mantida em um arquivo CSV. Já as dimensões em questão são o número de itens para as colunas e a quantidade de respondentes para as linhas.

A análise paralela foi feita indicando ao *paran* que é uma análise fatorial comum que será feita. É necessário especificar isto ao módulo, pois o padrão do pacote é a análise de componentes principais.

Após a seleção dos fatores relevantes ao fenômeno sendo medido, há de ser mapeado a relação entre os itens do instrumento de avaliação e os fatores. A essa altura, os fatores também precisam ser nomeados. Ambas essas atividades são feitas manualmente. A carga fatorial do item para cada fator selecionado foi interpretada. Paralelamente, a nomeação dos fatores dependeu de consultas à base teórica para encontrar na literatura os termos mais adequados para o fenômeno que está medindo aquele conjunto de itens.

Uma interpretação fatorial foi feita para a quantidade de fatores sugerida pelos critérios de Kaiser, teste de Scree e de análise paralela. Nem sempre esses critérios coincidem na quantidade de fatores selecionados, levando a mais de uma interpretação fatorial possível por rodada.

A principal restrição para a interpretação fatorial é que a variável precisa ter uma

carga fatorial acima de 0,3 para ser considerada significativamente correlata com tal fator. Isso se aplica para pesquisas cujo tamanho de amostra é inferior a 100 (HAIR *et al.*, 2009), como nesta pesquisa. Valores que exibem uma carga fatorial superior a 0,30 para dois fatores foram atribuídos ao fator para o qual a carga fatorial do item foi a mais alta.

Caso um ou mais itens não correspondam a algum fator em uma determinada interpretação fatorial, uma nova rodada de análise fatorial foi feita. As execuções subsequentes desconsideraram os itens que não atenderam os critérios estabelecidos. Esse processo foi repetido iterativamente até que uma análise fatorial não indicasse o descarte de um item do formulário.

### 6.5.5 Comparações entre Grupos Categóricos (Análise de Variância)

Esta pesquisa fez diversas comparações entre grupos categóricos, cujo processo de análise se encontra na Figura 5.1. Os dados foram classificados consoante as respostas das seções de dados demográficos e hábitos de jogo. Em seguida, foram comparados ao padrão de respostas dos itens de avaliação do jogo.

A comparação entre os grupos categóricos usou a técnica do teste Kruskal-Wallis. Os valores dos testes Kruskal-Wallis foram calculados com o auxílio do módulo *kruskal* da biblioteca *pingouin* (VALLAT, 2018). O *script* elaborado nessa pesquisa para a comparação entre grupos categóricos recebe um DataFrame com as respostas do formulário e iterativamente faz uma chamada à função *kruskal()* do *pingouin*. O campo categorizável de dados demográfico ou hábitos de jogo serviu como a variável independente, enquanto a distribuição das frequências de respostas para cada item de avaliação do jogo foi passado ao módulo como a variável dependente.

Após a execução, o módulo provê as métricas necessárias para testar a hipótese nula, de que as categorias influenciam os valores da variável dependente. Nesta pesquisa, a hipótese nula é rejeitada caso a distribuição de respostas de um item do formulário variar significativamente conforme a categoria. As métricas fornecidas pelo método *kruskal()* são os graus de liberdade, a estatística de teste H de Kruskal-Wallis e o valor-p, não corrigido, correspondente. A interpretação dos resultados para avaliar a aprovação ou rejeição da hipótese nula é feita então com a combinação dos graus de liberdade e estatística de teste H, comparando-os com a distribuição qui-quadrado. Em paralelo, o valor-p é comparado com o valor alfa estabelecido para a pesquisa.

O pressuposto que a estatística de teste H é uma aproximação da distribuição qui-quadrado exige que o tamanho amostral seja minimamente significativo. Portanto, apenas categorias cuja frequência de resposta tenham sido superiores a cinco para

um determinado item foram aceitas para análise de variância.

Os resultados para a métrica Kruskal-Wallis foram reportados e discutidos nas próximas seções. Serão destacadas as combinações de itens do formulário e dados demográficos e de hábitos de jogo cujos resultados diferem entre suas categorias observadas.

# Capítulo 7

## Artefato

### 7.1 Modelo de Avaliação de Jogos Educacionais Digitais Original

Este capítulo apresenta os artefatos gerados pelos dois ciclos de design e rigor. Inicialmente, é apresentado o modelo de avaliação de jogos educacionais digitais proveniente de teorias encontradas na revisão da literatura; posteriormente, é exposta uma versão que contempla apenas os critérios cujo conteúdo foi validado e, finalmente, a versão conclusiva do modelo, resultante de uma análise fatorial exploratória.

A estrutura do modelo de reação ao jogo educacional digital segue os moldes detalhados no capítulo de Método Proposto (Capítulo 6.1). As descrições detalhadas dos subcritérios que compõem cada critério do modelo original desta pesquisa estão dispostas no Apêndice A.

O modelo original serviu como uma taxonomia da avaliação de jogos educacionais digitais. O desenvolvimento de uma taxonomia é útil para o agrupamento de itens em categorias e identificação de critérios que teoricamente representam um fenômeno. O modelo inicial desta pesquisa foi dividido em três áreas importantes para a avaliação de um jogo educacional digital: jogabilidade, experiência de usuário e motivação para o aprendizado, como mostra a Figura 7.1. A jogabilidade e experiência do usuário medem a qualidade do jogo educacional digital, respectivamente, como um jogo e como um sistema de software, enquanto a motivação do aluno servirá como um indicador inicial para a qualidade do jogo como material educacional.

A **Jogabilidade** é interpretada como a qualidade das interações que um jogador tem com o sistema, outros jogadores, e o contexto (SÁNCHEZ *et al.*, 2012). Seus critérios são:

- **Diversão:** a principal razão pela qual pessoas decidem jogar, definida por KOSTER (2013) como o desejo de aprender padrões novos aplicados em con-

textos familiares. É necessário que um jogo apresente seu conteúdo com variações nos padrões para motivar pessoas a jogá-lo por um tempo prolongado. Isso exige a atenção do jogador, exigindo um aprendizado sobre como reagir às variações. A convergência entre aprendizado e diversão intrínseca em jogos, quando combinados com outros componentes críticos à educação, tornam jogos digitais excelentes materiais instrucionais. (KOSTER, 2013) (SÁNCHEZ *et al.*, 2012)

- **Comunicação:** as interações sociais encapsuladas nesse critério incluem, mas não se limitam a: as conversas entre jogadores, o compartilhamento de recursos e de informações no jogo, e as interações sobre o conteúdo ou temática do jogo externas a sessões do jogo. (SWEETSER e WYETH, 2005) (FU *et al.*, 2009) (SÁNCHEZ *et al.*, 2012)
- **Interações com o Sistema:** as características e qualidades das respostas do sistema às ações feitas pelo jogador. (SÁNCHEZ *et al.*, 2012)
- **Escolha:** a variedade e a qualidade das escolhas as quais o jogador tem acesso nas ações, estratégias e mecânicas à sua disposição para conquistar os seus objetivos. (KELLER, 1987)
- **Presença:** o quanto o jogador sente-se imerso e presente na experiência do jogo, abrangendo as sensações de presença física e presença social. (TAKATALO *et al.*, 2010)
- **Fluxo:** uma sensação prazerosa de completa absorção ao realizar alguma atividade, com a ausência de ansiedade e tédio. (CSIKSZENTMIHALYI, 2008) (TAKATALO *et al.*, 2010) (SWEETSER e WYETH, 2005)

A **Experiência de Usuário** é a qualidade das interações humano-computador de um sistema, atribuídas à qualidade do produto, um sistema de software, e à qualidade em uso do sistema. Essa área de avaliação tem como fundamentação teórica principal a característica de usabilidade do SQuaRE (ISO/IEC 25000), pertencente à avaliação da qualidade de produto de sistemas de software. Portanto, um dos critérios escolhidos para avaliar a experiência de usuário é a usabilidade em si, enquanto os outros critérios, a acessibilidade e a proteção de erros, são sub-características de usabilidade no SQuaRE. Os critérios de experiência de usuário são:

- **Usabilidade:** É o quanto usuários de um determinado público conseguem atingir efetivamente os seus objetivos de forma eficiente e satisfatória no contexto especificado para o sistema. (ISO/IEC 25000) (PETRI *et al.*, 2019)

- **Acessibilidade:** A acessibilidade desse modelo tem como fundamento os dois seguintes trechos da Lei Nº 13.146, de 6 de julho de 2015: a “possibilidade e condição de alcance para utilização, com segurança e autonomia, de ... informação e comunicação, inclusive seus sistemas e tecnologias” e “... garantir condições de acesso, permanência, participação e aprendizagem por meio de recursos de acessibilidade que eliminem as barreiras e promovam a inclusão plena”. (REIS, 2020) (ISO/IEC 25000) (PETRI *et al.*, 2019)
- **Proteção de Erros:** Devem ser dados pesos diferentes a duas categorias de erros que possivelmente ocorrerão no jogo. O primeiro são erros *leves* que atrapalham a experiência do usuário, mas não afetam o andamento do jogo; esses erros devem ser evitados, mas serão tolerados. Em contrapartida, os erros *graves*, disruptivos ou que alterem o percurso, e conseqüentemente o desempenho ou resultado dos jogadores, não podem ocorrer. (ISO/IEC 25000) (SWEETSER e WYETH, 2005) (PETRI *et al.*, 2019)

A **Motivação ao Aprendizado** é um pré-requisito para o aprendizado propriamente dito. O aluno deve estar intrinsecamente com a vontade e a intenção de fazer o esforço necessário para aprender e também encontrar-se em um ambiente favorável ao aprendizado.

- **Atenção:** O engajamento do aluno deve ser obtido e mantido, mas mantido em um nível moderado para evitar que a experiência torne-se ansiolítica. A atenção do aluno é fundamental, por ser um pré-requisito para o aprendizado. (KELLER, 1987) (CSIKSZENTMIHALYI, 2008) (TAKATALO *et al.*, 2010) (SWEETSER e WYETH, 2005) (PETRI *et al.*, 2019) (PASQUALOTTO *et al.*, 2022)
- **Relevância:** O material instrucional escolhido para ensinar o conteúdo deve ser perceptivelmente relacionado ao conteúdo esperado pelo aluno e ser relevante para o seu aprendizado. (KELLER, 1987) (TAKATALO *et al.*, 2010) (PETRI *et al.*, 2019)
- **Confiança:** O aluno deve sentir controle sobre as suas ações no jogo e ter consciência de que consegue superar objetivos claros caso se empenhe. Deve-se instigar autoconfiança e buscar um ambiente que deixe o aluno à vontade. (KELLER, 1987) (PETRI *et al.*, 2019)
- **Satisfação:** A experiência do jogo deve gerar um bem-estar, algo que é influenciado por recompensas e pela ausência de pressões ou influências negativas. (KELLER, 1987) (PETRI *et al.*, 2019) (ISO/IEC 25000)

Figura 7.1: Modelo de Reação a Jogos Educacionais Digitais

Objetivo de Medição: Reação do Aluno ao Jogo Educacional Digital	<b>Jogabilidade</b>	<b>Diversão:</b> Fiero, Animação, Relaxamento, Ausência de Medo
		<b>Comunicação:</b> Interações Competitivas, Interações Cooperativas, Compartilhamento, Contexto Social
		<b>Interações com o Sistema:</b> Velocidade, Imprevisibilidade
		<b>Escolha:</b> Variedade nas Ações do Agente, Variedade de Estratégias para Ganhar o Jogo
		<b>Presença:</b> Presença Física, Engajamento com o Personagem, Co-presença
		<b>Fluxo:</b> Competência, Sensação de Distorção de Tempo, Engajamento com o Mundo do Jogo
	<b>Experiência de Usuário (UX)</b>	<b>Usabilidade:</b> Inteligibilidade, Aprendizabilidade, Atratividade
		<b>Acessibilidade:</b> Acessibilidade para Pessoas com Daltonismo, Nitidez da Imagem, Textos Bem Formatos, Acessibilidade para Pessoas com Deficiência Auditiva, Volume Ajustável, Linguagem Compreensível
		<b>Proteção de Erros:</b> Proteção de Erros "Leves", Ausência de Erros Disruptivos
	<b>Motivação do Aluno</b>	<b>Atenção:</b> Concretude, Humor
		<b>Relevância:</b> Experiência, Temática Relevante, Mecânicas Relevantes
		<b>Confiança:</b> Requisitos de Aprendizado, Moderação de Expectativas, Autoconfiança
		<b>Satisfação:</b> Recompensas para Tarefas Interessantes, Recompensas para Tarefas Entediadas, Elogios a Resultados Positivos
		<b>Curiosidade:</b> Sensorial, Cognitiva
		<b>Desafio:</b> Objetivo, Demanda de Habilidade, Dificuldade Crescente, Ritmo do Jogo
	<b>Fantasia:</b> Intrínseca, Extrínseca	

- **Curiosidade:** A curiosidade é o instinto de exploração, seja aquele provocado por estímulos sensoriais ou um por vontade de aprender e descobrir informações novas. (MALONE, 1981) (KELLER, 1987) (DILLON, 2010) (SÁNCHEZ *et al.*, 2012)
- **Desafio:** O esforço causado por dificuldades e obstáculos de jogo mantém o engajamento de seus jogadores. O jogo deve ser difícil ao ponto de haver uma incerteza quanto ao seu resultado. (MALONE, 1981) (TAKATALO *et al.*, 2010) (SWEETSER e WYETH, 2005)
- **Fantasia:** A fantasia é a evocação de imagens mentais e sensações a partir de situações e objetos fictícios ou fantasiosos, que não ocorrem no ambiente real da pessoa. A fantasia pertencente ao jogo pode ser intrínseca às mecânicas ou extrínseca, de forma que as mecânicas e a fantasia independem uma da outra. (MALONE, 1981)

## 7.2 Modelo de Avaliação de Jogos Educacionais Digitais : Pós-Validação de Conteúdo

As dimensões do modelo foram reduzidas após a validação de conteúdo feito por especialistas. Apenas os critérios considerados essenciais à avaliação foram mantidos.



A Figura 7.2 exibe o modelo com apenas os critérios remanescentes após a etapa de validação de conteúdo, tais itens estão em destaque na Figura 7.3, que os exibe em conjunto aos demais no modelo inicialmente elaborado nesta pesquisa. Os critérios mantidos para a versão atual do modelo são:

- Para a área de jogabilidade, apenas o critério de animação.
- Na área de avaliação de experiência de usuário: inteligibilidade, aprendibilidade, nitidez da Imagem e linguagem compreensível.
- Finalmente, para a motivação ao aprendizado permanecem a maioria dos critérios: experiência, mecânicas relevantes, ritmo do Jogo, requisitos de aprendizado, curiosidade cognitiva, curiosidade sensorial, concretude, dificuldade crescente.

Figura 7.2: Modelo de Reação a Jogos Educacionais Digitais, exibindo apenas os critérios que permaneceram após a validação de conteúdo.

<b>Objetivo de Medição: Reação do Aluno ao Jogo Educacional Digital</b>	<b>Jogabilidade</b>	<b>Diversão:</b> Animação
	<b>Experiência de Usuário (UX)</b>	<b>Usabilidade:</b> Inteligibilidade, Aprendizabilidade
		<b>Acessibilidade:</b> Nitidez da Imagem, Linguagem Compreensível
	<b>Motivação do Aluno</b>	<b>Atenção:</b> Concretude
		<b>Relevância:</b> Experiência, Mecânicas Relevantes
		<b>Confiança:</b> Requisitos de Aprendizado
		<b>Curiosidade:</b> Sensorial, Cognitiva
		<b>Desafio:</b> Dificuldade Crescente, Ritmo do Jogo

Figura 7.3: Modelo de Avaliação de Jogos Educacionais Digitais original, com destaque aos critérios que permaneceram após a validação de conteúdo.

Objetivo de Medição: Reação do Aluno ao Jogo Educacional Digital	Jogabilidade	Diversão: Fiero, <i>Animação</i> , Relaxamento, Ausência de Medo
		Comunicação: Interações Competitivas, Interações Cooperativas, Compartilhamento, Contexto Social
		Interações com o Sistema: Velocidade, Imprevisibilidade
		Escolha: Variedade nas Ações do Agente, Variedade de Estratégias para Ganhar o Jogo
		Presença: Presença Física, Engajamento com o Personagem, Co-presença
		Fluxo: Competência, Sensação de Distorção de Tempo, Engajamento com o Mundo do Jogo
	Experiência de Usuário (UX)	Usabilidade: <i>Inteligibilidade</i> , <i>Aprendizibilidade</i> , Atratividade
		Acessibilidade: Acessibilidade para Pessoas com Daltonismo, <i>Nítidez da Imagem</i> , Textos Bem Formataados, Acessibilidade para Pessoas com Deficiência Auditiva, Volume Ajustável, <i>Linguagem Compreensível</i>
		Proteção de Erros: Proteção de Erros "Leves", Ausência de Erros Disruptivos
	Motivação do Aluno	Atenção: <i>Concretude</i> , Humor
		Relevância: <i>Experiência</i> , Temática Relevante, <i>Mecânicas Relevantes</i>
		Confiança: <i>Requisitos de Aprendizado</i> , Moderação de Expectativas, Autoconfiança
		Satisfação: Recompensas para Tarefas Interessantes, Recompensas para Tarefas Entediantes, Elogios a Resultados Positivos
		Curiosidade: <i>Sensorial</i> , <i>Cognitiva</i>
		Desafio: Objetivo, Demanda de Habilidade, <i>Dificuldade Crescente</i> , <i>Ritmo do Jogo</i>
	Fantasia: Intrínseca, Extrínseca	

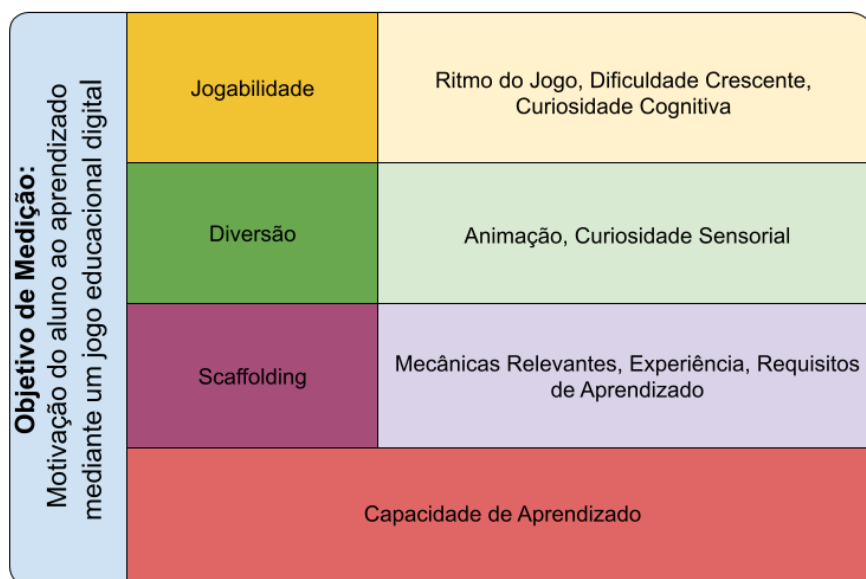
### 7.3 Modelo de Avaliação de Jogos Educacionais Digitais: Pós-Análise Fatorial Exploratória

O modelo de avaliação em sua versão final (Figura 7.4), no escopo desta dissertação, foi traçado posterior à análise fatorial exploratória. O estabelecimento desta versão foi feito especificamente após a interpretação fatorial. Na interpretação fatorial, cada item foi designado a um fator conforme a sua carga fatorial e os fatores encontrados na prática foram rotulados.

A análise fatorial exploratória prevê que itens cuja carga não corresponde a fatores selecionados sejam descartados. Dessa forma, o modelo ficou limitado a apenas nove dos critérios originais. Segue o artefato desta pesquisa em sua versão atualizada, estabelecida após uma análise fatorial exploratória.

Os fatores foram rotulados como jogabilidade, diversão, *scaffolding* e capacidade de aprendizado. Houve uma predominância de critérios inicialmente atribuídos ao fator de motivação ao aprendizado, o critério de animação é a única exceção. Isso reduz o escopo da avaliação, tornando-a mais focada na avaliação da motivação ao aprendizado do que a avaliação de jogos educacionais digitais em um contexto geral.

Figura 7.4: Modelo de Avaliação da Motivação ao Aprendizado via Jogos Educativos Digitais.



# Capítulo 8

## Resultados e Discussões

### 8.1 Resultado

Esta pesquisa foi composta de duas iterações dos ciclos de design e de rigor da metodologia DSR. O ciclo inicial foi composto de uma revisão da literatura e o artefato resultante foi a versão inicial do modelo de avaliação de jogos educacionais digitais, cujo resultado está exposto no Capítulo 7.1. Este capítulo, por sua vez, apresenta os dados resultantes da segunda iteração dos ciclos DSR de rigor e design; neste ciclo, foi conduzida a validação de conteúdo do formulário, aplicação do formulário e uma análise de dados posterior à aplicação do formulário.

Os resultados da análise da validação de conteúdo guiaram o refinamento do formulário de avaliação, indicando mudanças que deveriam ser feitas em seus itens. Na análise de dados após a distribuição do formulário de avaliação de jogos educacionais digitais, informações de confiabilidade, validade de construtos e de comparação entre grupos categóricos foram coletados e interpretados.

#### 8.1.1 Resultados Anteriores à Aplicação do Formulário

Inicialmente, a validade de conteúdo do formulário e seus itens foi determinada seguindo um método tradicional. O método tradicional proposto por LAWSHE (1975) pressupõe o cálculo dos coeficientes de validade de conteúdo de cada item e do índice de validade de conteúdo do formulário em sua totalidade.

As tabelas abaixo listam o coeficiente de validade de conteúdo (IVC) de cada área de avaliação do modelo (Tabelas 8.1, 8.2, 8.3), seguido do valor agregado de índice de validade de conteúdo das respectivas áreas (Tabela 8.4):

Tabela 8.1: Índice de Validade de Conteúdo para os critérios de Jogabilidade

Critério	Item	IVC
Engajamento com o Personagem	Ao longo do jogo senti uma identificação com o papel que estava desempenhando.	-0,5
Fiero	Senti alegria durante a minha experiência com o jogo.	-0,33

Imprevisibilidade nas Interações com o Sistema	Foi fácil prever as respostas do jogo às minhas ações, dificilmente fiquei surpreso.	-0,167
Ausência de Medo	Estive consciente de que minha experiência no jogo não envolvia riscos verdadeiros, e, portanto, não senti medo ao jogar.	-0,5
Sensação de Distorção de Tempo	Senti o tempo passar em um ritmo diferente enquanto jogava, parecia estar passando mais devagar ou mais rápido.	0,0
Variedade nas Ações do Agente	Pude alternar as estratégias para tentar ganhar o jogo, não me senti limitado a uma única maneira de jogar.	0,0
Relaxamento	Senti pouca ou nenhuma tensão durante o jogo.	-0,833
Interações Cooperativas	Gostei das interações colaborativas, quando jogadores atuam em grupo para alcançar um objetivo compartilhado, que tive com outros jogadores.	-0,167
Contexto Social	Tive conversas com colegas sobre o jogo e achei tais conversas produtivas para o meu aprendizado.	-0,5
Co-presença	Apesar de estar em um ambiente virtual, tive a sensação de estar compartilhando a minha experiência de forma vívida com outros.	-0,667
Competência	Senti competência em relação ao meu desempenho no jogo, tive as habilidades necessárias para jogar bem.	0,0
Engajamento com o Mundo do Jogo	Esqueci do mundo ao meu redor enquanto jogava, era como se eu estivesse presente no mundo do jogo.	0,0
Variedade nas Ações do Agente	O jogo me oferecia a todos os instantes uma quantidade desejável de ações alternativas para o meu personagem performar.	-0,167
Velocidade das Interações com o Sistema	O jogo respondeu às minhas ações com agilidade e velocidade.	0,167
Interações Competitivas	Os momentos em que competi com outros jogadores pelo mesmo objetivo foram positivos para a minha experiência.	-0,667
Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,333
Compartilhamento	A minha equipe compartilhou recursos e informações de forma ideal para conquistar os nossos objetivos.	-0,667
Presença Física	Durante o jogo, eu estava mais atento e comprometido com o ambiente do jogo do que com o seu entorno.	0,5

Tabela 8.2: Índice de validade de conteúdo para os critérios de Experiência de Usuário

Critério	Item	IVC
Volume Ajustável	É possível ajustar ou silenciar o volume do jogo, diferenciando entre os volumes dos efeitos sonoros, trilhas sonoras e narração.	-0,778
Proteção de Erros	O jogo apresentou erros que não afetaram a continuidade do jogo, mas levaram a comportamentos inesperados pelo sistema.	-0,111
Reconhecimento de Aprendibilidade	Aprendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.	0,333
Ausência de Erros Disruptivos	Presenciei algum erro que interrompeu ou terminou o jogo de forma inesperada.	0,111
Acessibilidade para Pessoas com Daltonismo	O jogo não passou informações e sinais importantes para o andamento do jogo exclusivamente por cor.	-0,555
Textos Bem Formataados	O texto estava bem formatado e contrastava com o plano de fundo, facilitando a sua leitura.	0,111
Aprendizagem	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo um domínio dos controles, antes que se tornasse desafiador.	0,111
Acessibilidade para Pessoas com Daltonismo	O jogo não dependeu de narrações, efeitos, músicas ou outros efeitos sonoros para passar informações importantes ao seu andamento.	-0,556
Nitidez da Imagem	O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	0,111
Linguagem Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	0,556
Atratividade	Achei a interface do jogo bonita, o aspecto visual ou sonoro cativou-me de uma forma positiva.	0,111

Tabela 8.3: Índice de Validade de Conteúdo para os critérios de Motivação ao Aprendizado

Critério	Item	IVC
Humor	Houve momentos de descontração durante o jogo.	0,111
Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,556
Demanda de Habilidade	Senti compatibilidade entre as demandas do jogo e as minhas habilidades para superá-las.	-0,333
Intervalo de Reforços Positivos	As recompensas dadas pelo jogo ficam progressivamente mais difíceis de conquistar.	-0,334
Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre o jogo.	0,111
Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,778
Recompensas para Tarefas Entediantes	As tarefas que considerei entediantes eram recompensadas de forma evidente e previsível.	-0,111

Autoconfiança	Tive confiança e independência suficientes para aplicar o conhecimento adquirido pelo jogo em situações realísticas.	-0,111
Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,333
Objetivo do Jogo	Achei os objetivos do jogo claros, inclusive os sub-objetivos ou a otimização do desempenho das tarefas.	0,111
Concretude	Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	0,333
Temática Relevante	A temática e a narrativa são relevantes para o aprendizado oferecido pelo jogo.	0,333
Elogios a Resultados Positivos	Recebia elogios por parte do jogo ao completar tarefas ou ter um bom desempenho no jogo.	-0,111
Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,333
Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,556
Recompensas para Tarefas Interessantes	Ao completar tarefas interessantes, o jogo me recompensava inesperadamente, variando de alguma forma na recompensa que era dada.	-0,111
Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados para avaliar o meu desempenho.	0,556
Conforto	O jogo não me causou nenhum momento em que fiquei fisicamente desconfortável.	0,778
Moderação de Expectativas	Senti que era possível ter um desempenho legal no jogo mediante o meu esforço.	-0,333
Evitar Influências Negativas	Ao jogar, não senti uma pressão ao ponto de que meu bem-estar e desempenho no jogo fossem prejudicados.	-0,556

Tabela 8.4: Índice de validade de conteúdo para cada área de avaliação

Área de Avaliação	IVC
Jogabilidade	-0,231
Experiência do Usuário	-0,0505
Motivação ao Aprendizado	0,0556

Esta pesquisa desenvolveu uma alternativa ao método tradicional de validade de conteúdo proposto por LAWSHE (1975). Foi aplicado o cálculo de grau de consenso, técnica derivada da lógica Fuzzy, para determinar se houve um consenso em relação à essencialidade de cada item para a avaliação de jogos educacionais digitais. Os itens aprovados, determinados seguindo o critério de consenso do Tipo 1 de lógica Fuzzy, foram especificados na Tabela 8.5.

Tabela 8.5: Conjunto de itens cujo consenso foi de que o item é essencial

Critério	Item	Distância ao Consenso	Grau de Certeza Médio	Valor p	Quantidade de permutações
Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,16	0,85	0,0547	18
Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,07	0,93	0	15
Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,12	0,89	0	17
Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,12	0,89	0	17
Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados para avaliar o meu desempenho.	0,12	0,89	0	17
Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,16	0,85	0,0547	18
Concretude	Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	0,16	0,85	0	18

Dificuldade	Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,16	0,85	0	18
Curiosidade	Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre o jogo.	0,18	0,84	0,0522	19
Linguagem	Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	0,15	0,86	0	17
Inteligibilidade		Aprendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.	0,18	0,84	0,0547	18
Capacidade de Aprendizado	de	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo um domínio dos controles, antes que se tornasse desafiador.	0,18	0,84	0	19
Nitidez da Imagem		O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	0,18	0,84	0,0522	19

A tabela exhibe a quantidade de permutações feitas para calcular o valor p do consenso em relação à essencialidade de cada item. O valor de permutações foi determinado por uma análise de potência e o valor p foi calculado como a quantidade de vezes que uma matriz com dados pseudo aleatórios teve um grau de consenso superior àquele observado.

A Tabela 8.5 lista apenas os itens cujo consenso foi de essencialidade à avaliação. A distância ao consenso e grau de certeza de todos os itens do formulário, bem como a alternativa mais respondida de cada, se encontra no apêndice desta dissertação (Tabelas D.4, D.9, D.13).

Estes resultados apresentados foram usados para determinar o conteúdo do formulário de avaliação. Após a administração do formulário em questão foram feitas diversas análises sobre as respostas, que serão apresentados na seção a seguir.

### 8.1.2 Resultados Posteriores à Aplicação do Formulário

O formulário de avaliação à reação do jogo educacional digital foi distribuído presencial e digitalmente para preenchimento após atividades envolvendo jogos educacionais digitais. Foram coletadas um total de 65 respostas, correspondendo a avaliação de 4 jogos educacionais: Cred Carbon Game (SOUSA *et al.*, 2023), DiscoverX (XEXÉO *et al.*, 2022), Mapa do Tesouro (NASCIMENTO, 2023) e Screener Timeline (REIS, 2023). Esse total de respostas, apesar de modesto, cumpre com os requisitos encontrados na literatura para a validação do formulário. Entretanto, seria desejável ter mais respostas; portanto, os impactos da quantidade de respondentes foram detalhados na seção de Discussão (Seção 8.2) e, posteriormente, identificado como uma melhoria para trabalhos futuros.

As respostas obtidas nesta pesquisa foram então usadas para analisar o formulário quanto à sua confiabilidade e validade como instrumento de medição. Foi feita também uma comparação entre grupos categóricos. Para auxiliar na discussão sobre

o padrão de respostas de cada item, foi elaborado um resumo dos dados estatísticos descritivos. Os dados estatísticos descritivos às respostas do formulário estão dispostos na tabela abaixo.

Tabela 8.6: Dados estatísticos descritivos das respostas ao formulário de avaliação da reação a jogos educacionais digitais.

Critério	Item	Média	Desvio padrão	Primeiro quartil (25%)	Mediana	Terceiro quartil (75%)
Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,523	1,17	0	1	1
Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	1,03	1,10	1	1	2
Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,692	1,30	0	1	2
Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,641	1,25	0	1	2
Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados para avaliar o meu desempenho.	0,862	1,33	0	1	2
Concretude	Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	0,662	1,30	0	1	2
Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,215	1,51	-1	0	2
Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,185	1,50	-1	0	1
Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,569	1,38	-1	1	2
Linguagem Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	0,846	1,51	0	2	2
Inteligibilidade	Aprendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.	0,2	1,20	-1	0	1
Capacidade de Aprendizado	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,523	1,50	-1	1	2
Nitidez da Imagem	O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	1,36	1,24	1	2	2

Ao converter as respostas da escala em valores numéricos, a escala variou entre -2 e 2, esses extremos representam as opções de discordo fortemente e concordo fortemente, respectivamente. Há uma distância unitária entre as opções, porém é importante reforçar que a distância entre as classificações não é quantificável. Os dados descritivos servem então apenas para nortear uma discussão a respeito das respostas.

Dados estatísticos descritivos permitem, entre outras finalidades, a visualização de medidas de dispersão para ver o quanto as respostas variam na prática. A variância nas respostas é algo essencial a instrumentos psicométricos. É desejado que itens que correspondem ao mesmo fator variem conjuntamente; para verificar isso, há de ser determinada a confiabilidade do instrumento de medição.

## Confiabilidade

A confiabilidade do formulário foi testada mediante três métricas de consistência interna, cujos valores estão listados na Tabela 8.8. O limite superior e inferior do intervalo de confiança de 95% foi de 0,821 e 0,914, respectivamente.



Tabela 8.7: Alfas de Cronbach caso determinados itens fossem descartados.

Critério	Item	Alfa de Cronbach ao descartar o item
Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,874
Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,88
Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,879
Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,870
Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados para avaliar o meu desempenho.	0,883
Concretude	Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	0,875
Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,978
Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,877
Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre o jogo.	0,876
Linguagem Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	0,895
Inteligibilidade	Apreendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.	0,881
Capacidade de Aprendizado	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo um domínio dos controles, antes que se tornasse desafiador.	0,876
Nitidez da Imagem	O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	0,895

O Alfa de Cronbach demonstra um pequeno grau de variação considerando as métricas caso um dos itens fosse descartado (Tabela 8.7). O valor máximo é de aproximadamente 0,90 ao descartar o item do *Nitidez da Imagem* enquanto há o mínimo de consistência interna nesse cenário ao remover o item de *Ritmo do Jogo*.

Quanto às demais métricas de consistência interna englobadas nessa dissertação, o valor de confiabilidade *split-half* foi de 0,807, enquanto o valor corrigido Spearman-Brown de confiabilidade *split-half* foi de 0,893.

Tabela 8.8: Resultados de confiabilidade do instrumento de medição desta pesquisa.

Métrica	Valor
Alfa de Cronbach	0,872
Split-Half	0,807
Valor corrigido Spearman-Brown	0,893

Todas as métricas de consistência interna previstas neste experimento obtiveram resultados no intervalo de 0,8 a 0,9. Assim, é possível concluir que há uma boa consistência interna entre os itens do formulário.

## Validade de Constructos (Análise Fatorial)

Antes da condução da análise fatorial foi averiguada a adequabilidade dos dados para a análise. A métrica Kaiser-Meyer-Olkin, KMO, foi usada para determinar a aptidão dos dados para a realização de uma análise fatorial. O valor do KMO para o conjunto de itens em sua totalidade foi de 0,77. Segue uma tabela dos valores de KMO para cada item do formulário (Tabela 8.9). Foi exposta como referência uma classificação dos valores obtidos de KMO. Os rótulos dados à classificação são

versões traduzidas daquelas propostas pelos autores da métrica KMO, em KAISER e RICE (1974).

Tabela 8.9: Métrica Kaiser-Meyer-Olkin dos itens do formulário de avaliação.

Critério	Item	KMO	Classificação
Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,82	Bom
Linguagem Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	0,73	Razoável
Nitidez da Imagem	O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	0,76	Razoável
Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,80	Bom
Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,66	Medíocre
Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,82	Bom
Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados para avaliar o meu desempenho.	0,79	Razoável
Concretude	Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	0,85	Bom
Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,92	Excelente
Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,69	Medíocre
Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,78	Razoável
Inteligibilidade	Aprendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.	0,75	Razoável
Capacidade de Aprendizado	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,42	Inapto para análises fatori-ais

O formulário, em sua totalidade, obteve um valor KMO razoável. Individualmente, a maioria dos itens alternou entre valores razoáveis e excelentes; apenas um item, de capacidade de aprendizado, apresentou inaptidão seguindo este critério. Estes resultados significam que uma proporção significativa de variância dos itens deve ser atribuída a uma variância conjunta, e conseqüentemente, há indícios de fatores subjacentes aos itens.

O critério de esfericidade de Bartlett também foi utilizado para verificar a adequabilidade dos dados. BARTLETT (1951) afirma que a sua estatística de teste possui uma distribuição qui-quadrado com  $\frac{p(p-1)}{2}$  graus de liberdade, no qual  $p$  representa a quantidade de variáveis. Segue então que os graus de liberdade para a esfericidade de Bartlett nessa pesquisa é 78.

$$gl = \frac{p(p-1)}{2} = \frac{13(13-1)}{2} = 78 \quad (8.1)$$

É possível consultar a tabela de distribuição de qui-quadrado para este par de valores a partir dos graus de liberdade calculados acima (Fórmula 8.1) e o valor alfa de poder estatístico selecionado para este experimento. O valor de qui-quadrado na distribuição assumindo um grau de liberdade de 78 e nível de significância, alfa, de 0,05, é 99,6. O valor qui-quadrado na distribuição resultante do teste de esfericidade de Bartlett foi de 360,5. Este valor é significativamente acima do valor crítico de qui-quadrado, denotando que a hipótese nula deve ser rejeitada.

A outra métrica de saída do teste de Bartlett, o valor de significância estatística

$p$ , foi de  $6,64e^{-38}$ . Este valor fica a algumas ordens de grandeza abaixo do valor de significância, alfa, escolhido para esse experimento, 0,05. Ambas as métricas, de qui-quadrado e o valor  $p$ , indicam que a hipótese nula deve ser rejeitada. Portanto, a matriz de correlação das respostas ao formulário de avaliação não se assemelham a uma matriz de identidade com as mesmas dimensões.

As duas métricas de verificação de adequabilidade dos dados a análise fatorial apresentam resultados satisfatórios. Portanto, foram conduzidas análises fatoriais confirmatórias e exploratórias para testar a matriz fatorial teórica e elaborar um novo conjunto fatorial, respectivamente.

A análise fatorial confirmatória conduzida para esta pesquisa produziu uma matriz de cargas fatoriais, apresentada abaixo na Tabela 8.10. Para análises fatoriais confirmatórias, o modelo fatorial teórico é usado para fixar em 0 a carga fatorial dos fatores além daquele que teoricamente representa cada item.

Tabela 8.10: Resultado da análise fatorial confirmatória.

Nome do critério	Jogabilidade	Experiência de Usuário	Motivação ao Aprendizado
Animação	0,97	0	0
Inteligibilidade	0	0,65	0
Capacidade de Aprendizado	0	0,86	0
Nitidez da Imagem	0	0,84	0
Linguagem Compreensível	0	0,93	0
Mecânicas Relevantes	0	0	0,75
Experiência	0	0	1,10
Curiosidade Sensorial	0	0	1,12
Curiosidade Cognitiva	0	0	0,66
Dificuldade Crescente	0	0	0,98
Requisitos de Aprendizado	0	0	0,69
Concretude	0	0	1,10
Ritmo do Jogo	0	0	0,20

Variáveis indicadoras com cargas fatoriais acima de 1,0 são consideradas impróprias, pois o limite da carga fatorial deve ficar restrita ao intervalo de 0,0 a 1,0. A Tabela 8.10 revela que três itens correspondentes ao fator de motivação ao aprendizado são impróprios para a verificação do ajuste dos dados ao modelo.

Os graus de liberdade da adequação dos dados ao modelo (*goodness-of-fit*) e as métricas referentes à adequabilidade selecionadas para essa pesquisa encontram-se na Tabela 8.11:

Tabela 8.11: Métricas de adequabilidade dos dados ao modelo teórico (*goodness-of-fit*).

Graus de Liberdade	$\chi^2$	valor $p$	GFI	AGFI
62	141,23	0,00	0,655	0,566

Todos os índices de adequabilidade dos dados ao modelo teórico apresentaram resultados insatisfatórios. O valor de qui-quadrado de aproximadamente 141 é significativamente superior ao valor crítico correspondente dado os graus de liberdade, 62, assumindo um poder estatístico de 0,05 ou até mesmo de 0,1. No contexto da verificação de adequabilidade (*goodness-of-fit*), é esperado que a hipótese nula seja

aceita. Portanto, o valor da estatística de teste deveria ser inferior ao equivalente da distribuição qui-quadrado.

Outrossim, o valor da métrica de Adjusted Goodness of Fit Index (AGFI), ficou distante de um valor aceitável para considerar o modelo teórico adequado aos resultados. O valor foi de aproximadamente 0,566, enquanto é esperado que o índice seja de ao menos 0,9.

Os valores do teste qui-quadrado em relação às cargas fatoriais dos fatores estão expostos na Tabela 8.12:

Tabela 8.12: Teste qui-quadrado das estimativas da análise fatorial confirmatória.

Fator	$\chi^2$	valor-p
Jogabilidade	11,62	0,476
Experiência do Usuário	7,54	0,82
Motivação ao Aprendizado	5,07	0,939

A adequabilidade do modelo de medição também foi calculada mediante as métricas de confiabilidade composta dos construtos e de variância média explicada. Os resultados desta verificação de adequabilidade estão apresentados na Tabela 8.13. Há uma ressalva ao cálculo das métricas, o fator de motivação ao aprendizado possui variáveis indicadoras com valores impróprios, como destacado anteriormente. Consequentemente, esses valores levariam a distorções nos resultados, e, portanto, não puderam ser inclusos no cálculo das métricas da adequabilidade.

Tabela 8.13: Métricas de adequabilidade do modelo de medição.

Fator	Confiabilidade composta do constructo	Variância média extraída
Jogabilidade	0,938	0,938
Experiência de usuário	0,893	0,681
Motivação ao aprendizado	0,808	0,492

O fator de jogabilidade possui apenas uma variável indicadora, então era esperado que a confiabilidade e a variância explicada média fossem altos. O fator de experiência de usuário atingiu níveis desejados para ambas as métricas, pois a confiabilidade composta e variância média extraída foram consideravelmente superiores ao limite mínimo de 0,5 sugeridos em HAIR *et al.* (2009). Por último, o constructo de motivação ao aprendizado conseguiu um nível adequado em sua confiabilidade composta, mas ficou marginalmente abaixo do limite mínimo em sua variância média extraída.

A última verificação de adequabilidade de um modelo é a do modelo estrutural (structural model fit), cujos resultados encontram-se na Tabela 8.14. Nela, é testada a significância estatística de cada coeficiente de estimativa da análise confirmatória. Segue o resultado para cada estimativa, considerando um nível de significância tradicional,  $\alpha = 0,05$  e um nível de significância conservador,  $\alpha = 0,01$ .

Tabela 8.14: Métricas de adequabilidade dos dados ao modelo estrutural (*structural model fit*). (Graus de liberdade = 64)

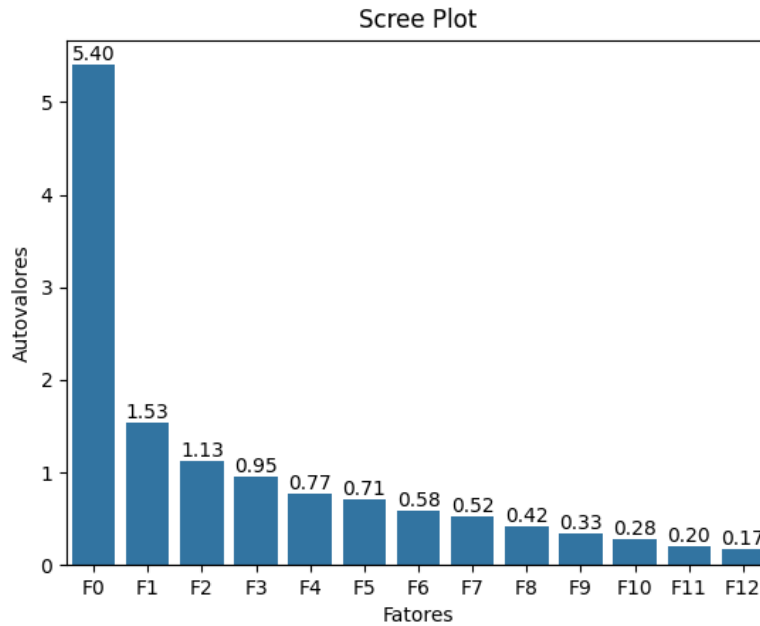
Critério	Item	Estatística t
Animação	Senti entusiasmo durante a minha experiência com o jogo.	N/A
Linguagem Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	4,14
Nitidez da Imagem	O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	4,23
Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	4,26
Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	4,14
Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	4,60
Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados pra avaliar o meu desempenho.	5,98
Concretude	Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	6,10
Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	5,79
Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	3,69
Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	4,33
Inteligibilidade	Aprendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.	5,99
Capacidade de Aprendizado	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	1,20

Os valores críticos da distribuição-t para 64 graus de liberdade são 1,67 para  $\alpha = 0,05$  e 2,39 para  $\alpha = 0,01$ , considerando que os testes são de apenas um dos extremos. A maioria das estimativas teve um valor t superior ao valor crítico, indicando significância estatística. Uma das exceções foi do critério de animação, cujo valor t não pôde ser calculado corretamente por ser a única variável do fator de jogabilidade. A outra foi do item de *capacidade de aprendizado*, no qual a hipótese nula foi aceita mesmo para o nível de significância de  $\alpha = 0,05$ .

Este experimento também conduziu uma análise fatorial de caráter exploratório. Um total de quatro rodadas de análise fatorial exploratória foram executadas ao longo dessa pesquisa. A primeira rodada foi conduzida com o conjunto completo das variáveis observadas na aplicação do formulário de avaliação. As execuções subsequentes foram compostas apenas das variáveis cuja carga fatorial foi suficientemente alta para algum dos fatores selecionados nas execuções que a antecederam. A necessidade de múltiplas rodadas de análise fatorial foi motivada pela exclusão de variáveis em rodadas anteriores. Caso um item não corresponda a um fator, uma nova análise deve ser feita com apenas àqueles que de fato medem um dos fatores. Nas situações na qual os critérios de seleção de fatores dessa pesquisa: o critério de Kaiser, teste de scree e análise paralela, divergiram na quantidade de fatores selecionado, foi feita uma interpretação fatorial para cada cenário. Caso a interpretação fosse aplicável, uma nova análise fatorial foi conduzida utilizando os itens que compuseram os fatores daquela análise.

A análise fatorial exploratória resultou na identificação de dois fatores usando o critério do Scree Plot (Figuras 8.1 e 8.4). A mesma quantidade de fatores foi

Figura 8.1: Gráfico de barras dos autovalores da primeira rodada de análise fatorial



selecionada pelo método de análise paralela, conduzida usando a matriz de correlação com todas as variáveis da avaliação. Os resultados da análise paralela estão disponíveis na Figura 8.2, nos quais os fatores cujo rótulo é "Adjusted Ev (retained)" representam aqueles selecionados pela análise. Por último, seguindo o critério de Kaiser, temos um total de três fatores selecionados. Apenas os três primeiros fatores possuem um autovalor superior a 1,0, o quarto fator na ordem decrescente tem um autovalor de 0,95.

Tabela 8.15: Resultados da análise paralela para a seleção de Fatores

Fator	Autovalor ajustado	Autovalor não ajustado	Erro sistemático (Bias)
0	4,39	5,44	1,05
1	0,14	0,95	0,80

Tabela 8.16: Matriz fatorial da pesquisa no cenário de três fatores selecionados.

Fator	Critério	Item	Carga Fatorial
F0	Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,31
	Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,50
	Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,78
	Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,47
	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,32
F1	Capacidade de Aprendizado	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,37
	Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,53
	Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,36

	Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados para avaliar o meu desempenho.	0,88
F2	Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,71
	Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,35
	Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,34
	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,56

O cenário com três fatores selecionados (Tabela 8.16) resultou em uma redução de 13 para 9 variáveis, pois as demais não tiveram uma carga fatorial suficientemente alta para algum dos fatores. Os critérios removidos das rodadas seguintes são: linguagem compreensível, nitidez da imagem, inteligibilidade e concretude. Outras variáveis mostraram uma carga significativa para múltiplos fatores. Nestes casos, a variável é interpretada como representante para o fator ao qual possui a carga mais elevada.

A perda nos itens do formulário é ainda mais drástica, assumindo que apenas dois fatores devem ser selecionados (Tabela 8.17). Neste caso, restariam apenas 8 itens. A diferença entre os cenários é que o critério de animação está incluso apenas quando há um terceiro fator.

Tabela 8.17: Matriz fatorial da pesquisa no cenário de dois fatores selecionados.

Fator	Critério	Item	Carga Fatorial
F0	Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,31
	Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,50
	Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,78
	Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,47
	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,32
F1	Capacidade de Aprendizado	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,37
	Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,53
	Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,36
	Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados pra avaliar o meu desempenho.	0,88

Devido à remoção dos critérios, faz-se necessário realizar uma nova instância da análise fatorial a partir de cada uma das interpretações fatoriais: a de dois e a de três fatores. Desta vez, a análise foi feita sem as variáveis que não obtiveram um nível suficiente para os fatores selecionados.

As rodadas posteriores seguiram os mesmos procedimentos da primeira rodada, a diferença esteve apenas nas variáveis incluídas nas análises. A Figura G.1 exhibe todas as execuções das análises fatoriais. O diagrama indica como *Sucesso* as interpretações fatoriais que mantiveram todas as variáveis analisadas naquela execução. As análises fatoriais cujo fator de seleção indicou a existência de apenas um fator, foram desconsideradas e indicadas no diagrama com um sinal de *proibido*, pois reduziriam o escopo a uma quantidade diminuta de variáveis.

Figura 8.2: Visualização da análise paralela, referente à primeira rodada .

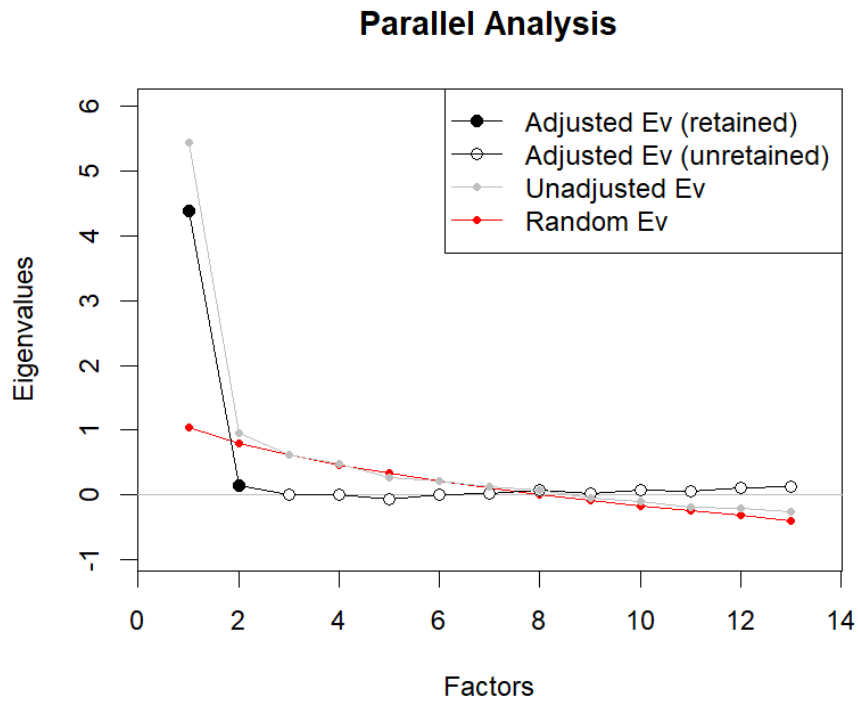


Figura 8.3: Scatterplot das cargas fatoriais, resultante da primeira rodada de análise de fatorial exploratória, assumindo uma matriz com três fatores.

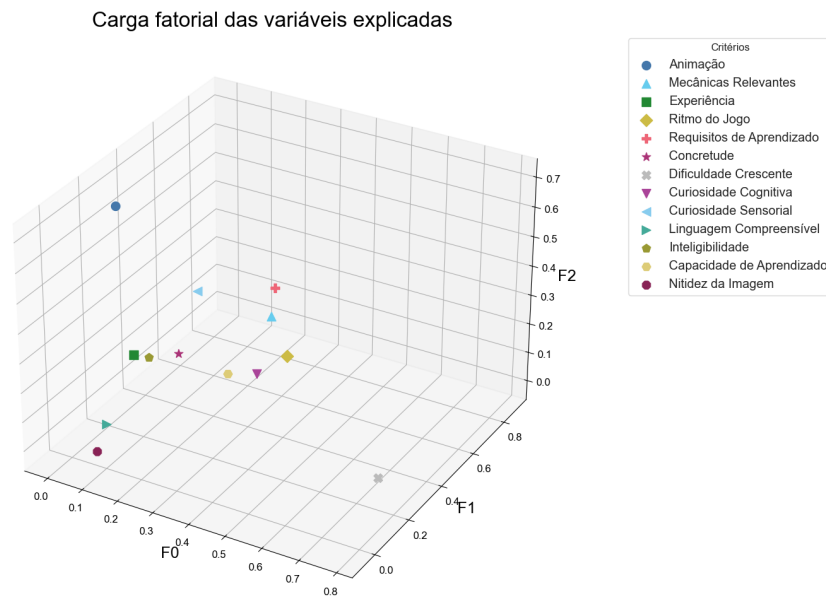
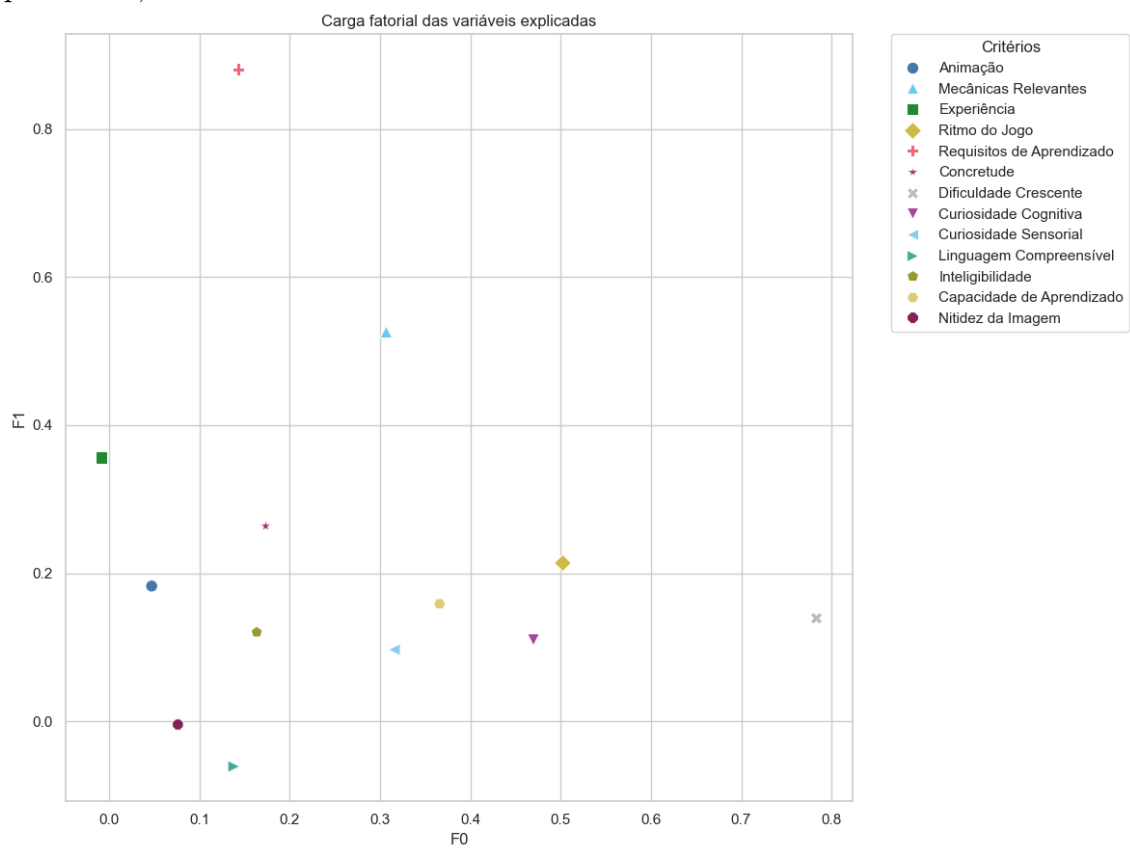




Figura 8.4: Scatterplot dos autovalores da primeira rodada de análise fatorial exploratória, assumindo dois fatores.



O diagrama informa também o valor das métricas de adequabilidade para análise fatorial de cada execução de avaliação fatorial feita. Além disso, são indicadas a quantidade de variáveis pertencentes àquela instância da análise. Um total de 6 instâncias de análise fatorial exploratória foram conduzidas nesta pesquisa. A matriz fatorial de cada uma das análises fatoriais exploratórias encontra-se no Apêndice G. Lá estão listados, para cada análise, os critérios que correspondem a cada fator selecionado e suas respectivas cargas fatoriais.

A interpretação fatorial escolhida pelo autor como a que melhor representou o modelo de avaliação desta pesquisa possui quatro fatores. Esta interpretação é fruto de uma das análises feitas na segunda rodada, feita a partir do cenário no qual três fatores foram selecionados na análise inicial (Figura 8.6). A seleção de quatro fatores deu-se pelo critério estabelecido pela análise paralela, cujo resultado está disponível na Figura 8.6.

Os motivos pelo qual essa interpretação foi a selecionada serão elaborados na próxima seção, de Discussão (Seção 8.2). Seguem os resultados para a interpretação fatorial definida como a ideal.

Tabela 8.18: Cargas fatoriais significativas para os fatores selecionados pela análise paralela na segunda rodada, no cenário no qual a primeira análise teve três fatores selecionados.

Fator	Critério	Item	Carga Fatorial
F0	Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,82
	Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,54
	Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,61
	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,39
F1	Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,83
	Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,39
	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,50
F2	Capacidade de Aprendizado	O jogo ensinou me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,33
	Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,31
	Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,58
	Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados pra avaliar o meu desempenho.	0,73
	Capacidade de Aprendizado	O jogo ensinou me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,33
F3	Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,51
	Capacidade de Aprendizado	O jogo ensinou me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,75

Tabela 8.19: Matriz fatorial escolhida como a ideal nesta pesquisa, rotulada e interpretada.

Fator	Critério	Item	Carga Fatorial
Jogabilidade	Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,82
	Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,54
	Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,61
Diversão	Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,83
	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,50
Scaffolding	Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,31

Experiência		Conseguí estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,58
Requisitos de Aprendizado		Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados pra avaliar o meu desempenho.	0,73
Capacidade de Aprendizado	-/-	O jogo ensinou me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,75

O fator cujo autovalor foi mais expressivo reuniu os critérios de "ritmo do jogo", "dificuldade crescente" e "curiosidade cognitiva". Esses três critérios são complementares, pois um jogo que cumpre este conjunto de critérios proporciona uma curva de aprendizado razoável ao jogador. Isto costuma ser alcançado em jogos que permitem estratégias variadas para conseguir um bom desempenho, exigindo uma motivação a aprender sobre como reagir aos diferentes cenários. Este fator foi denominado de *jogabilidade*, pois a moderação na dificuldade e na apresentação de novos desafios é intrínseco a um jogo com boa jogabilidade.

O fator de segundo maior autovalor foi dado o rótulo de *diversão*, pois a animação e curiosidade sensorial descrevem sensações positivas características a experiências divertidas. O terceiro fator foi interpretado como *scaffolding*. Finalmente, o quarto fator é homônimo ao critério que o compõe, pois um termo mais amplo poderia levar a interpretações errôneas sobre seu escopo. Fica a ressalva de que o critério de capacidade de aprendizado seria um ótimo candidato a pertencer ao fator de Scaffolding, pois o aprendizado mediante um tutorial ou manual retratam maneiras na qual Scaffolding pode ser aplicado em jogos segundo OBIKWELU (2017).

Figura 8.5: Gráfico de barras dos autovalores da segunda rodada de análise fatorial exploratória, assumindo uma matriz fatorial de três fatores na primeira.

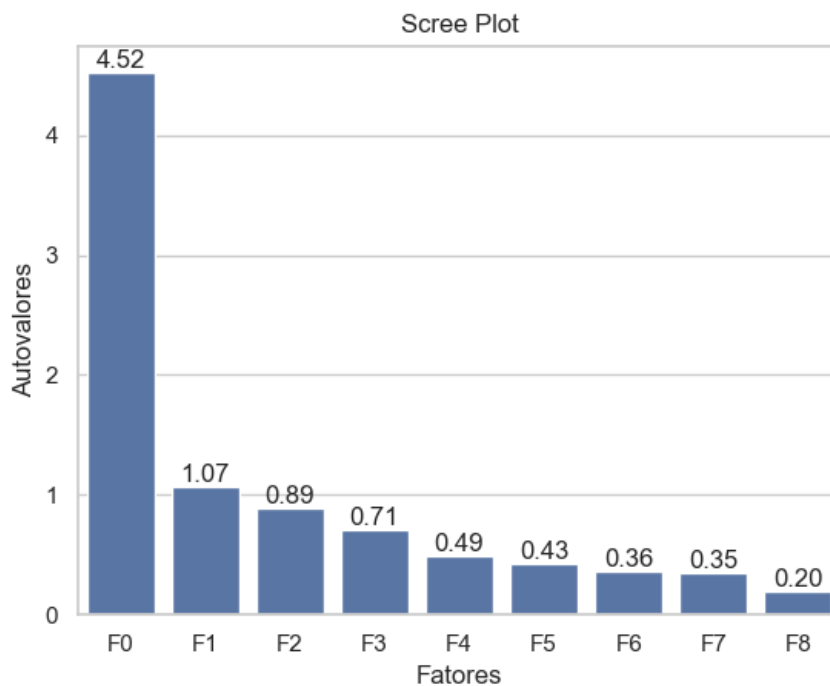
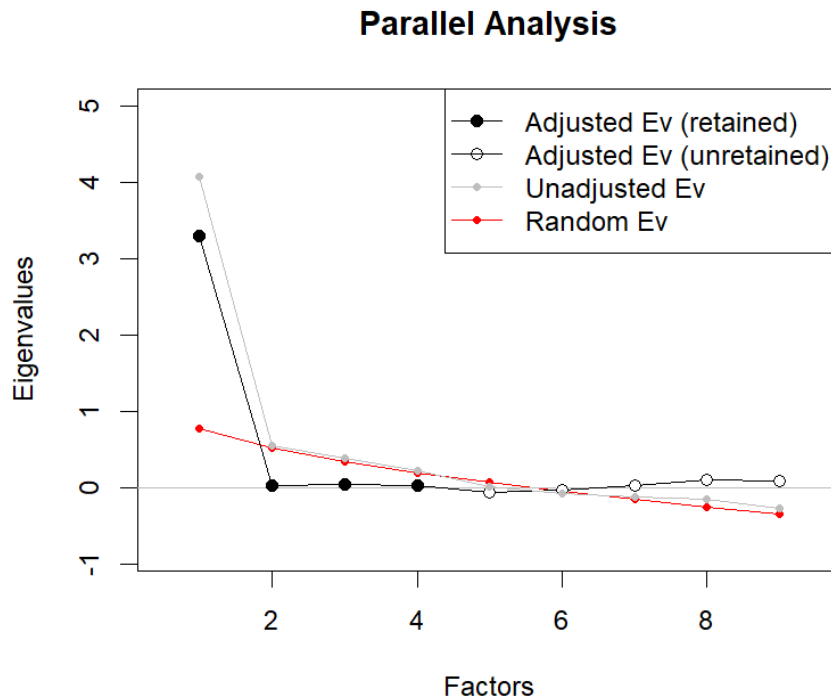


Figura 8.6: Visualização da segunda rodada de análise paralela, assumindo três fatores.



### Comparação Entre Grupos Categóricos

A comparação entre grupos categóricos testou a hipótese de que a distribuição de resposta a um item do formulário varia significativamente conforme as categorias de um determinado dado. As categorias correspondiam a características sociodemográficas ou de hábitos de jogo; apenas as categorias com pelo menos 5 indivíduos foram incluídas na análise.

Dado categórico	Item	Estatística H do teste Kruskal-Wallis	Valor p
Faixa Etária (graus de liberdade = 2)	Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	10,0	0,00673
	O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	6,46	0,0395
Área de Formação (graus de liberdade = 1)	Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	8,27	0,00403
	O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	4,06	0,0438
Maior titulação acadêmica (graus de liberdade = 3)	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados pra avaliar o meu desempenho.	8,14	0,0432
Renda mensal familiar (graus de liberdade = 5)	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	12,0	0,0354
	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	13,1	0,0226
Já jogou um jogo educacional? (graus de liberdade = 1)	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	5,25	0,0219

	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	4,23	0,0397
Concorda que jogos educacionais são divertidos? (graus de liberdade = 2)	Senti entusiasmo durante a minha experiência com o jogo.	6,92	0,0314
	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados pra avaliar o meu desempenho.	6,93	0,0313
	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	9,78	0,00753
	Aprendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.	7,64	0,0219
	O jogo começou fácil e gradativamente ficou mais difícil.	7,52	0,0233

Um total de 14 combinações das 104 testadas possuía diferenças significativas na distribuição de respostas entre as categorias. As 14 combinações com diferenças significativas pertencem a seis dos oito dados categóricos, do qual a pergunta sobre jogos educacionais serem divertidos contribuiu com 5 das combinações. Na perspectiva dos itens, apenas três dos treze que compõem o formulário não teve respostas significativamente diferentes para algum dado categórico.

Um exemplo de combinação de item do formulário e dado categórico cuja hipótese nula foi rejeitada foi do item de *animação*, "Senti entusiasmo durante a minha experiência com o jogo.", em combinação com a pergunta de hábitos de jogo sobre jogos educacionais serem divertidos. Para esta combinação, a estatística de teste  $H = 18,63$ , acima do valor referente aos graus de liberdade de 6 e de métrica  $\alpha = 0,05$ , 12.592. O valor-p da tupla sob análise, 0,0048 também esteve abaixo do alfa estabelecido para essa pesquisa, 0,05.

## 8.2 Discussão

O objetivo desta pesquisa foi o desenvolvimento de um modelo de avaliação da reação de alunos a jogos educacionais digitais. O modelo foi desenvolvido em conjunto a um formulário, cuja confiabilidade e validade há de ser comprovada para garantir que está medindo aquilo ao qual se propõe medir. Este capítulo discorrerá sobre os resultados da pesquisa, além de abordar percalços identificados em sua trajetória. Primeiro, será discutida a etapa de validade de conteúdo, cuja análise foi feita antes da aplicação do formulário de avaliação. Em seguida serão abordados a confiabilidade e a validade de construtos.

### 8.2.1 Validade de Conteúdo

O modelo de avaliação inicial desta dissertação foi discutido com especialistas do laboratório de pesquisa LUDES - UFRJ em um focus group. O modelo foi apresentado e em seguida foram discutidas sugestões para mudanças no modelo de avaliação. As contribuições mais notáveis foram quanto a área de experiência

de usuário, no qual um especialista sugeriu a inclusão de um item de *linguagem compreensível*, pertinente e considerado essencial pelos especialistas posteriormente, mas que não havia sido encontrado na literatura. Outro especialista notou que até então a área de avaliação de experiência de usuário tinha poucos critérios referentes a acessibilidade, e apontou a tese de REIS (2020) como uma referência interessante para sugestões sobre este tema aplicado a jogos educacionais digitais.

Na etapa subsequente desta pesquisa, na validação de conteúdo do formulário de avaliação, diversos comentários foram feitos por especialistas. Os comentários abrangeram o conteúdo dos critérios do modelo, entre os quais houve sugestões pertinentes a alterações no título ou descrição de critérios. Por exemplo, foi sugerido a mudança do critério de *Previsibilidade das Interações com o Sistema* para *Imprevisibilidade das Interações com o Sistema*, visto que a definição deste conceito preza justamente pela incerteza. Outra contribuição importante dos comentários foi a crítica a definições cuja explicação se favorecia de palavras já inclusas em seu título. Esta redundância deve ser evitada ao elaborar a definição de um termo.

A etapa de validade de conteúdo teve como um dos seus maiores obstáculos a obtenção de respostas suficientes. Convites foram enviados por mensagem em grupos cuja pauta é a ludologia, o estudo de jogos. Convites também foram enviados individualmente para especialistas da área de jogos e professores com experiência prática na aplicação de jogos educacionais. A taxa de resposta dos formulários impressos, entregues diretamente aos respondentes, foi muito superior às tentativas de distribuição em grupos digitais.

Alguns fatores podem ter influenciado negativamente a taxa de resposta. Entre eles estão a especificidade do tópico de pesquisa e a quantidade de itens sob-avaliação. Por mais que o formato padronizado do formulário tenha sido pensado para simplificar o processo e abreviar a sua duração, a leitura e avaliação de itens repetidamente pode ser um processo entediante, desincentivando-os a completar o questionário.

Outra dificuldade foi observada na etapa de validade de conteúdo, desta vez quanto ao rigor elevado do critério. Em um cenário no qual um total de 10 especialistas participem da validação de conteúdo de um questionário, é exigido que ao menos 8,1 especialistas concordem que o item é essencial, pois o valor crítico do IVC neste cenário é de 0,62 (LAWSHE, 1975). Uma quantidade ínfima de itens proposto para o formulário de avaliação de jogos educacionais digitais foi considerado apto seguindo a definição imposta por LAWSHE (1975). Esse resultado ocorreu devido a uma falta de unanimidade que os itens *essenciais* à medição, mesmo quando era unânime que o item não era *desnecessário*. Isto é, as respostas dos especialistas nesses casos está distribuída majoritariamente entre as categorias “útil, mas não essencial” e “essencial”.

O objetivo da etapa de validade de conteúdo é que os itens incluídos nos formulários de avaliação tenham seu conteúdo considerado por consenso entre os especialistas como *essencial*. O consenso e preferência em decisões de grupo podem ser medidas usando lógica fuzzy. Portanto, como alternativa à medida tradicional de IVC, foi identificada a preferência dos especialistas e calculado a distância ao consenso para cada item candidato. Os itens cujo consenso foi de essencialidade estão listados na Tabela 8.5. No total, 13 itens foram aprovados seguindo as métricas de consenso de lógica fuzzy. Há de se notar que a maioria das situações de preferência de grupo atingem consenso (ROSS, 2009).

Um total de 13 dos 49 itens foi aprovado quanto a sua essencialidade a avaliação de jogos educacionais digitais. A aprovação dos itens foi conforme a definição de consenso fuzzy do Tipo 1. Esta proporção corresponde a aproximadamente 26% dos itens candidatos a entrar no formulário inicialmente.

Esta proporção pode ser interpretada como uma quantidade razoável, pois uma quantidade exorbitante de itens pode causar a desistência de respondentes, assim reduzindo a taxa de respostas do formulário KITCHENHAM e PFLEEGER (2002b). Em contrapartida, a quantidade diminuta levou a uma redução do escopo, pois somente uma das áreas de avaliação se manteve com uma quantidade razoável de itens. Uma quantidade reduzida de itens também impacta negativamente a confiabilidade do instrumento de medição, pois a consistência interna de um formulário aumenta conforme a quantidade de itens englobados pelo formulário cresce (DEVELLIS e THORPE, 2021).

Cabe a ressalva que a predominância dos itens aprovados na etapa de validação de conteúdo foram da área da motivação ao aprendizado, representando 8 dos 13 itens. Sobraram então apenas itens representando o critério de *animação* para a área de jogabilidade e *linguagem compreensível, aprendibilidade, inteligibilidade e nitidez da imagem*, para a área de avaliação de experiência de usuário.

Enquanto o autor reconhece a importância das dimensões de jogabilidade para jogos e de experiência de usuário para sistemas digitais, é certo que a escala ficou sub-representada nestas áreas. É esperado que os itens predominantes de motivação ao aprendizado tenham uma correlação alta entre eles e correspondam aos principais fatores com maior carga na análise fatorial exploratória.

Outro efeito está na análise fatorial confirmatória. Conforme será discutido ao abordar a análise fatorial, a escassez de itens nessas áreas levou a uma insuficiência para se afirmar que a escala desta pesquisa mede os construtos de jogabilidade e de experiência de usuário. O instrumento de medição utilizado na coleta de dados teve seu escopo reduzido após a etapa de validação de conteúdo.

## 8.2.2 Confiabilidade

O instrumento de medição apresentou uma alta consistência interna. Isso é evidenciado pelo valor do Alfa de Cronbach,  $\alpha = 0,872$ , e pelo valor de confiabilidade *Split-Half* ajustado pela fórmula Spearman-Brown, 0,893. A consistência interna mostrou valores altos, mesmo considerando o piso do intervalo de confiança, pois a confiabilidade é significativamente superior a 0,60, definido como um limite inferior aceitável para estudos exploratórios, de acordo com ROBINSON *et al.* (1991).

Um dos efeitos benéficos de uma consistência interna alta é a sua implicação em um poder estatístico mais elevado caso seja assumido um tamanho de amostra fixo DEVELLIS e THORPE (2021). Esta relação é importante para esta pesquisa, visto que o tamanho amostral de resposta à avaliação foi abaixo do desejado.

Instrumentos de medição confiáveis indicam que o instrumento está de fato medindo algum fenômeno latente. A confiabilidade alta, decorrente de correlações fortes entre os itens do formulário, possibilitou que o experimento pudesse ser conduzido apesar de um número relativamente baixo de respondentes.

O valor de Alpha de Cronbach é influenciado pela quantidade de itens no formulário, assumindo que o resto mantenha-se fixo, aumentar a quantidade de itens resulta em uma consistência interna mais elevada (DEVELLIS e THORPE, 2021; HAIR *et al.*, 2009). A quantidade de itens nesta pesquisa foi de apenas 13, portanto, a consistência interna do conjunto de itens certamente não foi inflada meramente pelo seu tamanho.

O resultado satisfatório no aspecto de confiabilidade do formulário foi fundamental para que a sua validade também pudesse ser atestada. A validade de contrutos desta pesquisa foi testada mediante análise fatoriais, discutidas a seguir.

## 8.2.3 Análise Fatorial

### Confirmatória

A análise fatorial confirmatória apresentou resultados insatisfatórios, que serão discutidos adiante. Um dos fatores importantes para a condução da análise é o tamanho amostral.

O tamanho de amostra de respondentes ao formulário de avaliação, 65, cumpre os requisitos mínimos para a condução de uma análise fatorial confirmatória usando o método de estimativa de máxima verossimilhança. É exigido que a amostra seja de no mínimo 50 indivíduos (HAIR *et al.*, 2009). Entretanto, é recomendável que a amostra esteja em um intervalo de 100 a 200 indivíduos (HAIR *et al.*, 2009).

Agrava-se a esta situação a existência de um fator com apenas uma variável indicadora. Análises fatoriais confirmatórias cujo tamanho de amostra for inferior



a 100 devem ter ao menos três variáveis correspondentes a cada fator (DING *et al.*, 1995). Entretanto, o fator de jogabilidade possui apenas o item referente a animação no modelo fatorial inicialmente proposto nessa dissertação. Apesar de um nível adequado de itens correspondentes aos outros fatores, o fator de jogabilidade torna-se inadequado para uma análise fatorial confirmatória nesse contexto.

O fator de motivação ao aprendizado apresentou resultados impróprios para uma análise fatorial confirmatória, apesar de ter uma quantidade significativa de variáveis correspondentes. Isso afetou algumas das análises que compõem o escopo da análise fatorial confirmatória desta pesquisa. Mais especificamente, foram prejudicadas as análises de verificação do ajuste ao modelo teórico e ao modelo de medição.

Em resumo, a análise fatorial confirmatória apresentou diversos resultados que não atenderam as expectativas. Isso evidencia uma incompatibilidade entre os fatores teóricos e os itens que o compõem, apesar de ambos esses componentes terem sido devidamente fundamentados em taxonomias encontradas na literatura.

No caso dos itens, houve uma validação de especialistas quanto a sua essencialidade para a avaliação em questão. A análise fatorial exploratória desta pesquisa encontrou um conjunto de quatro fatores que correspondem a nove dos treze itens do modelo. Em um trabalho futuro, é recomendável fazer uma análise fatorial confirmatória para validar a matriz fatorial proposta a partir da análise exploratória desta pesquisa.

## **Exploratória**

A análise fatorial exploratória identificou a presença de um conjunto de quatro fatores que explicam a correlação entre os itens do formulário. Os fatores foram interpretados e rotulados como: jogabilidade, diversão, *scaffolding* e capacidade ao aprendizado. Coube ao autor rotular os fatores conforme os itens que o compõem. Uma autocrítica aos rótulos atribuídos é que o critério de capacidade ao aprendizado e *scaffolding* possuem uma interseção em teoria, apesar desta pesquisa assumir uma ortogonalidade na matriz fatorial.

Experimentos cujo tamanho amostral é inferior a 100 necessitam que itens tenham uma carga fatorial de ao menos 0,30 para serem considerados para um determinado fator (HAIR *et al.*, 2009). Isso levou à eliminação de alguns critérios do modelo na matriz fatorial final. Os itens que representam os critérios de linguagem compreensível, nitidez da imagem, inteligibilidade e concretude não são suficientemente correlatos com nenhum dos fatores selecionados.

Esta medida de eliminação de variáveis sem cargas fatoriais acima de 0,3 é feita para garantir significância prática. Entretanto, o nível necessário para uma carga fatorial ser considerada estatisticamente relevante é muito mais conservador. No contexto desta pesquisa, apenas variáveis cujas cargas fatoriais fossem acima de 0,7

seriam consideradas, visto que o tamanho da amostra foi de 65 respondentes. Esse limite foi calculado para diversos tamanhos de amostra por DIXON (1992), os limites definidos para diferentes magnitudes de carga fatorial estão apresentados na Tabela 8.21. HAIR *et al.* (2009) cita que no caso de cargas fatoriais, a relevância estatística deve ser apenas um "ponto de partida", e que cargas inferiores a esse limite podem ser consideradas.

Tabela 8.21: Identificação de cargas fatoriais estatisticamente relevantes por tamanho de amostra (DIXON, 1992) (HAIR *et al.*, 2009) .

Carga fatorial	Tamanho de Amostra Necessário para Significância Estatística
0,30	350
0,35	250
0,40	200
0,45	150
0,50	120
0,55	100
0,60	85
0,65	70
0,70	60
0,75	50

Alguns itens demonstraram cargas fatoriais significativamente altas para múltiplos fatores. Nesses casos, o item foi alocado ao fator ao qual possui a maior carga fatorial entre aqueles com carga significativamente alta.

Animação, o único critério pertencente à área de avaliação de jogabilidade no modelo pós-validação de conteúdo, mostrou uma carga fatorial elevada para o segundo fator do modelo. Este mesmo fator, rotulado como diversão, teve como um segundo critério a "curiosidade sensorial". Ambos os critérios que compõem o fator descrevem emoções positivas evocadas por componentes do jogo.

Outra observação interessante quanto a matriz fatorial está no fator de capacidade de aprendizado. O critério homônimo é o único remanescente da área de avaliação de experiência de usuário na matriz fatorial. O aprendizado ao qual o critério se refere é quanto a capacidade de aprender a usar o sistema e não algum conteúdo acadêmico. Portanto, este critério estar em um fator separado dos demais demonstra consistência com o modelo teórico. Este critério é correspondente à usabilidade no modelo desta pesquisa e naqueles que o fundamentam.

## 8.2.4 Comparação entre Grupos Categóricos

Não foi identificado uma diferença estatisticamente relevante para a grande maioria, 86,5%, das combinações de item vs. dado socio-demográfico. Um destaque há de ser dado às combinações nas quais os respondentes tiveram, agregadamente, respostas significativamente diferente de acordo com seu perfil socio-demográfico.

Uma combinação curiosa na qual houve uma diferença significativa entre as categorias foi das respostas à pergunta sobre hábitos de jogar. Mais especificamente,

na combinação entre a pergunta sobre o respondente acharem jogos educacionais divertidos ou não e o item sobre entusiasmo no formulário de avaliação do jogo.

De fato, este dado referente ao quanto os respondentes acham jogos educacionais divertidos mostrou uma variância significativa para cinco dos treze itens do formulário de avaliação de JEDs. Isto é, diferentes opiniões quanto ao material ser divertido exibiram padrões diferentes nas respostas da avaliação do jogo. Esta observação pode mostrar uma tendência a avaliar o jogo de forma mais positiva caso o respondente encare a atividade como divertida, e negativamente caso contrário.

# Capítulo 9

## Conclusões

A contribuição dessa pesquisa é oferecer um artefato capaz de avaliar JEDs de forma mais concisa do que modelos propostos anteriormente. Outra contribuição, além do artefato gerado, é agregar ao conhecimento científico. No ciclo inicial da pesquisa, mediante o desenvolvimento de um modelo de avaliação que agrupe uma quantidade maior e mais diversa do que suas alternativas existentes. Já na sua segunda iteração, foram testadas hipóteses baseadas em fenômenos encontrados na literatura para avaliação de jogos e motivação ao aprendizado, e uma análise exploratória identificou fenômenos latentes a partir dos dados coletados.

O artefato desenvolvido pela pesquisa foi concretizado em um instrumento de medição; foi elaborado um formulário cujos itens correspondem aos critérios do modelo de avaliação de JEDs. O questionário resultante desta dissertação demanda pouco tempo para seu preenchimento, e foca na motivação ao aprendizado por intermédio do jogo.

Outra contribuição relevante originada nesta dissertação é a aplicação de um método de validação de conteúdo utilizando a análise de consenso de lógica fuzzy. A análise de consenso, pertencente a área de tomada de decisão da lógica fuzzy, foi utilizada para determinar a preferência de especialistas quanto a essencialidade de itens para o objetivo da avaliação. Desta forma, a preferência utilizou um método alternativo àquele estabelecido por LAWSHE (1975), mas manteve as opções de grau de essencialidade ao objetivo de medição: desnecessário, “útil, mas não essencial” e essencial.

Em resumo, esta pesquisa desenvolveu um modelo, original e conciso, de avaliação de jogos educacionais digitais. Enquanto a proposta inicial era de uma avaliação abrangente, cobrindo aspectos de jogabilidade, experiência de usuário e de motivação ao aprendizado, as validações feitas durante a pesquisa tornaram-no um modelo focado na avaliação de motivação ao aprendizado mediante jogos educacionais digitais. O formulário vinculado ao modelo demonstrou confiabilidade enquanto instrumento de medição, além de validade de conteúdo e de construtos. A confirmação da confi-

abilidade e validade do modelo reforça sua robustez, assegurando que o formulário mede e identifica corretamente o fenômeno latente.

Em relação à validade de construtos, fica a ressalva à análise fatorial confirmatória, que evidenciou resultados insatisfatórios na adequação do modelo teórico pré-estabelecido aos dados coletados. As causas e implicações destes resultados insatisfatórios foram abordadas no capítulo de Discussão (Capítulo 8.2). Em paralelo, a análise exploratória produziu um resultado interessante, no qual nove itens remanescentes do formulário foram distribuídos em quatro fatores fortemente associados à motivação ao aprendizado.

O artefato nessa pesquisa serve como uma alternativa válida para a avaliação de jogos educacionais digitais. Algo que o diferencia é a sua brevidade, o que pode ter um efeito positivo na taxa de respostas (KITCHENHAM e PFLEEGER, 2002b). Há também uma ênfase mais focada no aspecto motivacional do aprendizado mediante o jogo.

## 9.1 Limitações

O tamanho de amostra do estudo foi de 65 respondentes para a avaliação do formulário. Apesar de não ter sido um empecilho, conforme limites encontrados na literatura para as análises feitas nesta pesquisa, seria desejável ter uma quantidade de respondentes mais elevada. Uma quantidade maior de especialistas também teria impactos positivos à pesquisa, pois um grau de divergência maior seria permitido seguindo o critério de validade de conteúdo tradicional, estabelecido por (LAWSHE, 1975).

Para remediar esta limitação de respondentes em ambas avaliações em trabalhos futuros, seria desenvolvida alguma estratégia para uma taxa de resposta mais elevada. Isto seria feito devido ao formulário de avaliação ter relativamente poucos respondentes proporcional a quantidade de convites à participação.

Outra limitação se dá em relação à diversidade da amostra. Apesar dos respondentes terem perfis sociodemográficos distintos, a grande maioria dos respondentes são alunos da mesma instituição de ensino.

Outra limitação da avaliação feita nesta dissertação é que por utilizar somente um formulário psicométrico, a avaliação limita-se à reação do aluno ao jogo. KIRKPATRICK e CRAIG (1970) define uma hierarquia de quatro níveis de avaliação no qual reação é apenas o primeiro. Portanto, faz-se necessário a introdução de algum outro instrumento de medição para atingir os níveis subsequentes de avaliação.

Sugestões para solucionar estas limitações em trabalhos futuros são levantadas na próxima seção, de trabalhos futuros.

## 9.2 Trabalhos Futuros

O instrumento final desta pesquisa resultou em um escopo mais específico do que o planejado, concentrando-se na avaliação do aspecto motivacional de jogos educacionais. Em uma nova iteração de desenvolvimento do formulário de avaliação, seria feita uma nova revisão bibliográfica para identificar teorias e critérios motivacionais preteridos em sua versão inicial. Além disso, uma gama maior de itens candidatos no escopo específico de aspectos motivacionais aplicados a jogos educacionais seria apresentado aos especialistas, resultando numa escala potencialmente mais robusta e consistente.

O guia proposto por DEVELLIS e THORPE (2021) tem como sua última etapa a otimização do tamanho da escala. Seguindo a orientação deste guia, em um trabalho futuro a tendência é que serão acrescentados itens. A quantidade de itens se provou escassa em diversos pontos desta pesquisa, seguramente o incremento do formulário após um novo ciclo de desenvolvimento será bem-vindo.

O refinamento do conjunto de itens da escala seria feito consoante o processo iterativo descrito em ALMANASREH *et al.* (2019). Este processo seria seguido ponderando a necessidade da validade de conteúdo no novo ciclo de desenvolvimento do formulário.

Nesta segunda iteração, os itens podem ser modificados, adicionados ou removidos para a reconstrução do instrumento de medição desenvolvido, para então passar por uma nova rodada no ciclo de validação de conteúdo. Em uma eventual segunda rodada, seria dado um enfoque na área de motivação ao aprendizado, adicionando variações dos itens que representam os critérios incluídos no modelo e removendo os itens que pertenciam às áreas de jogabilidade e experiência de usuário do modelo.

Além disso, seria estabelecido um limite inferior de 100 respostas e, de preferência, com uma proporção de dez vezes mais respondentes que itens no formulário. Este tamanho amostral seria primordial para aprimorar os resultados obtidos nas análises de confiabilidade e validade. A fim de conseguir mais respostas, uma gama maior de meios de distribuição seriam usados para difundir o uso do formulário para mais pessoas e instituições de ensino.

Uma análise qualitativa usando os métodos de *grounded theory* (GLASER e STRAUSS, 2017) enriqueceria as evidências apresentadas nesta dissertação. Assim, esta análise qualitativa poderia ser conduzida em paralelo a um novo ciclo de validação do modelo e formulário de avaliação. Os estágios iniciais da pesquisa, como a definição de um problema de investigação e do escopo da pesquisa são inerentes também ao *grounded theory* (FERNANDES e MAIA, 2001)

A teoria gerada na pesquisa feita a partir desta dissertação limita-se à construção de um modelo para a explicação de um fenômeno. Esta explicação do fenômeno

baseia-se em critérios encontrados na teoria e na descoberta de fatores formados por conjuntos mutuamente exclusivos de critérios.

O processo iterativo de comparação constante dos dados obtidos e conceitos abstratos que descrevem os dados é inerente ao *grounded theory*. Esta iteratividade possibilita a categorização de conceitos que possivelmente não foram encontrados nas referências na etapa de revisão bibliográfica. Assim posto, a geração de conhecimento com essa adição de análise qualitativa poderia mapear novos critérios para a análise de aspectos motivacionais de jogos educacionais. A análise qualitativa seria feita até que uma *saturação teórica* fosse atingida, na qual entrevistas novas resultassem em refinamentos irrisórios ao modelo conceitual construído durante o *grounded analysis*.

Outro trabalho futuro parte da avaliação de qualidade de software SQuaRe. Em complemento à avaliação da reação do aluno ao jogo, mediante o aspecto de motivação ao aprendizado, medir-se-ia a efetividade e eficiência das tarefas pertencentes ao jogo. Esses atributos são de avaliação de qualidade em uso, e a efetividade e eficiência seriam medidas a partir dos seguintes indicadores, respectivamente: a proporção de tarefas concluídas, e a quantidade de tempo e recursos consumidos no processo.

A avaliação também pode ser estendida para avaliar o efeito do jogo educacional no aprendizado do aluno. Uma maneira seria aplicar avaliações pré-teste e pós-teste sobre o conteúdo didático cujo aprendizado é esperado com o jogo. Nesse cenário, faz-se necessário a presença de um grupo de controle para a medição do efeito na aprendizagem. O grupo de controle participaria de alguma atividade didática, que não o jogo educacional digital, para aprender o conteúdo avaliado nas avaliações pré-teste e pós-teste.

Finalmente, também seria interessante elaborar uma escala Stapel que avaliasse os mesmos critérios do modelo de avaliação elaborado por esta pesquisa. Nesta situação, seriam dispostas afirmações correspondentes aos critérios do modelo e caberia ao respondente avaliar o quanto um determinado adjetivo é compatível com a afirmação. Os dados obtidos usando a escala Stapel poderiam então ser comparados àqueles analisados nesta dissertação.

# Referências Bibliográficas

- ADETUNJI, R. O., ADE-IBIJOLA, A., 2024, “Unlocking Learning: Investigating the Replayability of Educational Games”, *International Journal of Computer Games Technology*, v. 2024, n. 1, pp. 5876780.
- ALBAUM, G., 1997, “The Likert scale revisited”, *Market Research Society. Journal.*, v. 39, n. 2, pp. 1–21.
- ALMANASREH, E., MOLES, R., CHEN, T. F., 2019, “Evaluation of methods used for estimating content validity”, *Research in Social and Administrative Pharmacy*, v. 15, n. 2, pp. 214–221. ISSN: 1551-7411. doi: <https://doi.org/10.1016/j.sapharm.2018.03.066>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1551741118302687>>.
- ARTHUR JR, W., TUBRE, T., PAUL, D. S., et al., 2003, “Teaching Effectiveness: The relationship between reaction and learning evaluation criteria”, *Educational Psychology*, v. 23, n. 3.
- ASLAN, S., BALCI, O., 2015, “GAMED: digital educational game development methodology”, *Simulation*, v. 91, n. 4, pp. 307–319.
- BARTLETT, M., 1954, “A note on multiplying factors for various chi-squared approximations”, *Journal of the Royal Statistical Society*, v. 16, pp. 296–298.
- BARTLETT, M. S., 1951, “The effect of standardization on a  $\chi^2$  approximation in factor analysis”, *Biometrika*, v. 38, n. 3/4, pp. 337–344.
- BAUMEISTER, R. F., LEARY, M. R., 1997, “Writing narrative literature reviews”, *Review of general psychology*, v. 1, n. 3, pp. 311–320.
- BEZDEK, J. C., SPILLMAN, B., SPILLMAN, R., 1978, “A fuzzy relation space for group decision theory”, *Fuzzy Sets and systems*, v. 1, n. 4, pp. 255–268.
- BIGGS, J., MADNANI, N., 2022. “factor\_analyzer.factor\_analyzer Module”. Sep. Disponível em: <<https://factor-analyzer.readthedocs>.



io/en/latest/factor\_analyzer.html#module-factor{ }analyzer.  
factor{ }analyzer>.

- BLOOM, B. S., ENGELHART, M. D., FURST, E. J., et al., 1956, *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, NY, David McKay Company.
- BPMN, 2011. “Notation (BPMN) version 2.0”. .
- BRASIL, 2018. “LEI Nº 13.709, DE 14 DE AGOSTO DE 2018: Lei Geral de Proteção de Dados Pessoais (LGPD).” Disponível em: <[https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/L13709compilado.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm)>.
- BREUER, J., BENTE, G., 2010, “Why so serious? On the relation of serious games and learning”, *Journal for Computer Game Culture*, v. 4, pp. 7–24.
- BRINKERHOFF, R. O., 2005, “The success case method: A strategic evaluation approach to increasing the value and effect of training”, *Advances in Developing Human Resources*, v. 7, n. 1, pp. 86–101.
- BROWN, W., 1910, “Some experimental results in the correlation of mental abilities 1”, *British Journal of Psychology, 1904-1920*, v. 3, n. 3, pp. 296–322.
- BRUNNSTRÖM, K., BEKER, S. A., DE MOOR, K., et al., 2013, “Qualinet white paper on definitions of quality of experience”, *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*.
- BUJANG, M. A., OMAR, E. D., BAHARUM, N. A., 2018, “A review on sample size determination for Cronbach’s alpha test: a simple guide for researchers”, *The Malaysian journal of medical sciences: MJMS*, v. 25, n. 6, pp. 85.
- CAILLOIS, R., 2001, *Man, play, and games*. Urbana and Chicago, University of Illinois press.
- CALDIERA, V. R. B. G., ROMBACH, H. D., 1994, “The goal question metric approach”, *Encyclopedia of software engineering*, pp. 528–532.
- CARLTON, G., 2024. “2024 Online Learning Statistics.” 8.
- CATTELL, R. B., 1966, “The scree test for the number of factors”, *Multivariate behavioral research*, v. 1, n. 2, pp. 245–276.

- CHINOSI, M., TROMBETTA, A., 2012, “BPMN: An introduction to the standard”, *Computer Standards & Interfaces*, v. 34, n. 1, pp. 124–134. ISSN: 0920-5489. doi: <https://doi.org/10.1016/j.csi.2011.06.002>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0920548911000766>>.
- COHEN, J., 2013, *Statistical power analysis for the behavioral sciences*. New York, Routledge.
- COLLINS, L. A., SMITH, A. J., HANNON, P. D., 2006, “Applying a Synergistic Learning Approach in Entrepreneurship Education”, *Management Learning*, v. 37, n. 3, pp. 335–354. doi: 10.1177/1350507606067171. Disponível em: <<https://doi.org/10.1177/1350507606067171>>.
- COLMAN, A. M., NORRIS, C. E., PRESTON, C. C., 1997, “Comparing rating scales of different lengths: Equivalence of scores from 5-point and 7-point scales”, *Psychological Reports*, v. 80, n. 2, pp. 355–362.
- CRESPI, I., 1961, “Use of a scaling technique in surveys”, *Journal of Marketing*, v. 25, n. 5, pp. 69–72.
- CRONBACH, L. J., 1951, “Coefficient alpha and the internal structure of tests”, *psychometrika*, v. 16, n. 3, pp. 297–334.
- CSIKSZENTMIHALYI, M., 2008, *FLOW: The Psychology of Optimal Experience*. 1 ed. New York, Harper Perennial Modern Classics.
- DE OLIVEIRA, S., 2022, “O uso pedagógico do jogo de tabuleiro”. In: *Jogos de Tabuleiro na Educação*, cap. 6, pp. 101–110, São Paulo, SP, Devir.
- DEVELLIS, R. F., THORPE, C. T., 2021, *Scale development: Theory and applications*. Thousand Oaks, CA, Sage publications.
- DILLON, R., 2010, *On the Way to Fun: an emotion-based approach to successful game design*. Boca Raton, FL, CRC Press.
- DING, L., VELICER, W. F., HARLOW, L. L., 1995, “Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices”, *Structural Equation Modeling: A Multidisciplinary Journal*, v. 2, n. 2, pp. 119–143.
- DINNO, A., 2009, “Implementing Horn’s parallel analysis for principal component analysis and factor analysis”, *The Stata Journal*, v. 9, n. 2, pp. 291–298.

- DIXON, W. J., 1992, *BMDP statistical software manual: to accompany BMDP release 7*, v. 1. Berkeley, CA, Univ of California Press.
- DUNN, W. N., 2015, *Public policy analysis*. New York, Routledge.
- EMMERICH, K., BOCKHOLT, M., 2016, “Serious games evaluation: processes, models, and concepts”. In: *Entertainment Computing and Serious Games*, Springer, pp. 265–283.
- FABRIGAR, L. R., WEGENER, D. T., MACCALLUM, R. C., et al., 1999, “Evaluating the use of exploratory factor analysis in psychological research.” *Psychological methods*, v. 4, n. 3, pp. 272.
- FERNANDES, E. M., MAIA, Â., 2001, “Grounded theory”. In: *Métodos e técnicas de avaliação : contributos para a prática e investigação psicológicas*, pp. 49–76, Braga, Universidade do Minho. Centro de Estudos em Educação e Psicologia (CEEP).
- FISHER, R. A., 1922, “On the mathematical foundations of theoretical statistics”, *Philosophical transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, v. 222, n. 594-604, pp. 309–368.
- FISHER, R. A., 1925, *Statistical methods for research workers*. 1 ed. Edimburgo, Escócia, Oliver and Boyd.
- FONTOURA JUNIOR, P. H. F., 2018, *Recomendações para o desenvolvimento de jogos educacionais: aspectos para a inclusão de pessoas com deficiência visual*. Dissertação de mestrado, Universidade Federal de São Carlos.
- FORNELL, C., 1983, “Issues in the application of covariance structure analysis: A comment”, *Journal of consumer research*, v. 9, n. 4, pp. 443–448.
- FU, F.-L., SU, R.-C., YU, S.-C., 2009, “EGameFlow: A scale to measure learners’ enjoyment of e-learning games”, *Computers & Education*, v. 52, n. 1, pp. 101–112.
- GARRIS, R., AHLERS, R., DRISKELL, J. E., 2002, “Games, motivation, and learning: A research and practice model”, *Simulation & gaming*, v. 33, n. 4, pp. 441–467.
- GEE, J. P., 2003, “What video games have to teach us about learning and literacy”, *Computers in entertainment (CIE)*, v. 1, n. 1.

- GERBING, D. W., HAMILTON, J. G., 1996, “Viability of exploratory factor analysis as a precursor to confirmatory factor analysis”, *Structural Equation Modeling: A Multidisciplinary Journal*, v. 3, n. 1, pp. 62–72.
- GLASER, B., STRAUSS, A., 2017, *Discovery of grounded theory: Strategies for qualitative research*. Routledge.
- GOMES, A. S., GOMES, C. R. A., 2020, “Estrutura do Método científico: Por uma epistemologia da Informática na Educação”, *Metodologia de Pesquisa Científica em Informática na Educação: Concepção de Pesquisa*. Porto Alegre: SBC. Recuperado em, v. 6.
- GRANT, J. S., DAVIS, L. L., 1997, “Selection and use of content experts for instrument development”, *Research in nursing & health*, v. 20, n. 3, pp. 269–274.
- GREGOR, S., HEVNER, A. R., 2013, “POSITIONING AND PRESENTING DESIGN SCIENCE RESEARCH FOR MAXIMUM IMPACT”, *MIS Quarterly*, v. 37, n. 2, pp. 337–355.
- GUTTMAN, L., 1944, “A basis for scaling qualitative data”, *American sociological review*, v. 9, n. 2, pp. 139–150.
- GUTTMAN, L., 1954, “Some necessary conditions for common-factor analysis”, *Psychometrika*, v. 19, n. 2, pp. 149–161.
- HAIR, J. F., BLACK, W. C., BABIN, B. J., et al., 2009, *Análise multivariada de dados*. Porto Alegre, RS, Bookman editora.
- HAYNES, S. N., RICHARD, D., KUBANY, E. S., 1995, “Content validity in psychological assessment: A functional approach to concepts and methods.” *Psychological assessment*, v. 7, n. 3, pp. 238.
- HAYTON, J. C., ALLEN, D. G., SCARPELLO, V., 2004, “Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis”, *Organizational research methods*, v. 7, n. 2, pp. 191–205.
- HEVNER, A. R., MARCH, S. T., PARK, J., et al., 2004, “Design science in information systems research”, *MIS quarterly*, pp. 75–105.
- HOEFT, F., WATSON, C. L., KESLER, S. R., et al., 2008, “Gender differences in the mesocorticolimbic system during computer game-play”, *Journal of psychiatric research*, v. 42, n. 4, pp. 253–258.
- HOLTON III, E. F., 1996, “The flawed four-level evaluation model”, *Human resource development quarterly*, v. 7, n. 1, pp. 5–21.

- HORN, J. L., 1965, “A rationale and test for the number of factors in factor analysis”, *Psychometrika*, v. 30, pp. 179–185.
- HOSS, M., TEN CATEN, C. S., 2010, “Processo de Validação Interna de um Questionário em uma Survey Research sobre ISO 9001: 2000”, *Produto & Produção*, v. 11, n. 2.
- HSU, S. H., WEN, M.-H., WU, M.-C., 2009, “Exploring user experiences as predictors of MMORPG addiction”, *Computers & Education*, v. 53, n. 3, pp. 990–999.
- HUIZINGA, J., 2019, *Homo ludens: O jogo como elemento da cultura*. São Paulo, SP, Perspectiva.
- HUNICKE, R., LEBLANC, M., ZUBEK, R., 2004, “MDA: A formal approach to game design and game research”. In: *Proceedings of the AAAI Workshop on Challenges in Game AI*, v. 4, p. 1722, San Jose, CA, 7.
- IGOLKINA, A. A., MESHCHERYAKOV, G., 2020, “semopy: A Python Package for Structural Equation Modeling”, *Structural Equation Modeling: A Multidisciplinary Journal*, v. 0, n. 0, pp. 1–12. doi: 10.1080/10705511.2019.1704289. Disponível em: <<https://doi.org/10.1080/10705511.2019.1704289>>.
- ISO/IEC 25000, 2020. “Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Guide to SQuaRE”. 06. Disponível em: <<https://www.iso.org/standard/64764.html>>.
- ISO/IEC 9126, 1991. “Information technology - Software evaluation - Quality characteristics and guide-lines for their use”. 12.
- KAISER, H. F., 1958, “The varimax criterion for analytic rotation in factor analysis”, *Psychometrika*, v. 23, n. 3, pp. 187–200.
- KAISER, H. F., 1960, “The application of electronic computers to factor analysis”, *Educational and psychological measurement*, v. 20, n. 1, pp. 141–151.
- KAISER, H. F., RICE, J., 1974, “Little jiffy, mark IV”, *Educational and psychological measurement*, v. 34, n. 1, pp. 111–117.
- KASURINEN, J., 2016, “Games as software: Similarities and differences between the implementation projects”. In: *Proceedings of the 17th International Conference on Computer Systems and Technologies 2016*, pp. 33–40, 06.

- KELLER, J. M., 1987, "Development and use of the ARCS model of instructional design", *Journal of instructional development*, v. 10, n. 3, pp. 2–10.
- KIRKPATRICK, D., KIRKPATRICK, J., 2006, *Evaluating training programs: The four levels*. San Francisco, CA, Berrett-Koehler Publishers.
- KIRKPATRICK, D. L., CRAIG, R., 1970, "Evaluation of training", *Evaluation of short-term training in rehabilitation*, p. 35.
- KITCHENHAM, B. A., PFLEEGER, S. L., 2002a, "Principles of survey research part 2: designing a survey", *ACM SIGSOFT Software Engineering Notes*, v. 27, n. 1, pp. 18–20.
- KITCHENHAM, B. A., PFLEEGER, S. L., 2002b, "Principles of survey research: part 3: constructing a survey instrument", *ACM SIGSOFT Software Engineering Notes*, v. 27, n. 2, pp. 20–24.
- KORHONEN, H., PAAVILAINEN, J., SAARENPÄÄ, H., 2009, "Expert review method in game evaluations: comparison of two playability heuristic sets". In: *Proceedings of the 13th international MindTrek conference: Everyday life in the ubiquitous era*, pp. 74–81, 11.
- KOSTER, R., 2013, *Theory of fun for game design*. "O'Reilly Media, Inc."
- KRATHWOHL, D. R., 2002, "A Revision of Bloom's Taxonomy: An Overview", *THEORY INTO PRACTICE*, v. 41, n. 4.
- KRUSKAL, W. H., WALLIS, W. A., 1952, "Use of ranks in one-criterion variance analysis", *Journal of the American statistical Association*, v. 47, n. 260, pp. 583–621.
- KUDER, G. F., RICHARDSON, M. W., 1937, "The theory of the estimation of test reliability", *Psychometrika*, v. 2, n. 3, pp. 151–160.
- LAWSHE, C. H., 1975, "A quantitative approach to content validity", *Personnel psychology*, v. 28, n. 4, pp. 563–575.
- LEE, L.-C., HAO, K.-C., 2015, "Designing and evaluating digital game-based learning with the ARCS motivation model, humor, and animation", *International Journal of Technology and Human Interaction (IJTHI)*, v. 11, n. 2, pp. 80–95.
- LEFEVER, S., DAL, M., MATTHÍASDÓTTIR, Á., 2007, "Online data collection in academic research: advantages and limitations", *British journal of educational technology*, v. 38, n. 4, pp. 574–582.

- LIKERT, R., 1932, “A technique for the measurement of attitudes.” *Archives of psychology*.
- LOMBARD, M., DITTON, T., 1997, “At the heart of it all: The concept of presence”, *Journal of computer-mediated communication*, v. 3, n. 2, pp. JCMC321.
- LÜDECKE, D., BEN-SHACHAR, M. S., PATIL, I., et al., 2021, “performance: An R Package for Assessment, Comparison and Testing of Statistical Models”, *Journal of Open Source Software*, v. 6, n. 60, pp. 3139. doi: 10.21105/joss.03139.
- MACCALLUM, R. C., WIDAMAN, K. F., ZHANG, S., et al., 1999, “Sample size in factor analysis.” *Psychological methods*, v. 4, n. 1, pp. 84.
- MALONE, T. W., 1981, “Toward a theory of intrinsically motivating instruction”, *Cognitive science*, v. 5, n. 4, pp. 333–369.
- MATSUNAGA, M., 2010, “How to Factor-Analyze Your Data Right: Do’s, Don’ts, and How-To’s.” *International journal of psychological research*, v. 3, n. 1, pp. 97–110.
- MAYER, R. E., 1997, “Multimedia learning: Are we asking the right questions?” *Educational psychologist*, v. 32, n. 1, pp. 1–19.
- MILLER, G. A., 1956, “The magical number seven, plus or minus two: Some limits on our capacity for processing information.” *Psychological review*, v. 63, n. 2, pp. 81.
- NASCIMENTO, P. V. M., 2023. “Desenvolvimento de um jogo digital educacional para o ensino de pensamento computacional concorrente”. .
- NEMOTO, T., BEGLAR, D., 2013, “Likert-scale questionnaires”. In: *JALT 2013 conference proceedings*, v. 108, pp. 1–6, 10.
- NOVAK, T. P., HOFFMAN, D. L., YUNG, Y.-F., 2000, “Measuring the customer experience in online environments: A structural modeling approach”, *Marketing science*, v. 19, n. 1, pp. 22–42.
- NUNNALLY, J. C., 1978, “An overview of psychological measurement”, *Clinical diagnosis of mental disorders: A handbook*, pp. 97–146.
- OBIKWELU, C. O., 2017, *Evaluating scaffolding in serious games with children*. Tese de Doutorado, University of Central Lancashire.

- OF VIRGINIA LIBRARY, U., Year the webpage was last updated or accessed. “Getting Started with the Kruskal-Wallis Test”. Disponível em: <<https://library.virginia.edu/data/articles/getting-started-with-the-kruskal-wallis-test>>.
- OLIVEIRA, R. N., ROCHA, R. V., 2021, “AvaliaJS: Planejamento da Avaliação do Desempenho de Alunos em Jogos Sérios”. In: *Anais dos Workshops do X Congresso Brasileiro de Informática na Educação*, pp. 101–110, Porto Alegre, RS, 11. SBC.
- ORLICK, T. D., 1981, “Positive socialization via cooperative games.” *Developmental Psychology*, v. 17, n. 4, pp. 426.
- PANDAS DEVELOPMENT TEAM, T., 2020. “pandas-dev/pandas: Pandas”. fev. Disponível em: <<https://doi.org/10.5281/zenodo.3509134>>.
- PASQUALOTTO, A., PARONG, J., GREEN, C. S., et al., 2022, “Video Game Design for Learning to Learn”, *International Journal of Human-Computer Interaction*, pp. 1–18.
- PEARSON, K., 1900, “X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling”, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, v. 50, n. 302, pp. 157–175.
- PEFFERS, K., TUUNANEN, T., ROTHENBERGER, M. A., et al., 2007, “A design science research methodology for information systems research”, *Journal of management information systems*, v. 24, n. 3, pp. 45–77.
- PETRI, G., VON WANGENHEIM, C. G., BORGATTO, A. F., 2019, “MEEGA+: Um Modelo para a Avaliação de Jogos Educacionais para o ensino de Computação”, *Revista Brasileira de Informática na Educação*, v. 27, n. 03, pp. 52–81.
- PRASLOVA, L., 2010, “Adaptation of Kirkpatrick’s four level model of training criteria to assessment of learning outcomes and program evaluation in higher education”, *Educational assessment, evaluation and accountability*, v. 22, pp. 215–225.
- REIS, A. P. D. S., 2020, *Recomendações de design para jogos educacionais Inclusivos para diferentes dispositivos*. Tese de Mestrado, Universidade Federal de São Carlos.



- REIS, H. B. D. C., 2023. “Jogos de timeline: elaboração e desenvolvimento de um jogo educativo para ensino do processo de descoberta e desenvolvimento de fármacos”. .
- ROBINSON, J. P., SHAVER, P. R., WRIGHTSMAN, L. S., 1991, “Criteria for scale selection and evaluation”, *Measures of personality and social psychological attitudes*, v. 1, n. 3.
- ROSS, T. J., 2009, *Fuzzy logic with engineering applications*. Chichester, West Sussex PO19 8SQ, England, John Wiley & Sons.
- SAARILUOMA, P., JOKINEN, J. P., 2014, “Emotional dimensions of user experience: A user psychological analysis.” *International Journal of Human-Computer Interaction*.
- SÁNCHEZ, J. L. G., VELA, F. L. G., SIMARRO, F. M., et al., 2012, “Playability: analysing user experience in video games”, *Behaviour & Information Technology*, v. 31, n. 10, pp. 1033–1054.
- SAVI, R., ULBRICHT, V. R., 2008, “Jogos digitais educacionais: benefícios e desafios”, *Renote*, v. 6, n. 1.
- SAVI, R., VON WANGENHEIM, C. G., BORGATTO, A. F., 2011, “A model for the evaluation of educational games for teaching software engineering”. In: *2011 25Th brazilian symposium on software engineering*, pp. 194–203. IEEE, 11.
- SCHALOCK, R. L., 2001, *Outcome-based evaluation*. New York, Springer Science & Business Media.
- SEABOLD, S., PERKTOLD, J., 2010, “statsmodels: Econometric and statistical modeling with python”. In: *9th Python in Science Conference*, 6.
- SOLINGEN, R., BERGHOUT, E., 1999, *The Goal/Question/Metric Method: A Practical Guide for Quality Improvement of Software Development*. McGraw-Hill.
- SOUSA, D. L. D., MAIA, J. R. C., SILVA, T. D. O., 2023. “Cred Carbon Game: uma abordagem interativa sobre questões ambientais”. .
- SPEARMAN, C., 1910, “Correlation calculated from faulty data”, *British journal of psychology*, v. 3, n. 3, pp. 271.

- STEINMAYR, R., SPINATH, B., 2009, “The importance of motivation as a predictor of school achievement”, *Learning and individual differences*, v. 19, n. 1, pp. 80–90.
- STRAUB, D., BOUDREAU, M.-C., GEFEN, D., 2004, “Validation guidelines for IS positivist research”, *Communications of the Association for Information systems*, v. 13, n. 1, pp. 24.
- STROTHER, J. B., 2002, “An assessment of the effectiveness of e-learning in corporate training programs”, *The International Review of Research in Open and Distributed Learning*, v. 3, n. 1.
- STUDENT, 1908, “The probable error of a mean”, *Biometrika*, pp. 1–25.
- SUHR, D. D., 2006, “Exploratory or confirmatory factor analysis?” *Statistics and Data Analysis*.
- SWEETSER, P., WYETH, P., 2005, “GameFlow: a model for evaluating player enjoyment in games”, *Computers in Entertainment (CIE)*, v. 3, n. 3, pp. 3–3.
- TAHERDOOST, H., 2019, “What is the best response scale for survey and questionnaire design; review of different lengths of rating scale/attitude scale/Likert scale”, *Hamed Taherdoost*, pp. 1–10.
- TAKATALO, J., HÄKKINEN, J., KAISTINEN, J., et al., 2010, “Presence, involvement, and flow in digital games”. In: *Evaluating user experience in games*, Springer, pp. 23–46.
- TEN BERGE, J. M., ZEGERS, F. E., 1978, “A series of lower bounds to the reliability of a test”, *Psychometrika*, v. 43, pp. 575–579.
- THURSTONE, L. L., 1947, *Multiple-factor analysis; a development and expansion of The Vectors of Mind*. Chicago, IL, University of Chicago Press.
- TOKARIEVA, A. V., VOLKOVA, N. P., HARKUSHA, I. V., 2019, “Educational digital games: models and implementation”, *Educational Dimension*.
- UNESCO, 2022. “Education: From disruption to recovery”. 6. Disponível em: <<https://en.unesco.org/covid19/educationresponse>>.
- VALLAT, R., 2022. “pinguin.cronbach\_alpha”. Disponível em: <<https://pinguin-stats.org/build/html/generated/pinguin.cronbach%5Falpha.html>>.

- VALLAT, R., 2018, “Pingouin: statistics in Python.” *J. Open Source Softw.*, v. 3, n. 31, pp. 1026.
- VAZ, A., XEXÉO, G., 2022, “Desenvolvimento de um Modelo de Reação a Jogos Educacionais Digitais”. In: *Anais Estendidos do XXI Simpósio Brasileiro de Jogos e Entretenimento Digital*, pp. 623–632. SBC.
- VITIELLO, P., 2022, “Adesão e resistência na experiência do jogar educativo”. In: *Jogos de Tabuleiro na Educação*, cap. 1, pp. 51–53, São Paulo, SP, Devir.
- VYGOTSKY, L. S., 1978, “Interaction between Learning and Development”. In: *Mind in Society: Development of Higher Psychological Processes*, pp. 79–91, Cambridge, MA, Harvard University Press. ISBN: 9780674576285. Disponível em: <<http://www.jstor.org/stable/j.ctvjf9vz4.11>>.
- WAKITA, T., UESHIMA, N., NOGUCHI, H., 2012, “Psychological distance between categories in the Likert scale: Comparing different numbers of options”, *Educational and psychological measurement*, v. 72, n. 4, pp. 533–546.
- WES MCKINNEY, 2010, “Data Structures for Statistical Computing in Python”. In: Stéfan van der Walt, Jarrod Millman (Eds.), *Proceedings of the 9th Python in Science Conference*, pp. 56 – 61, 6. doi: 10.25080/Majora-92bf1922-00a.
- WILSON, F. R., PAN, W., SCHUMSKY, D. A., 2012, “Recalculation of the critical values for Lawshe’s content validity ratio”, *Measurement and evaluation in counseling and development*, v. 45, n. 3, pp. 197–210.
- WINN, B. M., 2009, “The design, play, and experience framework”. In: *Handbook of research on effective electronic gaming in education*, IGI Global, pp. 1010–1024, Hershey, PA.
- WITECK, G. R., ALVES, A. C., BERNARDO, M. H., 2021, “Bloom Taxonomy, Serious Games and Lean Learning: What Do These Topics Have in Common?” In: *Learning in the Digital Era: 7th European Lean Educator Conference, ELEC 2021, Trondheim, Norway, October 25–27, 2021, Proceedings 7*, pp. 308–316. Springer, 10.
- WOOD, D., BRUNER, J. S., ROSS, G., 1976, “The role of tutoring in problem solving”, *Journal of child psychology and psychiatry*, v. 17, n. 2, pp. 89–100.

XEXÉO, G., MANGELI, E., NASCIMENTO, P., et al., 2022, “Um Jogo Digital para Divulgar o Processo de Descoberta e Desenvolvimento de Fármacos”. In: *Simpósio Brasileiro de Jogos e Entretenimento Digital (SBGames)*, pp. 273–277. SBC, 10.

ZADEH, L. A., 1988, “Fuzzy logic”, *Computer*, v. 21, n. 4, pp. 83–93.

# Apêndice A

## Modelo de Avaliação de Jogos Educativos Digitais

### A.1 Versão Inicial

Figura A.1: Modelo de Reação a Jogos Educacionais Digitais

<b>Objetivo de Medição:</b> Reação do Aluno ao Jogo Educacional Digital	
<b>Jogabilidade</b>	<b>Diversão:</b> Fiero, Animação, Relaxamento, Ausência de Medo
	<b>Comunicação:</b> Interações Competitivas, Interações Cooperativas, Compartilhamento, Contexto Social
	<b>Interações com o Sistema:</b> Velocidade, Imprevisibilidade
	<b>Escolha:</b> Variedade nas Ações do Agente, Variedade de Estratégias para Ganhar o Jogo
	<b>Presença:</b> Presença Física, Engajamento com o Personagem, Co-presença
	<b>Fluxo:</b> Competência, Sensação de Distorção de Tempo, Engajamento com o Mundo do Jogo
<b>Experiência de Usuário (UX)</b>	<b>Usabilidade:</b> Inteligibilidade, Aprendizibilidade, Atratividade
	<b>Acessibilidade:</b> Acessibilidade para Pessoas com Daltonismo, Nitidez da Imagem, Textos Bem Formataados, Acessibilidade para Pessoas com Deficiência Auditiva, Volume Ajustável, Linguagem Compreensível
	<b>Proteção de Erros:</b> Proteção de Erros "Leves", Ausência de Erros Disruptivos
	<b>Atenção:</b> Concretude, Humor
<b>Motivação do Aluno</b>	<b>Relevância:</b> Experiência, Temática Relevante, Mecânicas Relevantes
	<b>Confiança:</b> Requisitos de Aprendizizado, Moderação de Expectativas, Autoconfiança
	<b>Satisfação:</b> Recompensas para Tarefas Interessantes, Recompensas para Tarefas Entediantes, Elogios a Resultados Positivos
	<b>Curiosidade:</b> Sensorial, Cognitiva
	<b>Desafio:</b> Objetivo, Demanda de Habilidade, Dificuldade Crescente, Ritmo do Jogo
	<b>Fantasia:</b> Intrínseca, Extrínseca

## A.2 Área de Avaliação 1: Jogabilidade

Tabela A.1: Composição do Critério de Diversão

Sub-Critérios	Descrição
<b>Fiero</b>	É a emoção de orgulho e satisfação ao conquistar uma vitória. Somam-se a essa definição as emoções de <i>joy</i> e <i>happiness</i> encontradas em diversas fontes.
<b>Animação</b>	Sensação de excitação ou entusiasmo, alcançado quando a pessoa sente-se cheia de vida e de energia.
<b>Relaxamento</b>	Estado emocional na qual a pessoa tem uma ausência das sensações de tensão e ansiedade.
<b>Ausência de Medo</b>	É necessário que o jogador saiba que está jogando em um ambiente livre de risco e portanto não sinta pavor ao ponto de desmotivar o seu aprendizado e atrapalhar a sua experiência com o jogo.

Tabela A.2: Composição do Critério de Comunicação

Sub-Critérios	Descrição
<b>Interações Competitivas</b>	Esse critério medirá a reação do aluno aos momentos de disputa com outros agentes do jogo.
<b>Interações Cooperativas</b>	Esse critério medirá a reação do aluno aos momentos no qual houve a <b>cooperação</b> ou <b>colaboração</b> junto a outros jogadores. A cooperação é quando outras pessoas contribuem para a conquista de algum objetivo pessoal de um jogador, enquanto a colaboração é o esforço em conjunto de uma equipe para a conquista de algum objetivo em comum.
<b>Compartilhamento</b>	Jogos colaborativos frequentemente contam com o compartilhamento de recursos entre jogadores e sempre com a troca de informações. Esse critério medirá a reação dos jogadores a maneira com que os recursos e informações foram compartilhados para superar obstáculos.
<b>Contexto Social</b>	Além das sessões do jogo, na qual o jogador está diretamente envolvido com o sistema, a experiência de jogos é complementada por discussões, eventos e outras circunstâncias que remetem ao jogo. Esse critério mede a reação do aluno a tais interações com outros jogadores abordando temas e conteúdos do jogo.

Tabela A.3: Composição do Critério de Interações com o Sistema

<b>Sub-Critérios</b>	<b>Descrição</b>
<b>Velocidade das Interações com o Sistema</b>	O jogo deve responder de forma rápida e ágil às ações feitas pelo jogador, procedendo de forma fluída.
<b>Imprevisibilidade das Interações com o Sistema</b>	É desejável que o jogo responda às ações do jogador com um certo grau de variabilidade ou aleatoriedade, de forma que o jogador não saiba prever com insuspeição qual é a melhor forma de agir naquele cenário.

Tabela A.4: Composição do Critério de Escolha

<b>Sub-Critérios</b>	<b>Descrição</b>
<b>Variedade nas Ações do Agente</b>	O jogador deve ter um amplo repertório de opções a cada interação com o sistema.
<b>Variedade de Estratégias para Ganhar o Jogo</b>	Devem haver múltiplas táticas que tenham chance de vitória, de forma que não seja possível descobrir uma sequência de ações predeterminadas que sempre lhe garanta o sucesso.

Tabela A.5: Composição do Critério de Fluxo

<b>Sub-Critérios</b>	<b>Descrição</b>
<b>Competência</b>	Mede os efeitos que o jogador sente em relação às suas habilidades se o jogo tem objetivos claros.
<b>Sensação de Distorção de Tempo</b>	Engajamento com o jogo ao ponto de sentir que o tempo está passando de forma mais rápida ou mais devagar.
<b>Engajamento com o Mundo do Jogo</b>	O jogador sente-se mais envolvido com o ambiente do jogo do que com a sua vida real enquanto joga.



Tabela A.6: Composição do Critério de Presença

Sub-Critérios	Descrição
<b>Presença Física</b>	Sensação de estar em um ambiente realístico e vivido.
<b>Engajamento com o Personagem</b>	É o quanto as mecânicas e a narrativa do jogo cativam o jogador ao seu papel. A identificação com o papel do personagem dado ao jogador em jogos colaborativos e a sua contribuição para uma eventual vitória também são fatores importantes para o engajamento com o personagem.
<b>Co-presença</b>	Sensação de estar compartilhando a experiência do jogo com outros agentes. Em jogos cooperativos, é também a noção de estar em um grupo com objetivos compartilhados e agir em interesse do grupo de acordo com as dinâmicas e mecânicas do jogo.

### A.3 Área de Avaliação 2: Experiência de Usuário

Tabela A.7: Composição do Critério de Usabilidade

Sub-Critérios	Descrição
<b>Inteligibilidade</b>	Utilidade do sistema comparado ao objetivo desejado pelo usuário; no contexto de jogos educacionais é a capacidade do jogo em auxiliar na conquista dos objetivos de aprendizado do aluno. A utilidade pode ser tanto para objetivos atuais quanto para o futuro.
<b>Aprendizibilidade</b>	Capacidade do sistema de ensinar o usuário a aprender como usá-lo.
<b>Atratividade</b>	O sistema é considerado atrativo se possuir uma interface gráfica e recursos audiovisuais que encantem os jogadores. Alguns fatores que podem contribuir para a atratividade do sistema são: o esquema de cores, o posicionamento dos elementos na tela, os efeitos sonoros, a trilha sonora, entre outros.

Tabela A.8: Composição do Critério de Acessibilidade

Sub-Critérios	Descrição
<b>Acessibilidade para Pessoas com Daltonismo</b>	Nenhuma informação fundamental ao andamento do jogo deve ser passada apenas por cor.
<b>Nitidez da Imagem</b>	As imagens do jogo devem ser nítidas e adaptáveis ao tamanho de tela do usuário.
<b>Textos Bem Formatados</b>	O texto do jogo deve ter uma formatação padronizada e com alto contraste com o plano de fundo, facilitando o seu reconhecimento.
<b>Acessibilidade para Pessoas com Deficiência Auditiva</b>	Nenhuma informação deve ser transmitida apenas por recursos sonoros.
<b>Volume Ajustável</b>	Deve ser possível regular e silenciar o volume do jogo, diferenciando as configurações do volume para efeitos sonoros, trilha sonora e narração do jogo.
<b>Linguagem Compreensível</b>	O jogo deve empregar vocabulário compatível com o público alvo do jogo, considerando a idade e familiaridade dos alunos com o conteúdo sendo ensinado.

Tabela A.9: Composição do Critério de Proteção de Erros

Sub-Critérios	Descrição
<b>Proteção de Erros "Leves"</b>	O jogador não deveria ser exposto a falhas que atrapalhem a sua experiência de aprendizagem ou de jogabilidade.
<b>Ausência de Erros Disruptivos</b>	Não podem haver falhas que mudem o fluxo ou provoquem uma interrupção no jogo

## A.4 Área de Avaliação 3: Motivação ao Aprendizado

Tabela A.10: Composição do Critério de Atenção

Sub-Critérios	Descrição
<b>Concretude</b>	Uso de exemplos e representações visuais para clarificar a importância dos conceitos ou princípios sendo ensinados.
<b>Humor</b>	Uso de analogias, jogo de palavras ou outras técnicas para deixar o jogo com um clima mais descontraído e com momentos engraçados.

Tabela A.11: Composição do Critério de Relevância

Sub-Critérios	Descrição
<b>Experiência</b>	O conhecimento prévio e os eventos passados do aluno devem ser relacionados com a instrução que ele receberá, de forma clara ou com o uso de analogias.
<b>Temática Relevante</b>	O assunto e a narrativa do jogo são compatíveis e relevantes ao objetivo de aprendizagem do aluno.
<b>Mecânicas Relevantes</b>	As ações oferecidas ao jogador e as dinâmicas permitidas e proporcionadas por tais ações são compatíveis e importantes ao objetivo de aprendizagem do aluno.

Tabela A.12: Composição do Critério de Confiança

Sub-Critérios	Descrição
<b>Requisitos de Aprendizado</b>	O material instrucional precisa ter objetivos educacionais claros e explicar os critérios de avaliação para o aluno quando possível.
<b>Moderação de Expectativas</b>	Entendimento do aluno de que mediante o seu esforço ele conseguirá conquistar os objetivos do jogo e ajudá-los a traçar um plano de forma realística.
<b>Autoconfiança</b>	Deixar que os alunos ganhem independência em ambientes de baixo risco para aplicarem o conhecimento em situações realísticas.

Tabela A.13: Composição do Critério de Satisfação

Sub-Critérios	Descrição
<b>Recompensas para Tarefas Interessantes</b>	As atividades consideradas mais significativas devem ser retribuídas de forma imprevisível para o jogador.
<b>Recompensas para Tarefas Entediadas</b>	As atividades chatas e rotineiras devem ser recompensadas extrinsecamente e de forma previsível.
<b>Elogios a Resultados Positivos</b>	Verbalmente enaltecer o progresso e os feitos do aluno e dar um retorno que ajude o aluno quando útil.
<b>Evitar Influências Negativas</b>	Poupar o aluno de qualquer forma de ameaça ou pressão descabida, pois tais atitudes interferem negativamente no seu desempenho.
<b>Intervalos de Reforço</b>	Dar reforços positivos com frequência no começo do aprendizado e torná-los intermitentes de acordo com a competência do aluno na tarefa. Variar tanto no período entre reforços quanto na sua magnitude.
<b>Conforto</b>	A comodidade física que a pessoa sentiu ao jogar.

Tabela A.14: Composição do Critério de Curiosidade

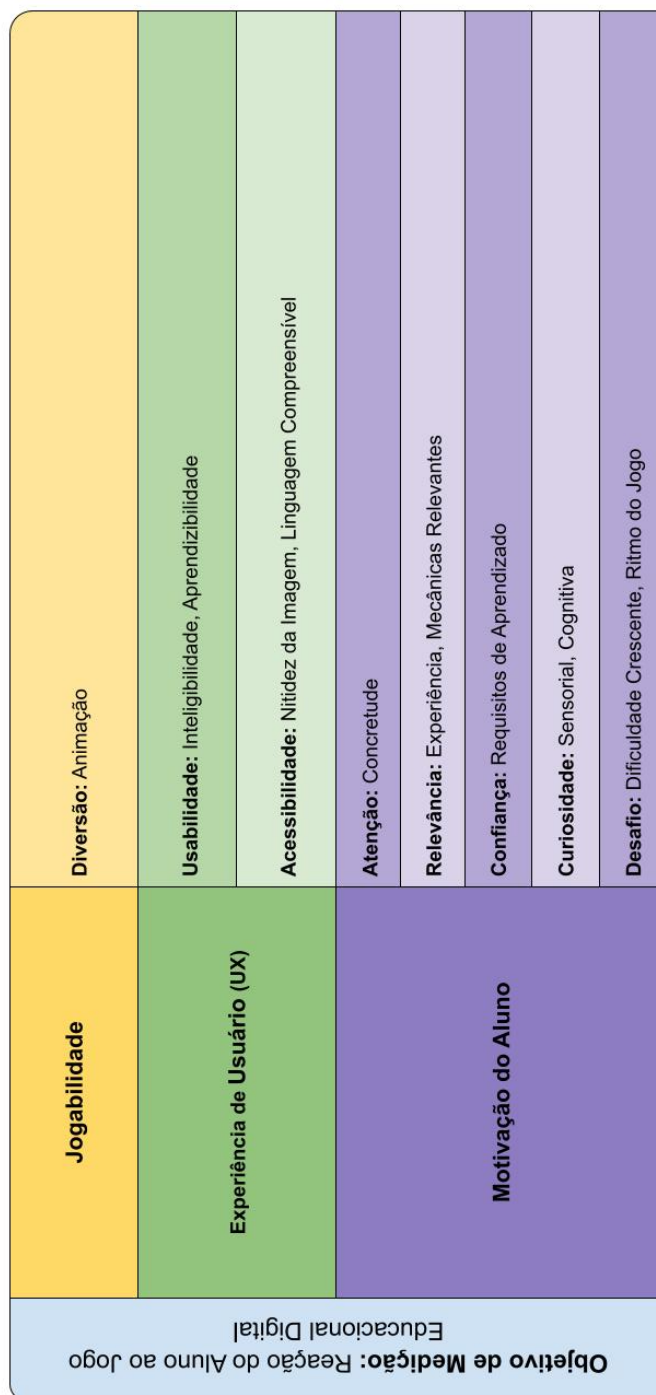
Sub-Critérios	Descrição
<b>Curiosidade Sensorial</b>	É o desejo de explorar, descobrir algo novo, provocado pelo aspecto audiovisual do jogo. Os componentes audiovisuais do jogo podem ser usados para cumprir um papel decorativo, funcionar como um sistema representativo de informação, contribuir para a temática do jogo ou servir como recompensa.
<b>Curiosidade Cognitiva</b>	É o desejo de explorar, descobrir algo novo, provocado pela narrativa e pelas mecânicas. É a vontade do jogador de saciar-se com mais informações sobre algum conceito para suprir um conhecimento que demonstrou-se incompleto ou incongruente.

Tabela A.15: Composição do Critério de Desafio

Sub-Critérios	Descrição
<b>Objetivo</b>	É fundamental que um jogo tenha uma meta clara a ser cumprida; em adição à meta principal sub-objetivos ou o incentivo à otimização de desempenho podem servir como metas secundárias. Uma melhora na pontuação do jogo ou o cumprimento das tarefas de forma mais rápida são aspectos que o jogo pode incentivar seus jogadores a aprimorar.
<b>Demanda de Habilidade</b>	O jogo deve exigir o uso e aprimoramento de alguma habilidade por parte do jogador.
<b>Dificuldade Crescente</b>	O jogo deve começar fácil para dar domínio dos controles e de como jogar o jogo, mas progressivamente aumentar a complexidade exigida para a resolução dos obstáculos impostos pelo jogo.
<b>Ritmo do Jogo</b>	O jogo deve introduzir conceitos e conteúdos em uma cadência moderado, encontrando um equilíbrio entre um andamento que leve a um processo de aprendizado que exija muito esforço e um que cause o tédio.

## A.5 Versão Final

Figura A.2: Modelo de Reação a Jogos Educacionais Digitais, exibindo apenas os critérios que permaneceram após a validação de conteúdo.



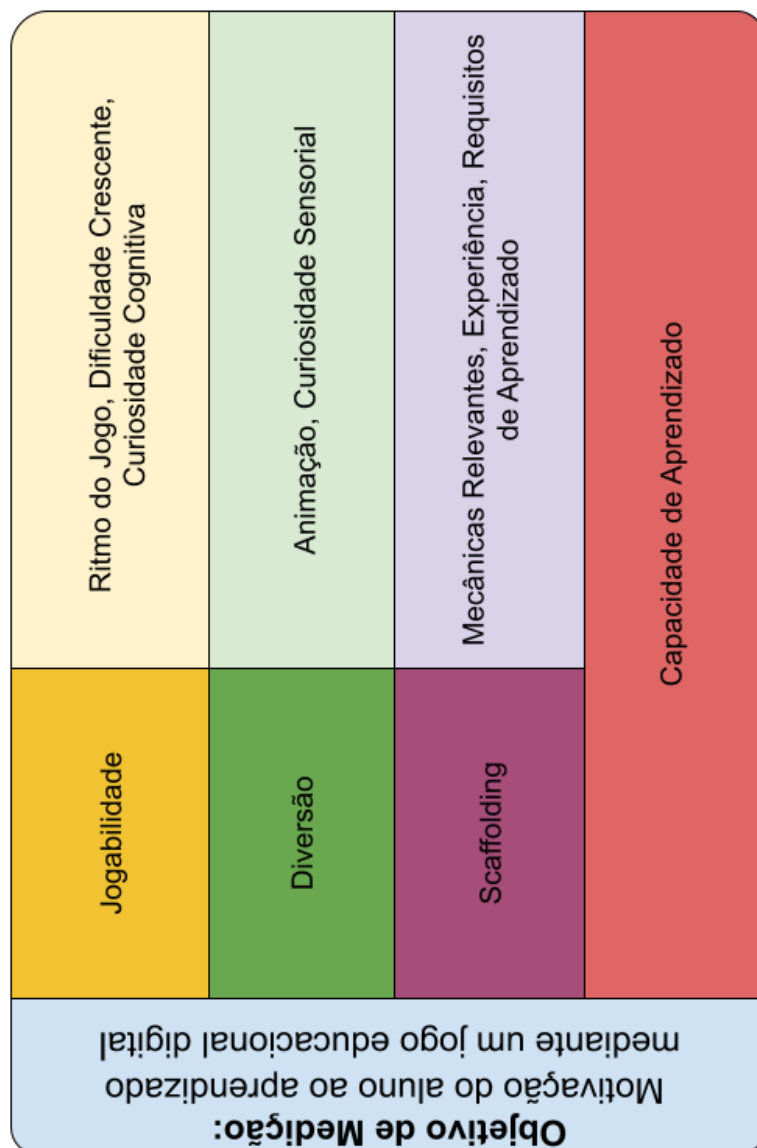
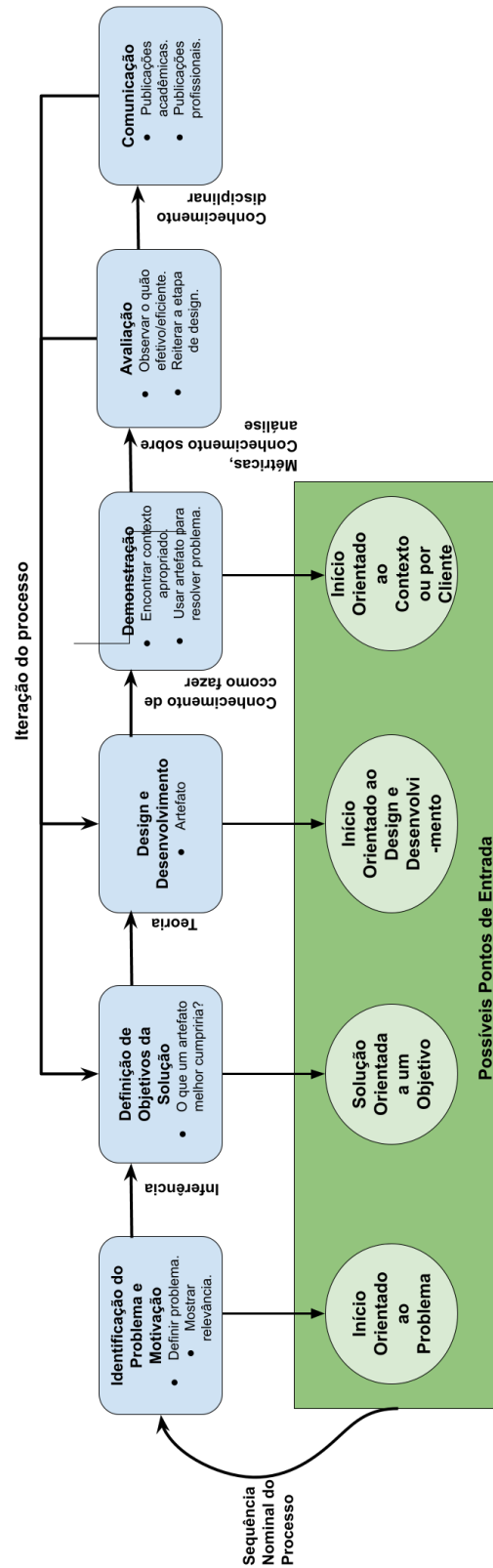


Figura A.3: Modelo de Avaliação da Motivação ao Aprendizado via Jogos Educacionais Digitais.

Figura A.4: Modelo do processo *Design Science Research Methodology*, adaptado de PEFFERS *et al.* (2007).



# Apêndice B

## Diagramas BPMN

### B.1 Desenho de Pesquisa de Análises Fatoriais

Figura B.1: *Overview* do desenho de pesquisa de análises fatoriais, adaptado de HAIR *et al.* (2009).

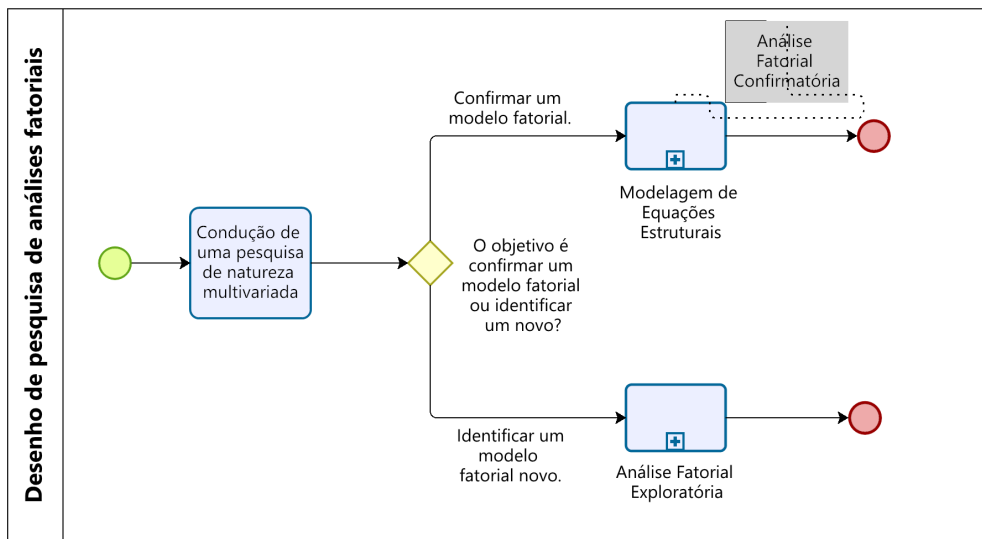




Figura B.2: Parte inicial do desenho de pesquisa de análises fatoriais, adaptado de HAIR *et al.* (2009).

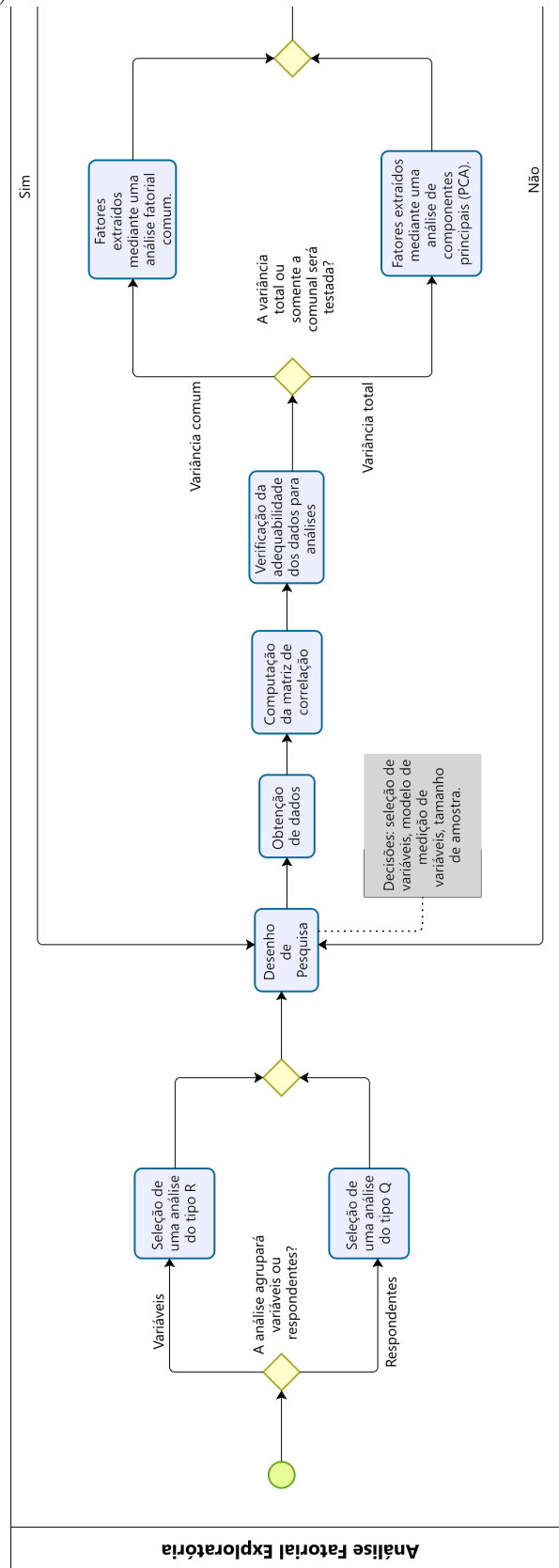
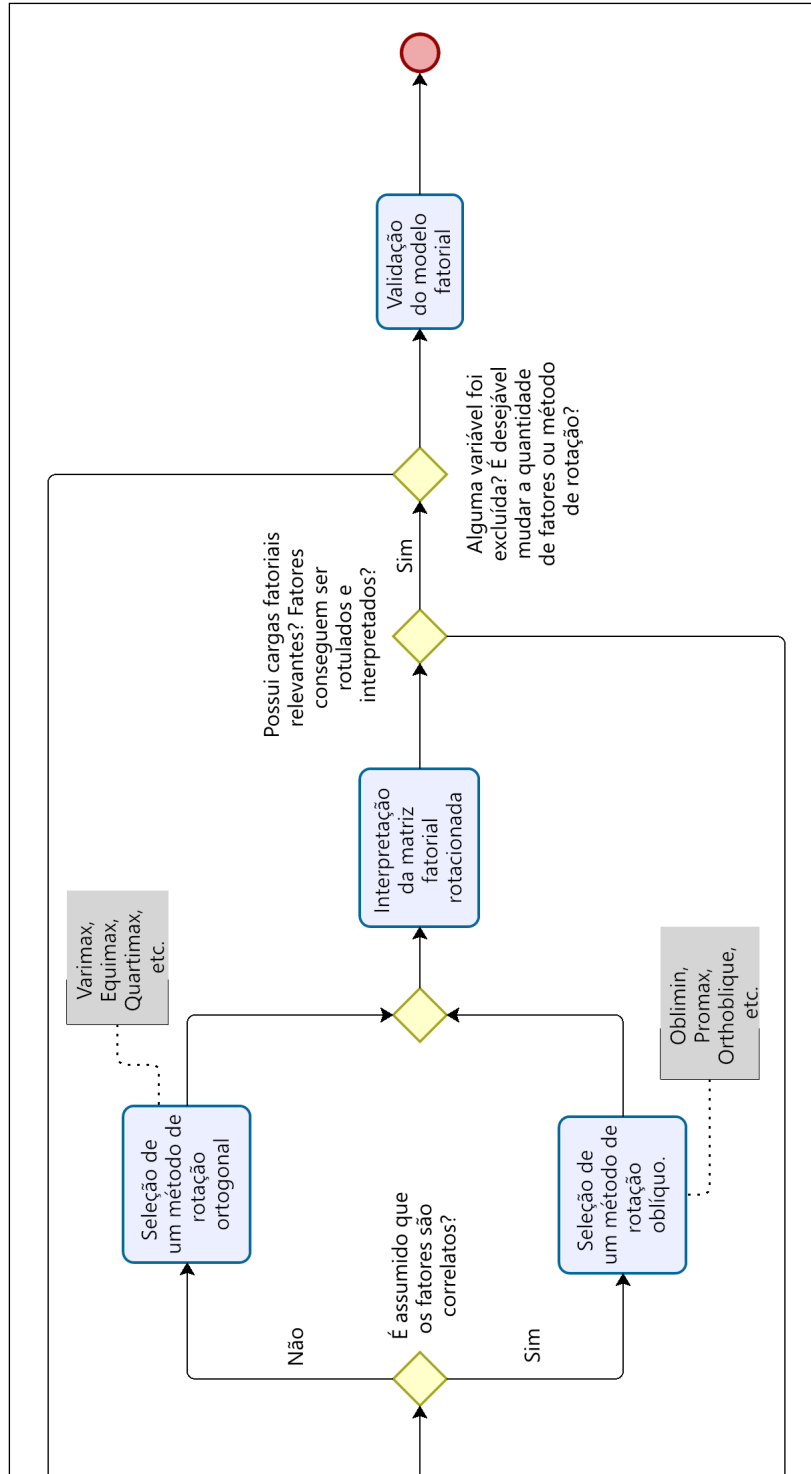


Figura B.3: Parte final do desenho de pesquisa de análises fatoriais, adaptado de HAIR *et al.* (2009).



## B.2 Processo de Desenvolvimento do Modelo e Formulário de Avaliação de Jogos Educacionais Digitais

Figura B.4: *Overview* do processo de desenvolvimento do modelo e formulário de avaliação de jogos educacionais digitais.

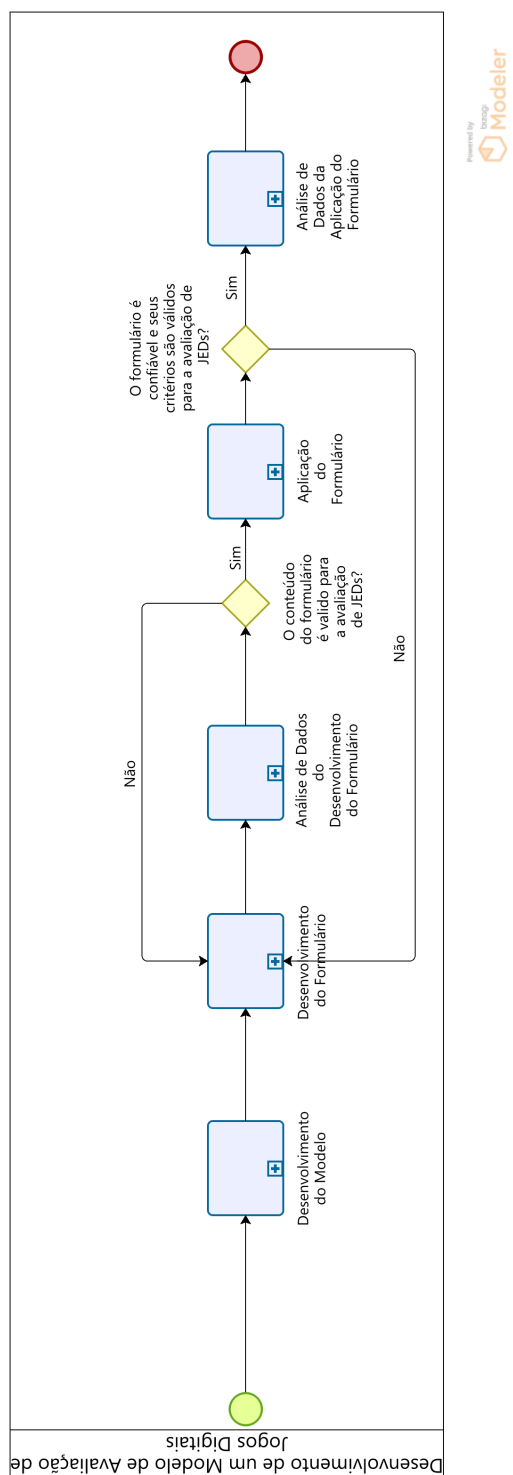


Figura B.5: Processo iterativo do desenvolvimento do modelo de avaliação de jogos educacionais digitais.

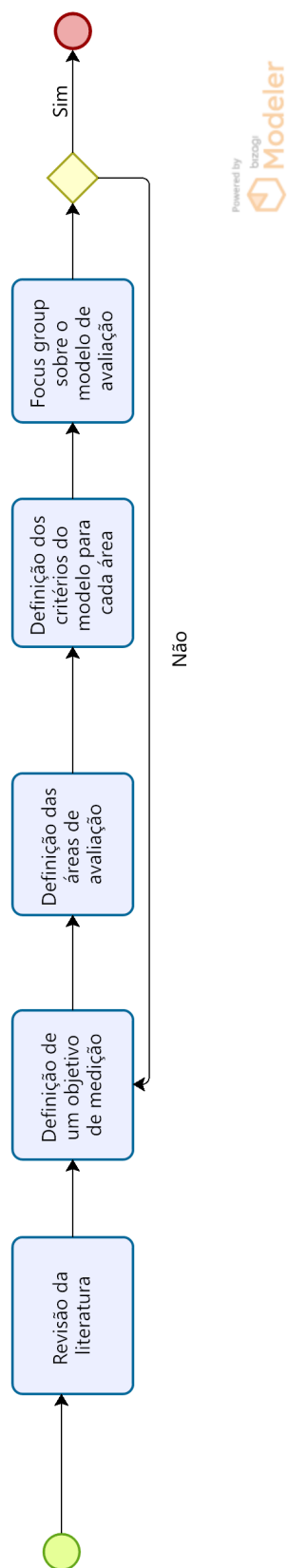


Figura B.6: Processo do desenvolvimento do formulário de avaliação de jogos educacionais digitais.

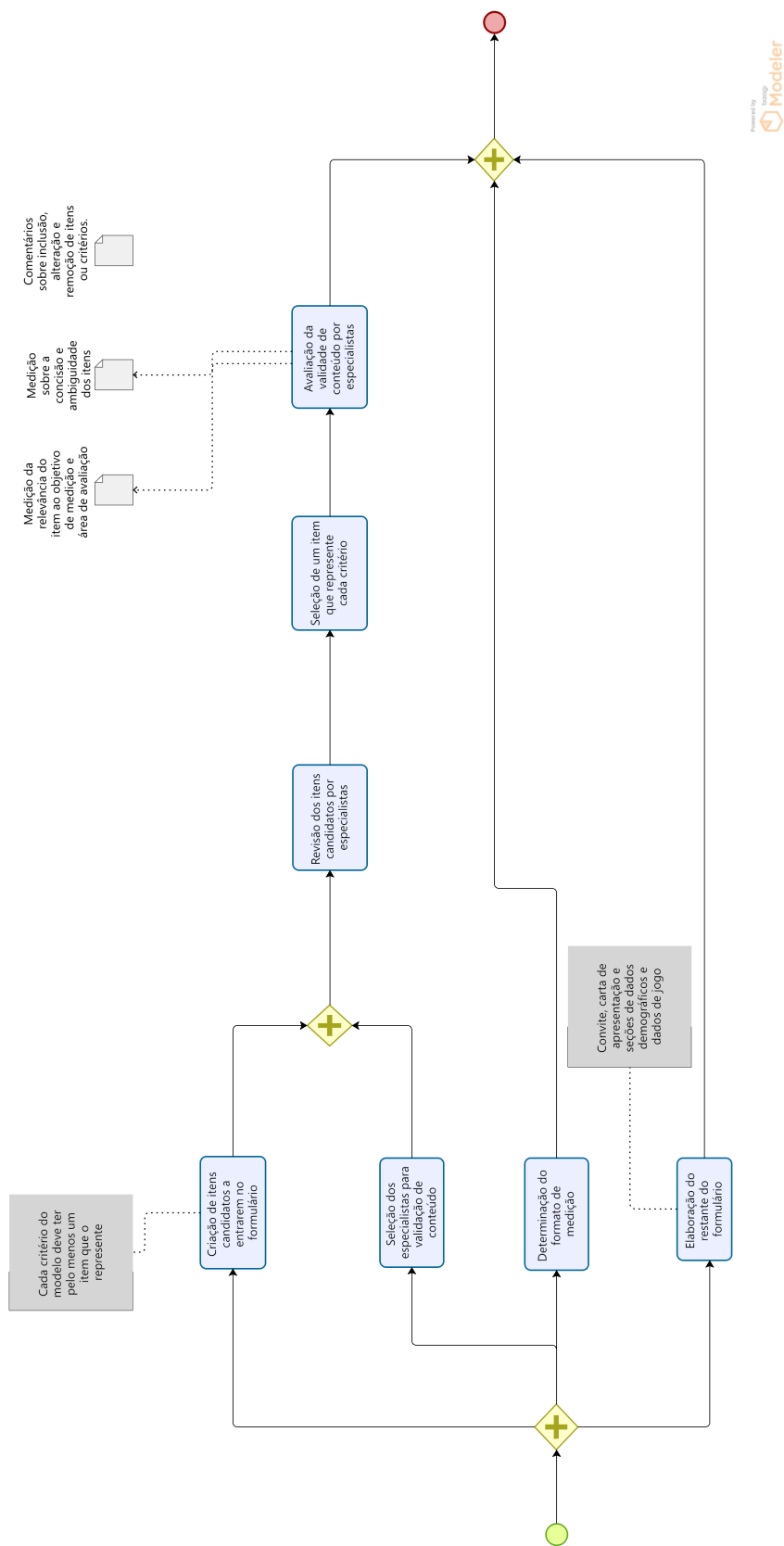


Figura B.7: Processo de análise de dados anterior à aplicação do formulário de avaliação de jogos educacionais digitais.

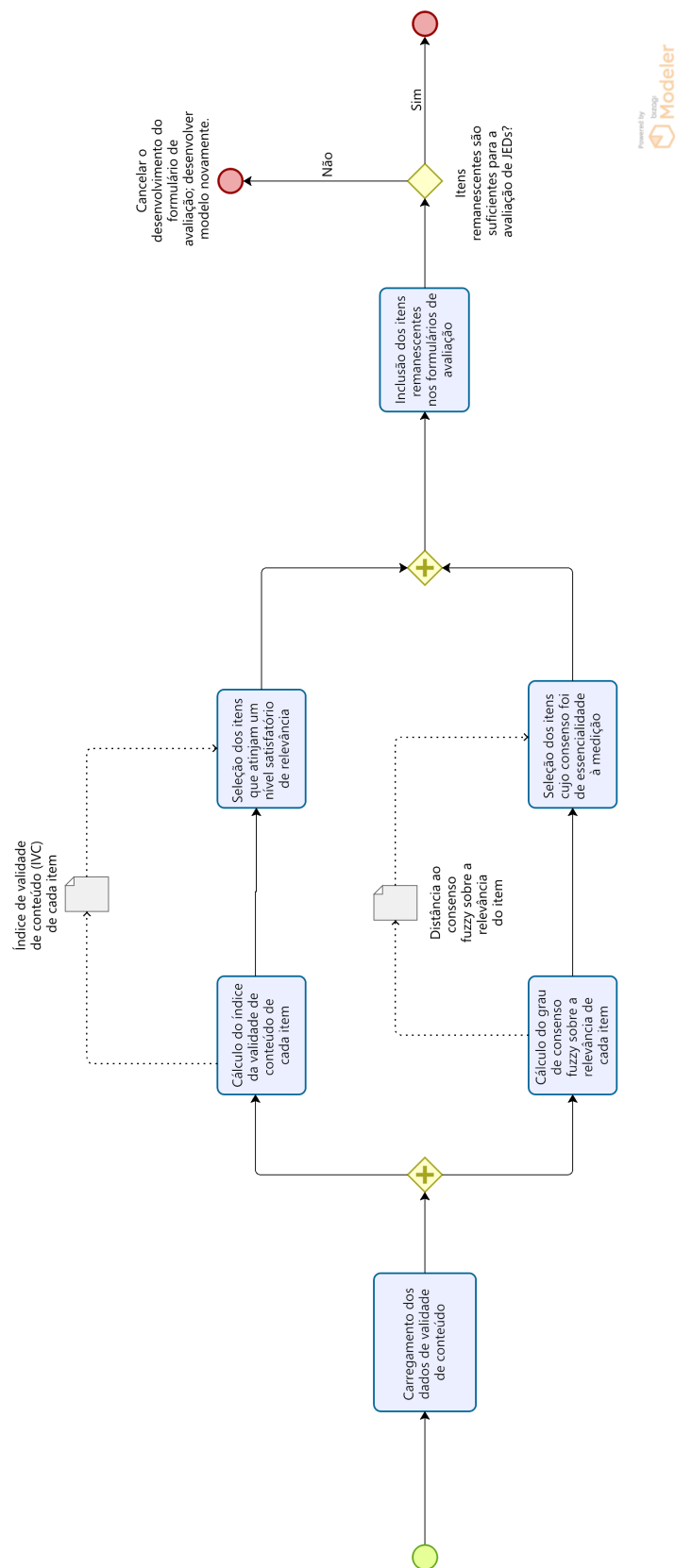


Figura B.8: Processo de aplicação do formulário de avaliação de jogos educacionais digitais.

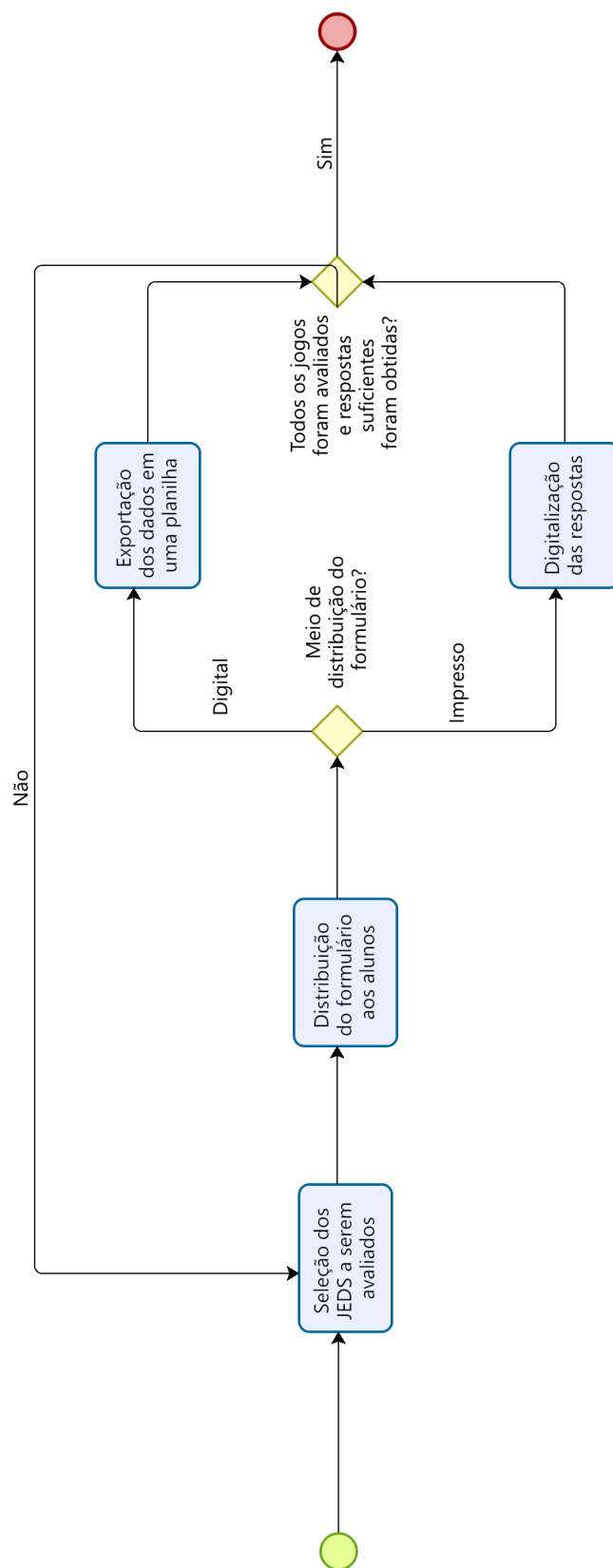


Figura B.9: Processo de análise de dados posterior à aplicação do formulário de avaliação de jogos educacionais digitais.

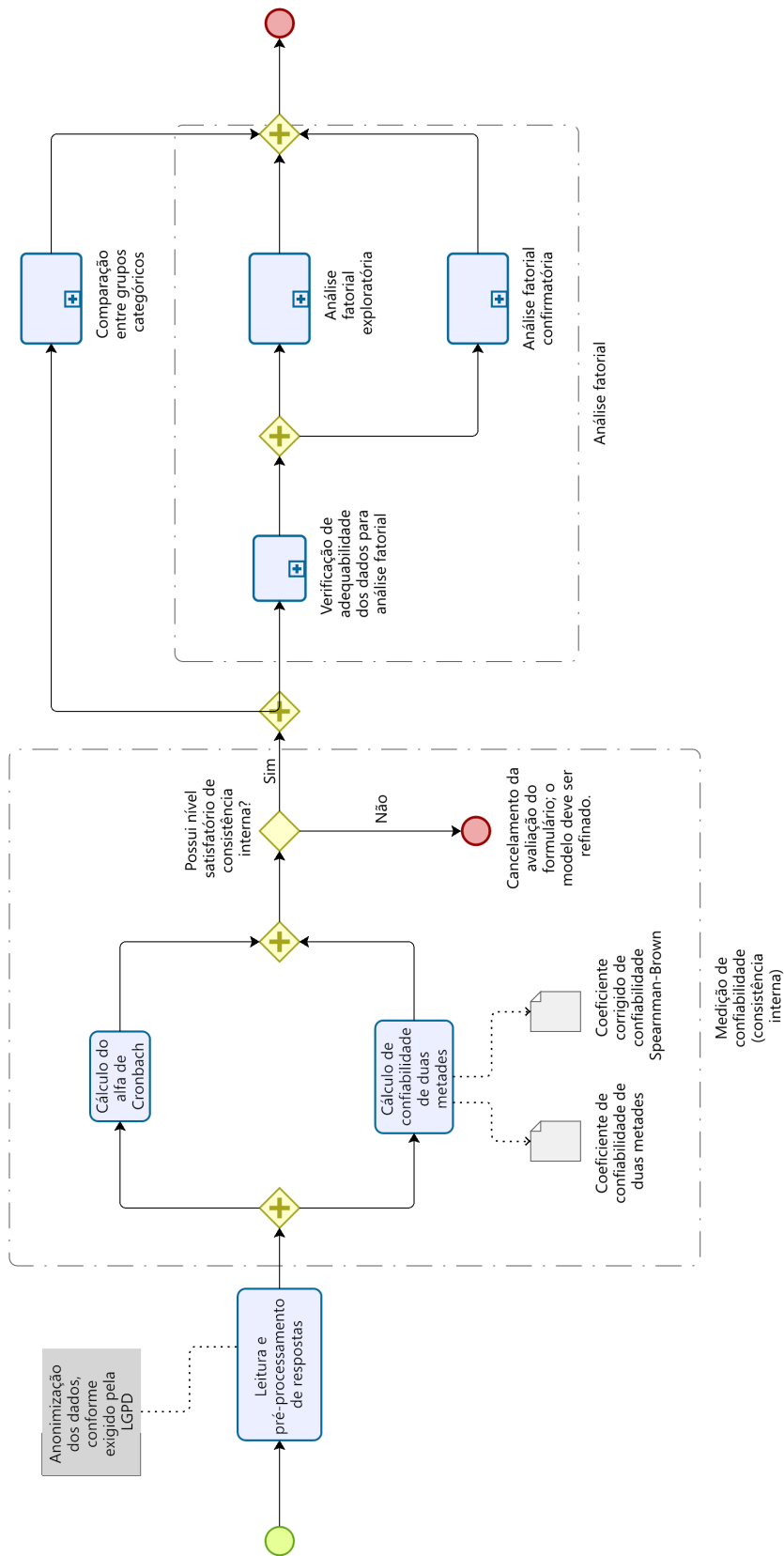




Figura B.10: Subprocesso da análise de adequabilidade dos dados da avaliação a uma análise fatorial.

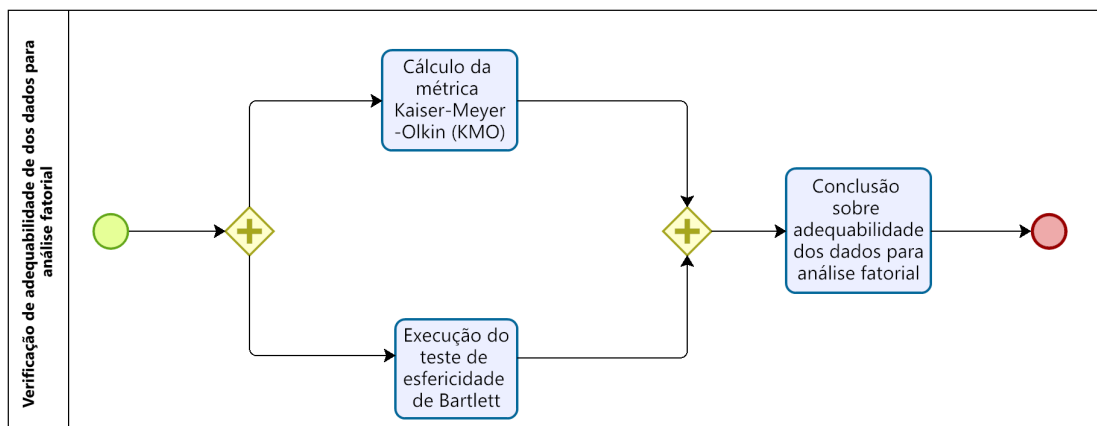


Figura B.11: Diagrama da análise fatorial confirmatória.

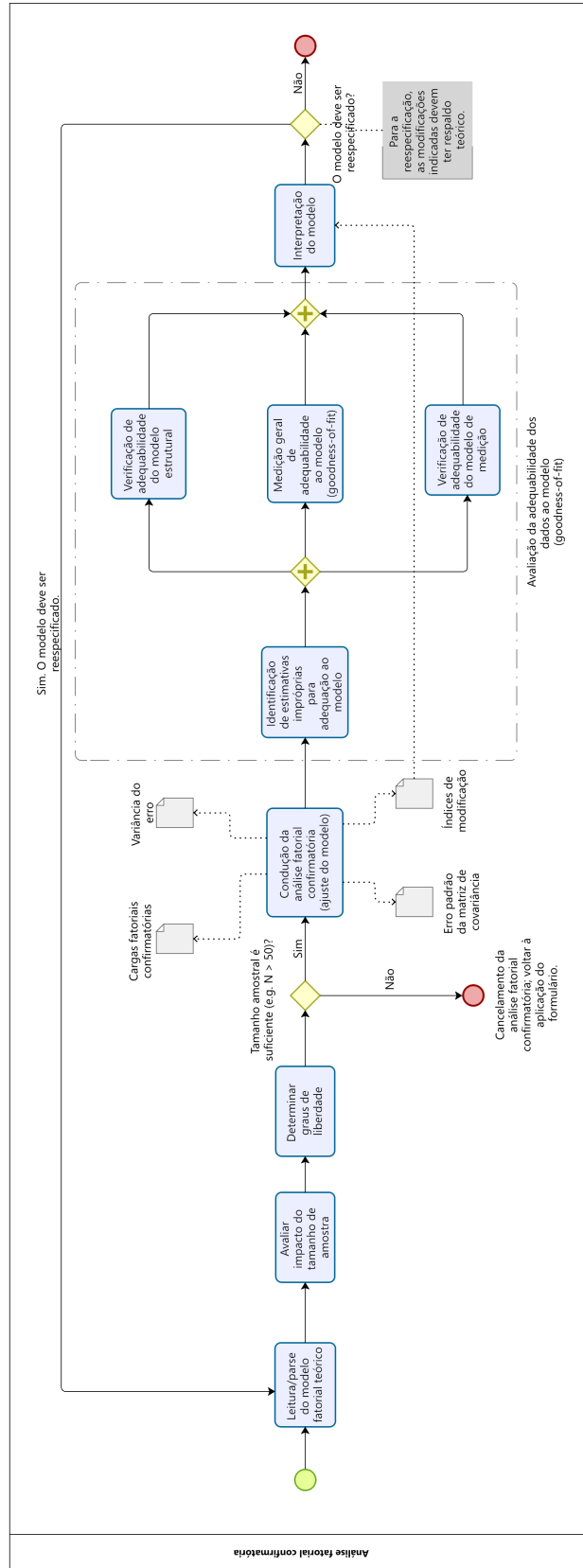


Figura B.12: Diagrama da análise fatorial exploratória.

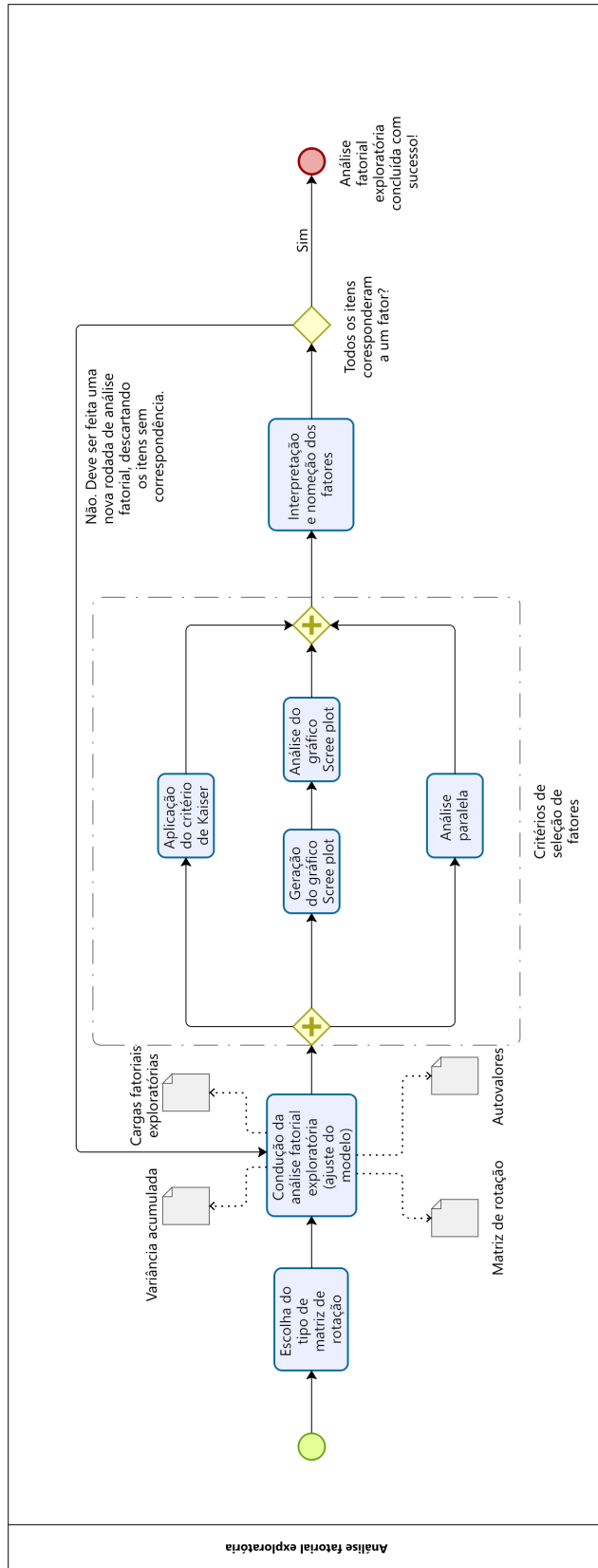
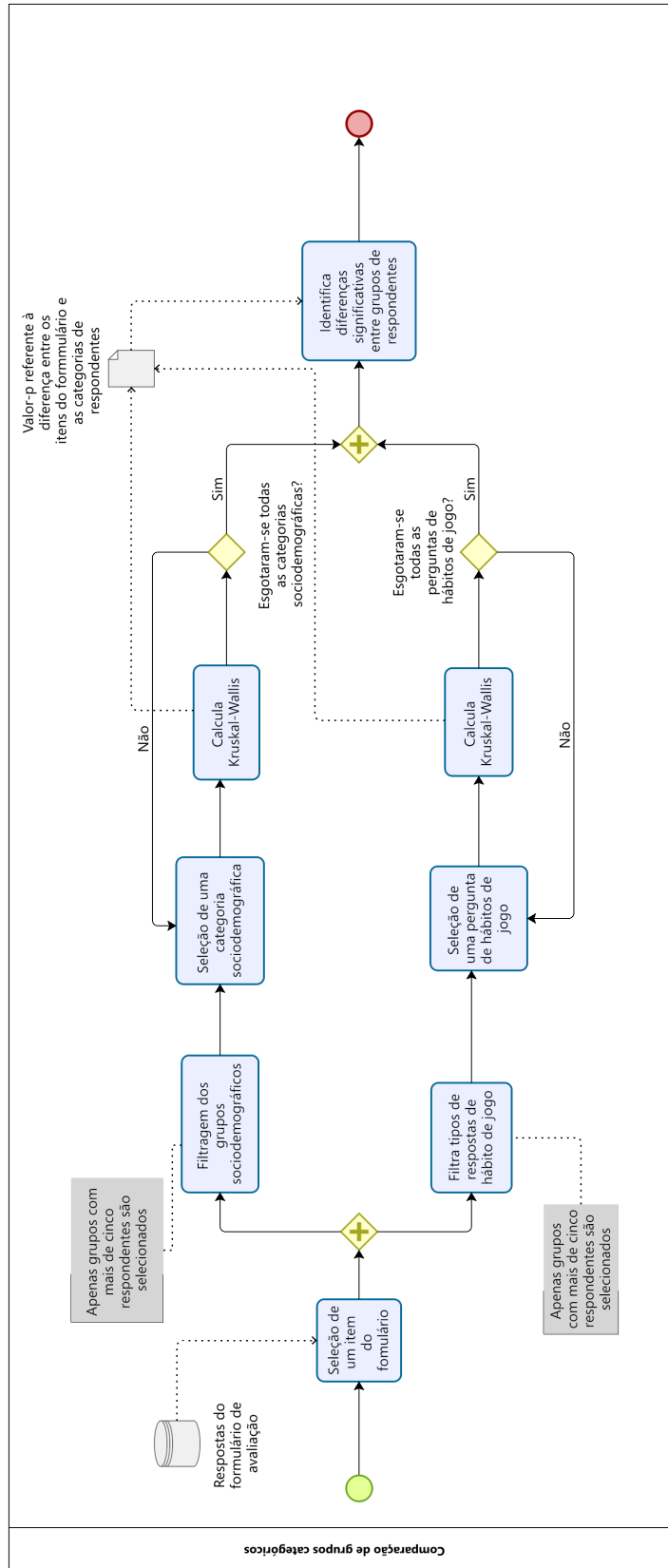


Figura B.13: Diagrama do subprocesso de comparação entre grupos categóricos.



# Apêndice C

## Itens do Formulário de Avaliação de Jogos Educacionais Digitais

### C.1 Versão anterior à primeira rodada de Validade de Conteúdo

#### C.1.1 Jogabilidade

Tabela C.1: Itens da Área de Avaliação de Jogabilidade

Critério	Item
<b>Animação</b>	Senti entusiasmo durante a minha experiência com o jogo.
<b>Engajamento com o Personagem</b>	Ao longo do jogo senti uma identificação com o papel que estava desempenhando.
<b>Fiero</b>	Senti alegria durante a minha experiência com o jogo.
<b>Imprevisibilidade nas Interações com o Sistema</b>	Foi fácil prever as respostas do jogo às minhas ações, dificilmente fiquei surpreso.
<b>Ausência de Medo</b>	Estive consciente de que minha experiência no jogo não envolvia riscos verdadeiros, e, portanto, não senti medo ao jogar.
<b>Sensação de Distorção de Tempo</b>	Senti o tempo passar em um ritmo diferente enquanto jogava, parecia estar passando mais devagar ou mais rápido.
<b>Compartilhamento</b>	A minha equipe compartilhou recursos e informações de forma ideal para conquistar os nossos objetivos.
<b>Variabilidade de Estratégias para Ganhar o Jogo</b>	Pude alternar as estratégias para tentar ganhar o jogo, não me senti limitado a uma única maneira de jogar.
<b>Velocidade das Interações com o Sistema</b>	O jogo respondeu às minhas ações com agilidade e velocidade.
<b>Variabilidade nas Ações do Agente</b>	O jogo me oferecia a todos os instantes uma quantidade desejável de ações alternativas para o meu personagem performar.
<b>Relaxamento</b>	Senti pouca ou nenhuma tensão durante o jogo.
<b>Interações Cooperativas</b>	Gostei das interações colaborativas, quando jogadores atuam em grupo para alcançar um objetivo compartilhado, que tive com outros jogadores.
<b>Interações Competitivas</b>	Os momentos em que competi com outros jogadores pelo mesmo objetivo foram positivos para a minha experiência.
<b>Contexto Social</b>	Tive conversas com colegas sobre o jogo e achei tais conversas produtivas para o meu aprendizado.
<b>Presença Física</b>	Durante o jogo, eu estava mais atento e comprometido com o ambiente do jogo do que com o seu entorno.
<b>Co-presença</b>	Apesar de estar em um ambiente virtual, tive a sensação de estar compartilhando a minha experiência de forma vívida com outros.
<b>Engajamento com o Mundo do Jogo</b>	Esqueci do mundo ao meu redor enquanto jogava, era como se eu estivesse presente no mundo do jogo.
<b>Competência</b>	Senti competência em relação ao meu desempenho no jogo, tive as habilidades necessárias para jogar bem.

## C.1.2 Experiência de Usuário

Tabela C.2: Itens de Avaliação da Experiência do Usuário

Critério	Item
Volume Ajustável	É possível ajustar ou silenciar o volume do jogo, diferenciando entre os volumes dos efeitos sonoros, trilhas sonoras e narração.
Proteção de Erros	O jogo apresentou erros que não afetaram a continuidade do jogo, mas levaram a comportamentos inesperados pelo sistema.
Reconhecimento de Adequabilidade	Apreendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.
Ausência de Erros Disruptivos	Presenciei algum erro que interrompeu ou terminou o jogo de forma inesperada.
Acessibilidade para Pessoas com Daltonismo	O jogo não passou informações e sinais importantes para o andamento do jogo exclusivamente por cor.
Textos Bem Formatados	O texto estava bem formatado e contrastava com o plano de fundo, facilitando a sua leitura.
Capacidade de Aprendizado	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo um domínio dos controles, antes que se tornasse desafiador.
Acessibilidade para Pessoas com Deficiência Auditiva	O jogo não dependeu de narrações, efeitos, músicas ou outros efeitos sonoros para passar informações importantes ao seu andamento.
Nitidez da Imagem	O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.
Linguagem Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.
Atratividade	Achei a interface do jogo bonita, o aspecto visual ou sonoro cativou-me de uma forma positiva.

## C.1.3 Motivação ao Aprendizado

Tabela C.3: Itens de Avaliação de Motivação ao Aprendizado

Critério	Item
Humor	Houve momentos de descontração durante o jogo.
Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.
Demanda de Habilidade	Senti compatibilidade entre as demandas do jogo e as minhas habilidades para superá-las.
Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados para avaliar o meu desempenho.
Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.
Intervalo de Reforços Positivos	As recompensas dadas pelo jogo ficam progressivamente mais difíceis de conquistar.
Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.
Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre o jogo.
Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.
Temática Relevante	A temática e a narrativa são relevantes para o aprendizado oferecido pelo jogo.
Evitar Influências Negativas	Ao jogar, não senti uma pressão ao ponto de que meu bem-estar e desempenho no jogo fossem prejudicados.
Autoconfiança	Tive confiança e independência suficientes para aplicar o conhecimento adquirido pelo jogo em situações realísticas.
Recompensas para Tarefas Entediantes	As tarefas que considerei entediantes eram recompensadas de forma evidente e previsível.
Recompensas para Tarefas Interessantes	Ao completar tarefas interessantes, o jogo me recompensava inesperadamente, variando de alguma forma na recompensa que era dada.
Conforto	O jogo não me causou nenhum momento em que fiquei fisicamente desconfortável.
Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.
Objetivo do Jogo	Achei os objetivos do jogo claros, inclusive os sub-objetivos ou a otimização do desempenho das tarefas.
Concretude	Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.
Elogios a Resultados Positivos	Recebia elogios por parte do jogo ao completar tarefas ou ter um bom desempenho no jogo.
Moderação de Expectativas	Senti que era possível ter um desempenho legal no jogo mediante o meu esforço.

# Apêndice D

## Formulário do Modelo de Avaliação de Jogos Educacionais Digitais

### D.1 Formulários de Validade de Conteúdo

Figura D.1: Carta de apresentação do formulário de validade de conteúdo para a *Motivação ao Aprendizado*.

**Termo de Consentimento Livre e Esclarecido**

Olá,

Sou Alexandre Vaz, mestrando pelo PESC/UFRJ, sob a orientação do Professor Geraldo Xexéo. Estou desenvolvendo um *Formulário de Avaliação de Reação a Jogos Educacionais Digitais* e convido-lhe a participar da revisão de especialistas para validar o conteúdo do formulário.

Você foi escolhido para essa pesquisa por sua experiência na área de Educação, Ludologia ou experiência com o desenvolvimento/aplicação de Jogos Educacionais. O objetivo desse formulário é validar os itens que serão incluídos em uma escala para a medição de avaliação da reação de alunos a jogos educacionais digitais. Para avaliar a validade do conteúdo dos itens candidatos a entrarem na escala, este formulário terá duas perguntas por item candidato a entrar na escala de avaliação:

- A primeira medirá a relevância do item quanto ao fenômeno que pretende avaliar, a **motivação ao aprendizado**.
- A segunda perguntará se o item está ambíguo ou prolixo.

No final do formulário, respondente poderão sugerir outros critérios considerados essenciais ou úteis para a avaliação de jogos educacionais digitais e comentar sobre os itens apresentados neste formulário.

Os critérios desse formulário pretendem avaliar a **motivação ao aprendizado**, interpretada nesse modelo como a vontade intrínseca e a intenção de fazer o esforço necessário para aprender e também a realização da atividade em um ambiente favorável ao aprendizado. O formulário será aplicado logo após a atividade envolvendo o jogo educacional digital, avaliando a reação do aluno a aspectos do jogo relacionados à sua motivação ao aprendizado. Os itens passando por uma avaliação da validade de conteúdo devem ser relevantes para a medição do fenômeno de **motivação ao aprendizado**, itens que tiverem um nível insatisfatório de validade de conteúdo serão descartados do formulário.

Sua participação é muito importante para a implementação dessa pesquisa, mas a pesquisa é feita de forma voluntária, então saiba que você pode desistir a qualquer momento caso não queira mais fazer parte da pesquisa.

As informações pertencentes a esse estudo poderão ser analisadas por especialistas autorizados, mas de forma codificada, visando a privacidade e confidencialidade da sua identidade. Garantimos que na eventual publicação dos dados e resultados do estudo o seu nome será ocultado. Os procedimentos dessa pesquisa seguirão rigorosamente as normas indicadas pela Lei Geral de Proteção de Dados (LGPD, Lei nº 13.709/2018).

As despesas necessárias para a realização dessa pesquisa não estão sob sua responsabilidade, mas também não haverá alguma compensação financeira pela sua participação.

Obrigado pela leitura dessas informações e por sua participação!

## D.1.1 Respostas

### Área de Avaliação I: Jogabilidade (Gráficos)

Figura D.2: Distribuição das respostas de validade de conteúdo na área de avaliação de *Jogabilidade*.

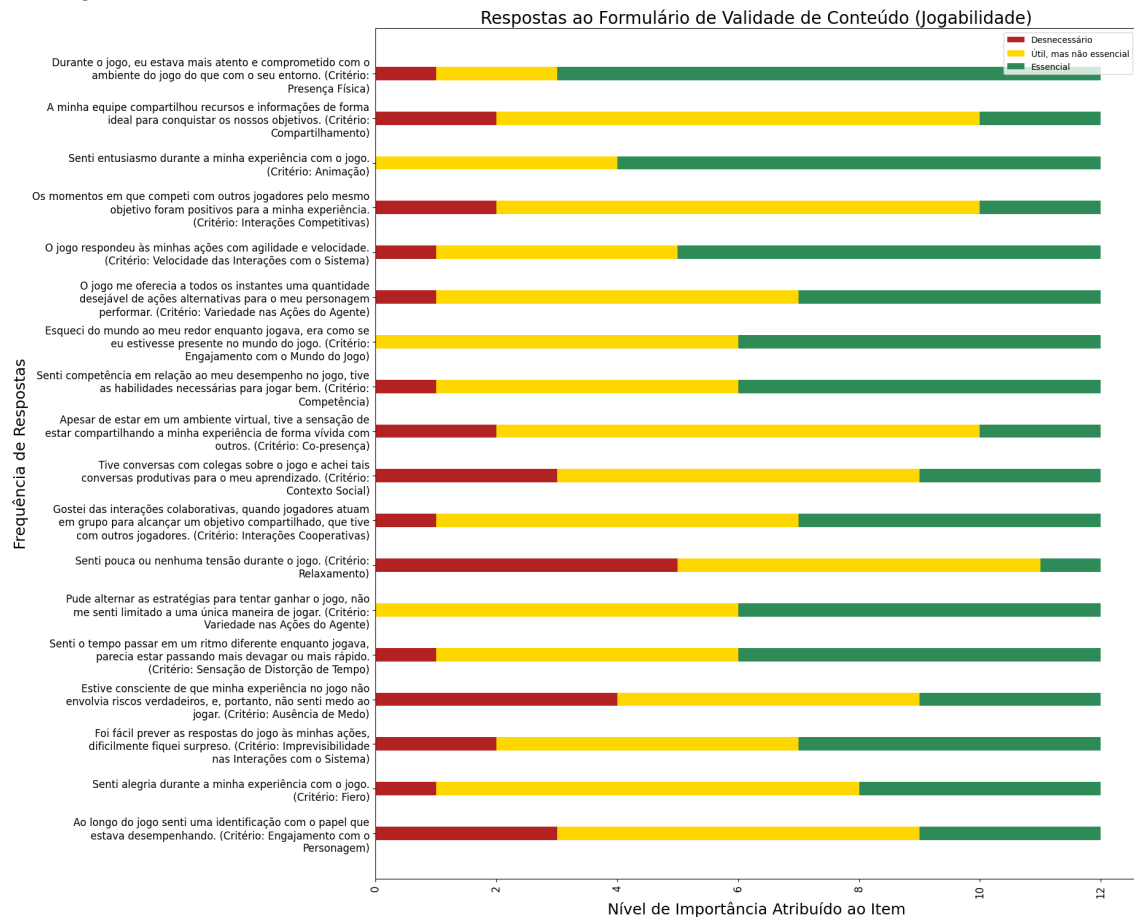




Figura D.3: Índice de Validade de Conteúdo da área de avaliação de *Jogabilidade*.

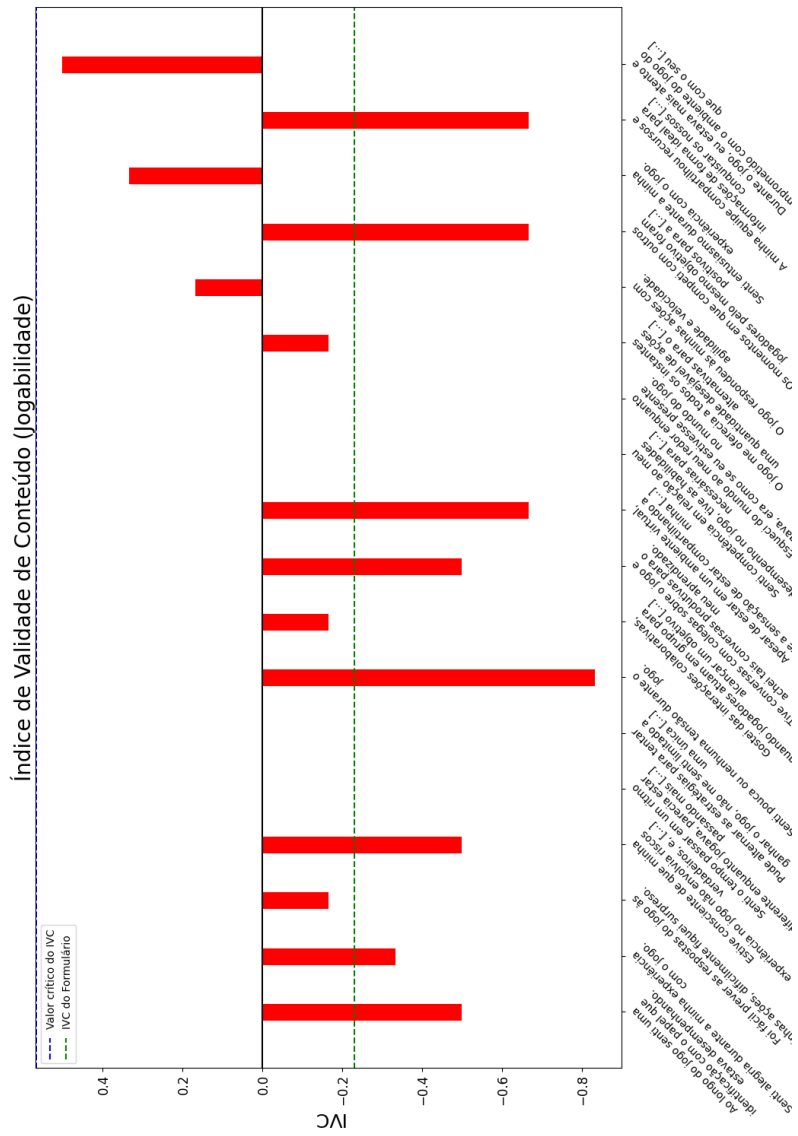




Figura D.5: Distância ao Consenso para a área de avaliação de jogabilidade.

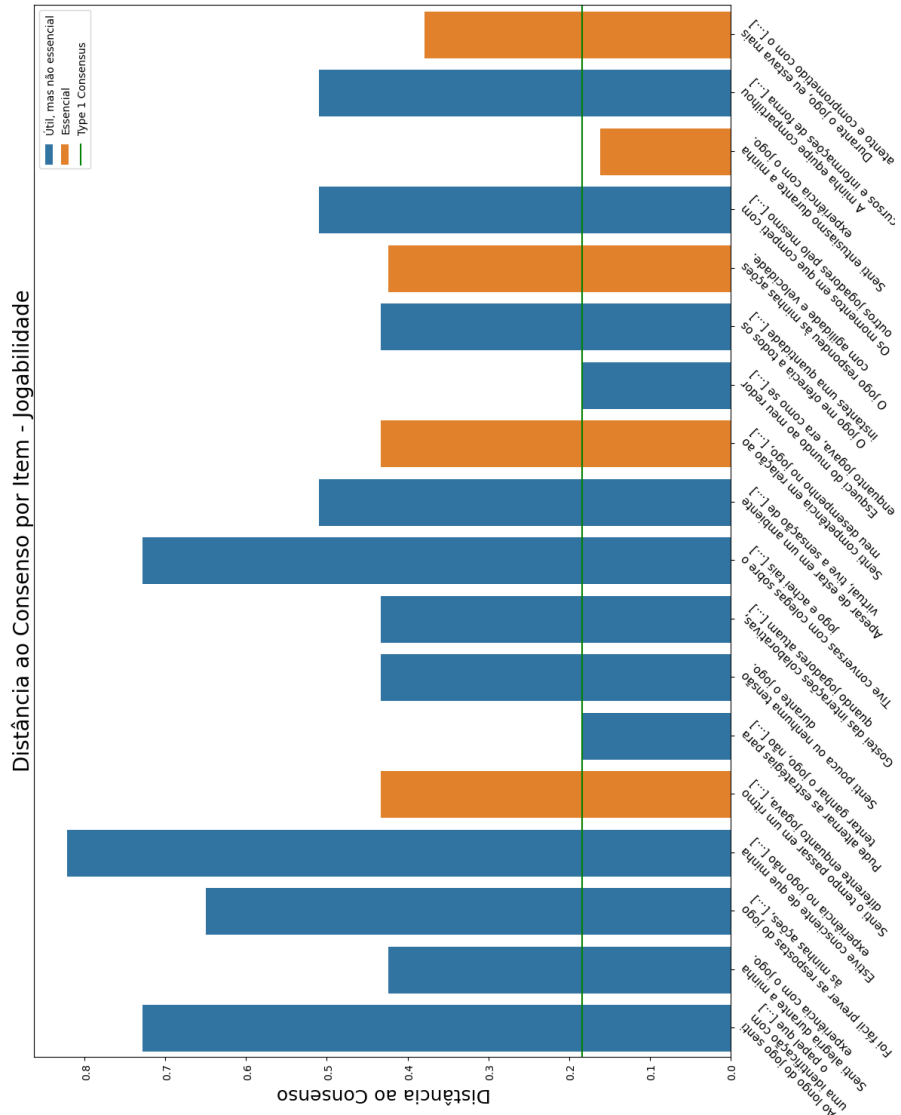
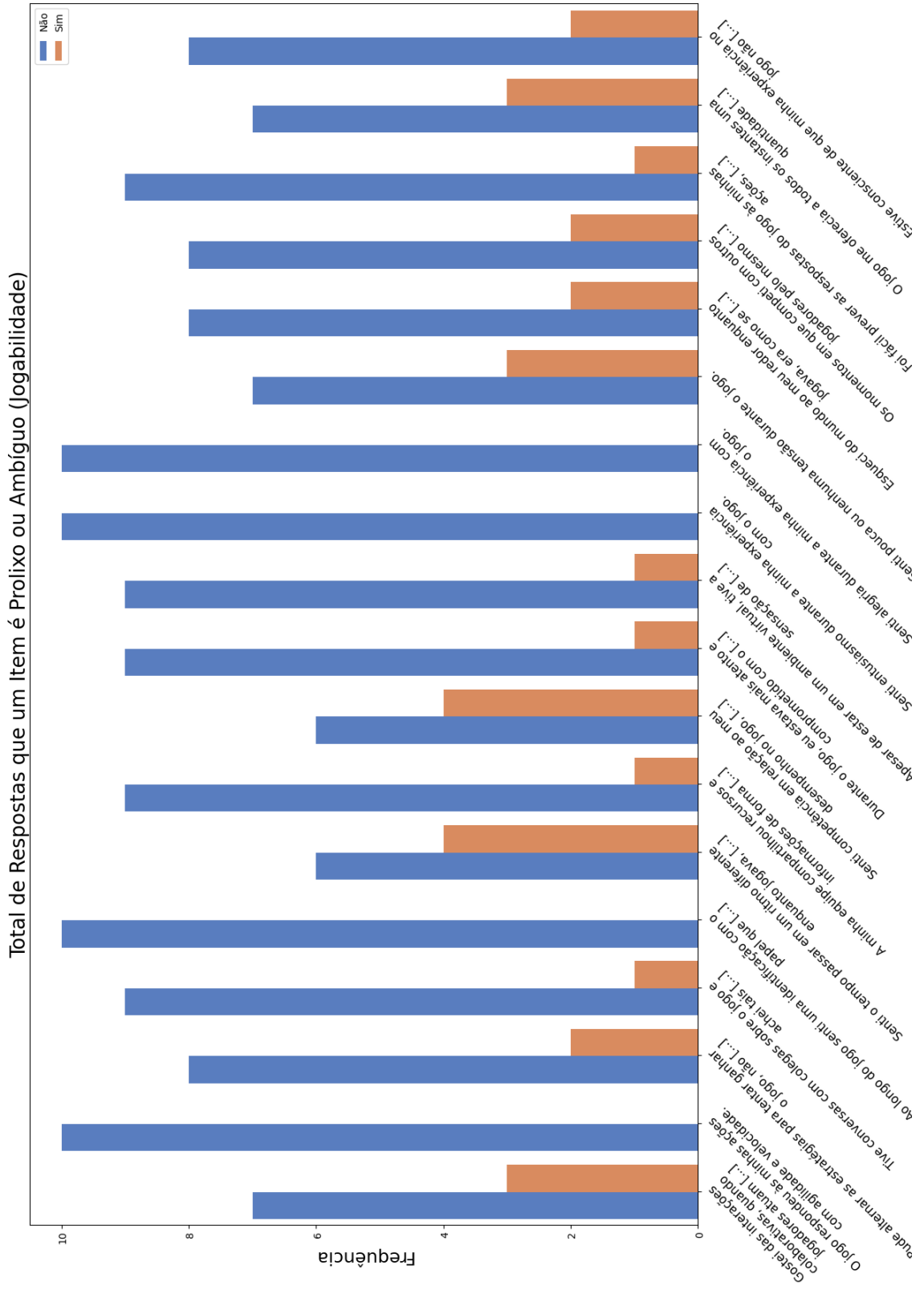


Figura D.6: Ambiguidade dos itens da área de avaliação de jogabilidade.



## D.1.2 Área de Avaliação II: Experiência de Usuário (Gráficos)

Figura D.7: Distribuição das respostas de validade de conteúdo na área de *Experiência de Usuário*.

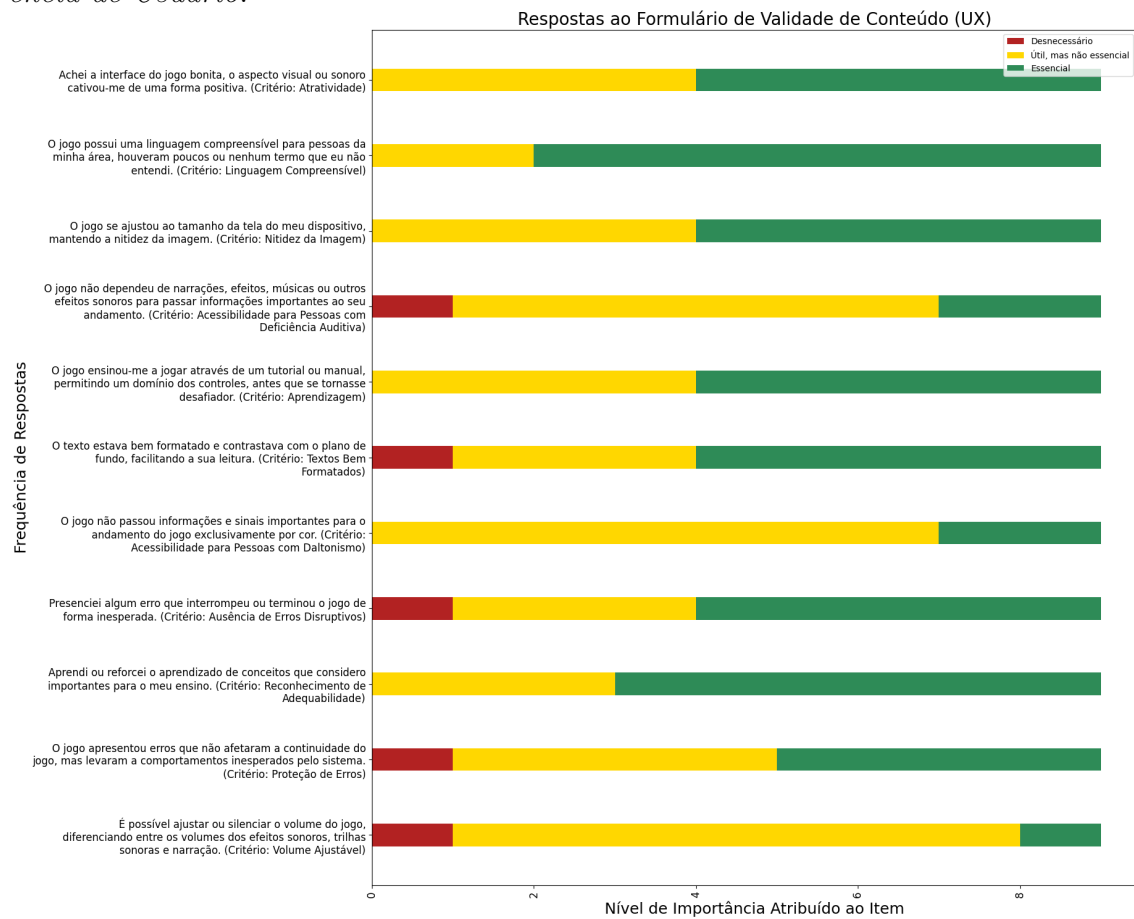
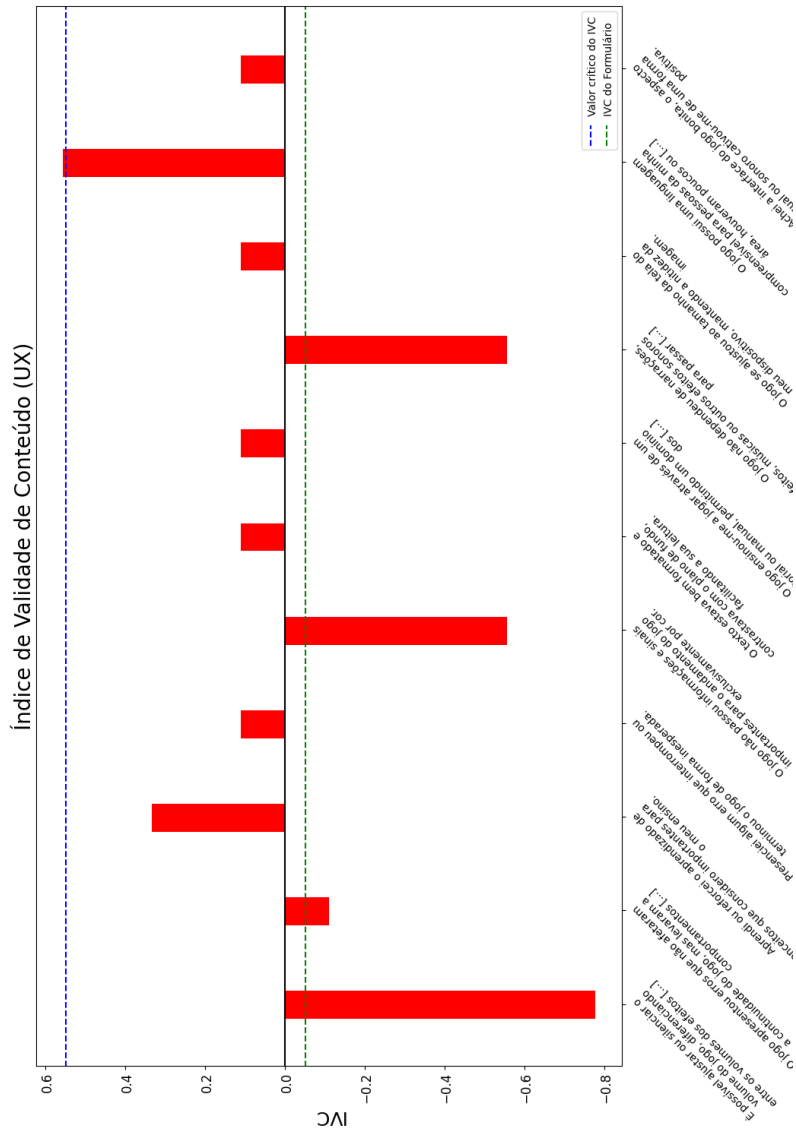


Figura D.8: Índice de Validade de Conteúdo da área de avaliação de *Experiência de Usuário*.



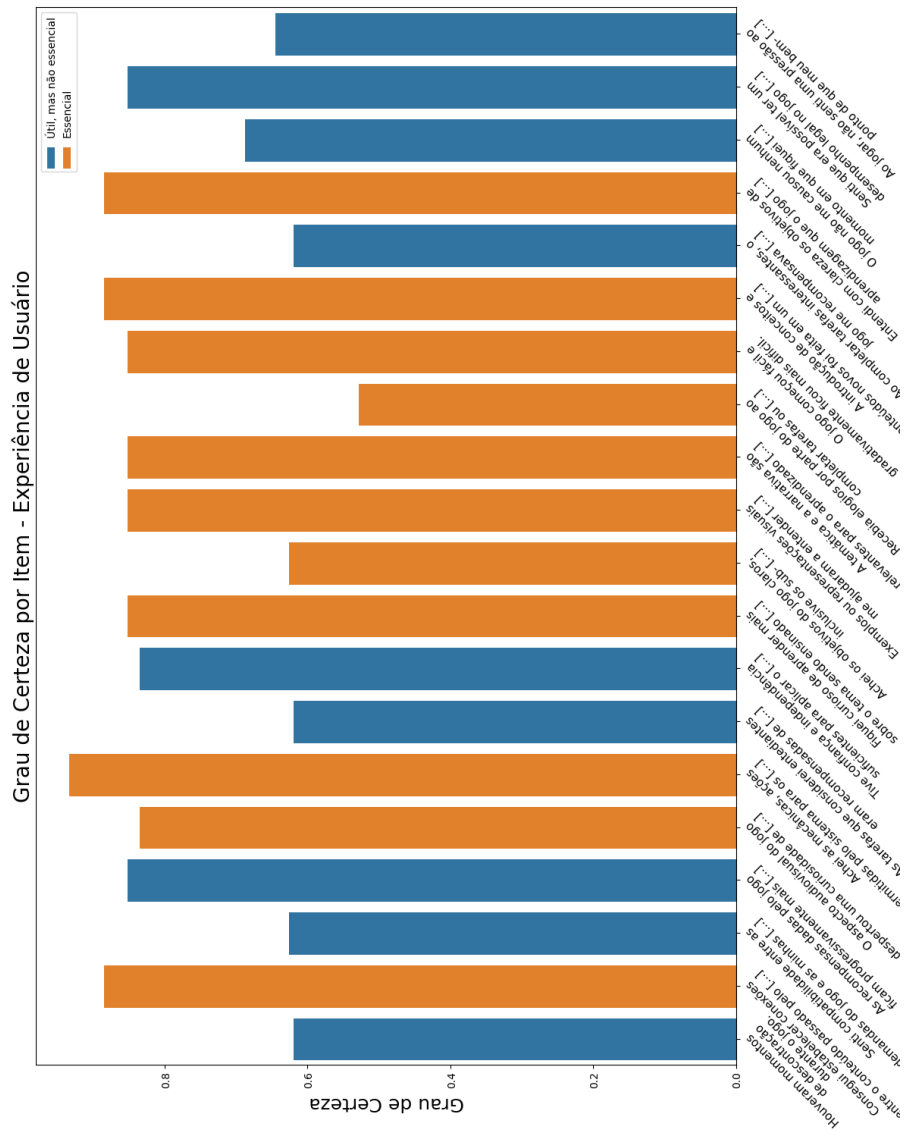


Figura D.9: Grau de Certeza para a Área de Avaliação de Experiência de Usuário

Figura D.10: Ambiguidade dos itens da área de avaliação de *Experiência de Usuário*.

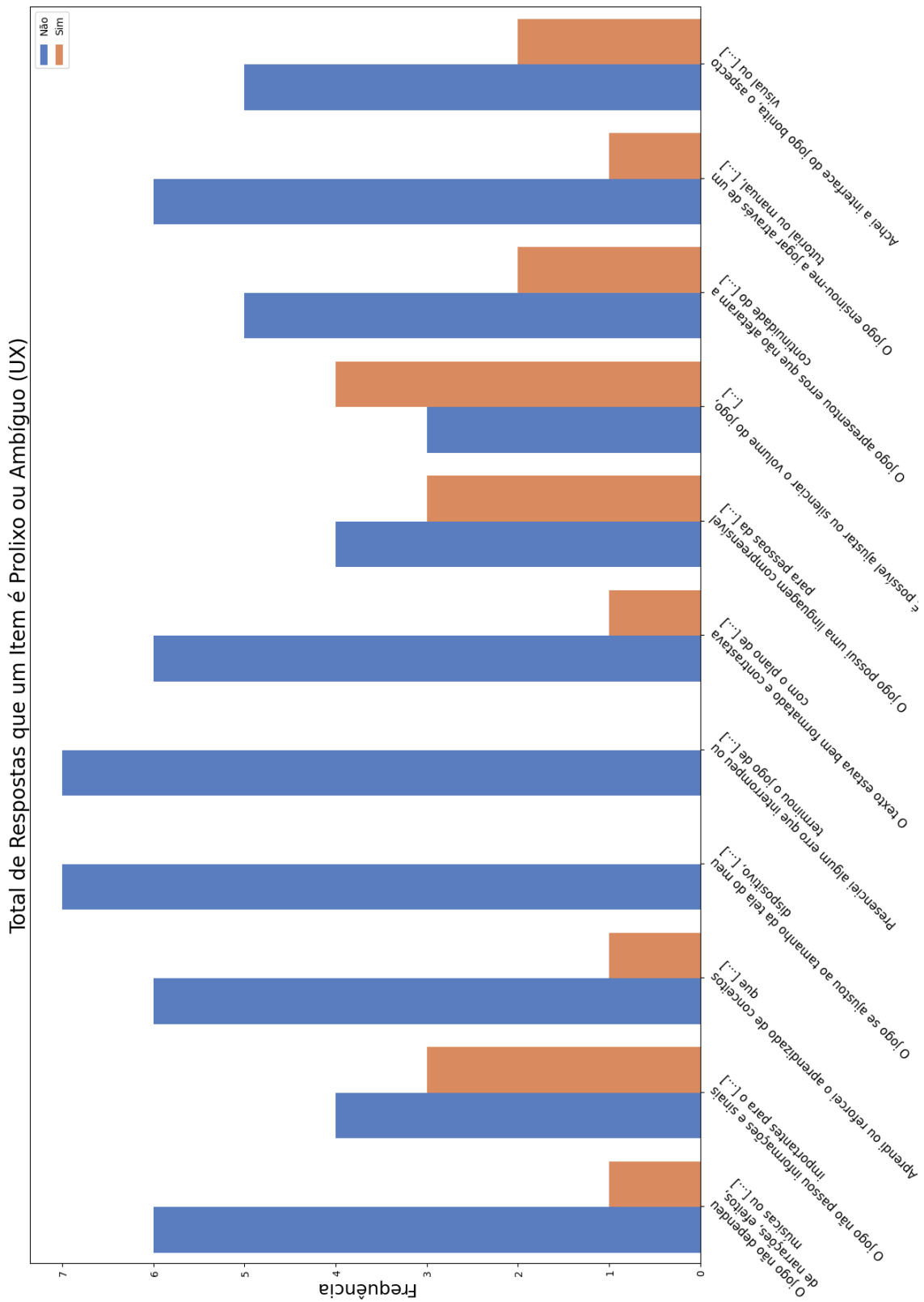




Figura D.11: Distribuição das respostas de validade de conteúdo na área de avaliação de *Motivação ao Aprendizado*.

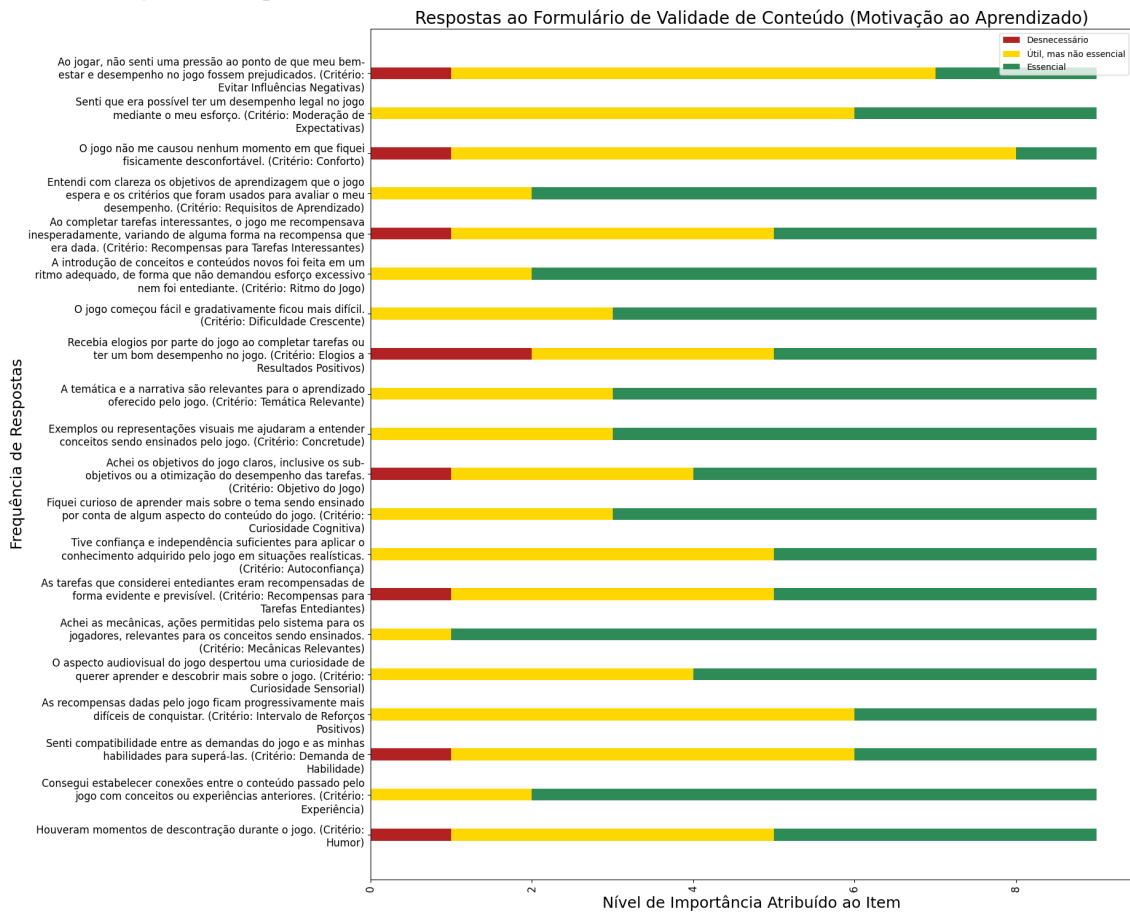




Figura D.13: Grau de Certeza para a área de avaliação de *Motivação ao Aprendizado*.

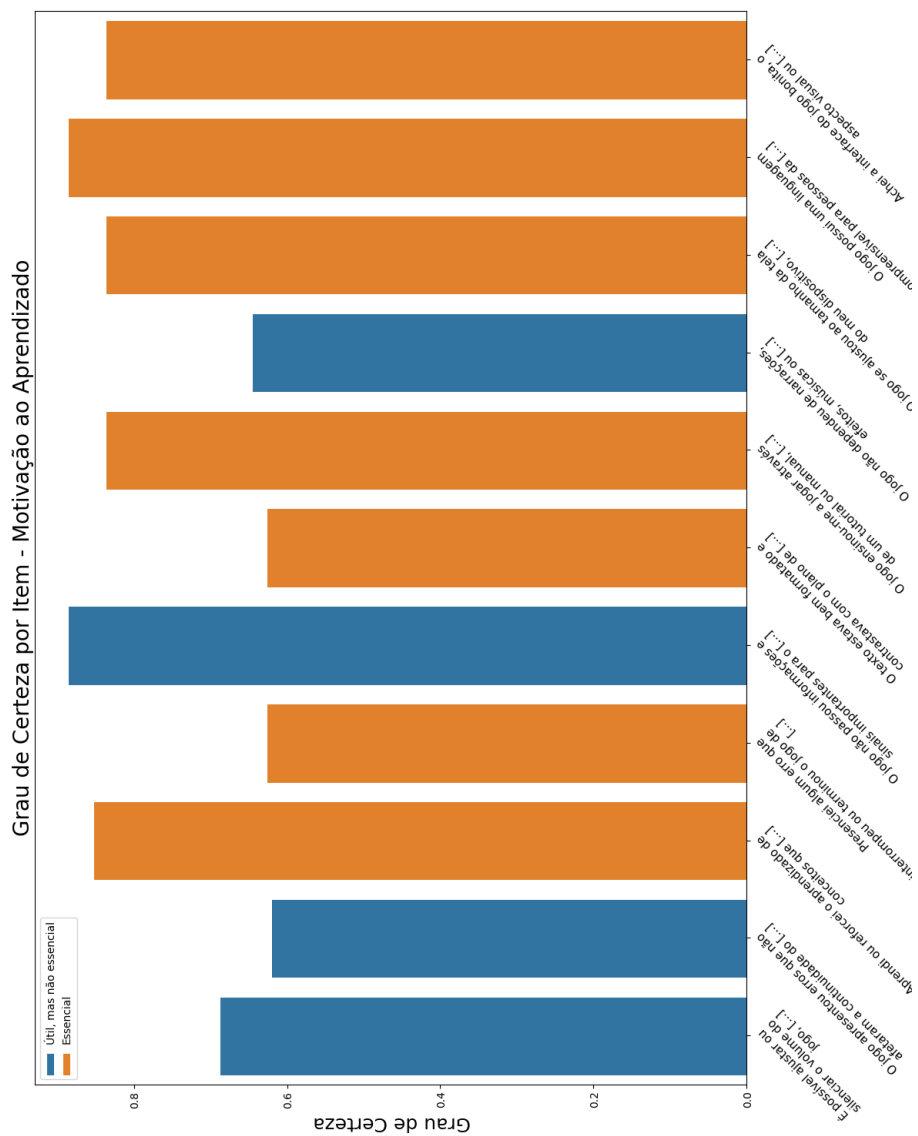
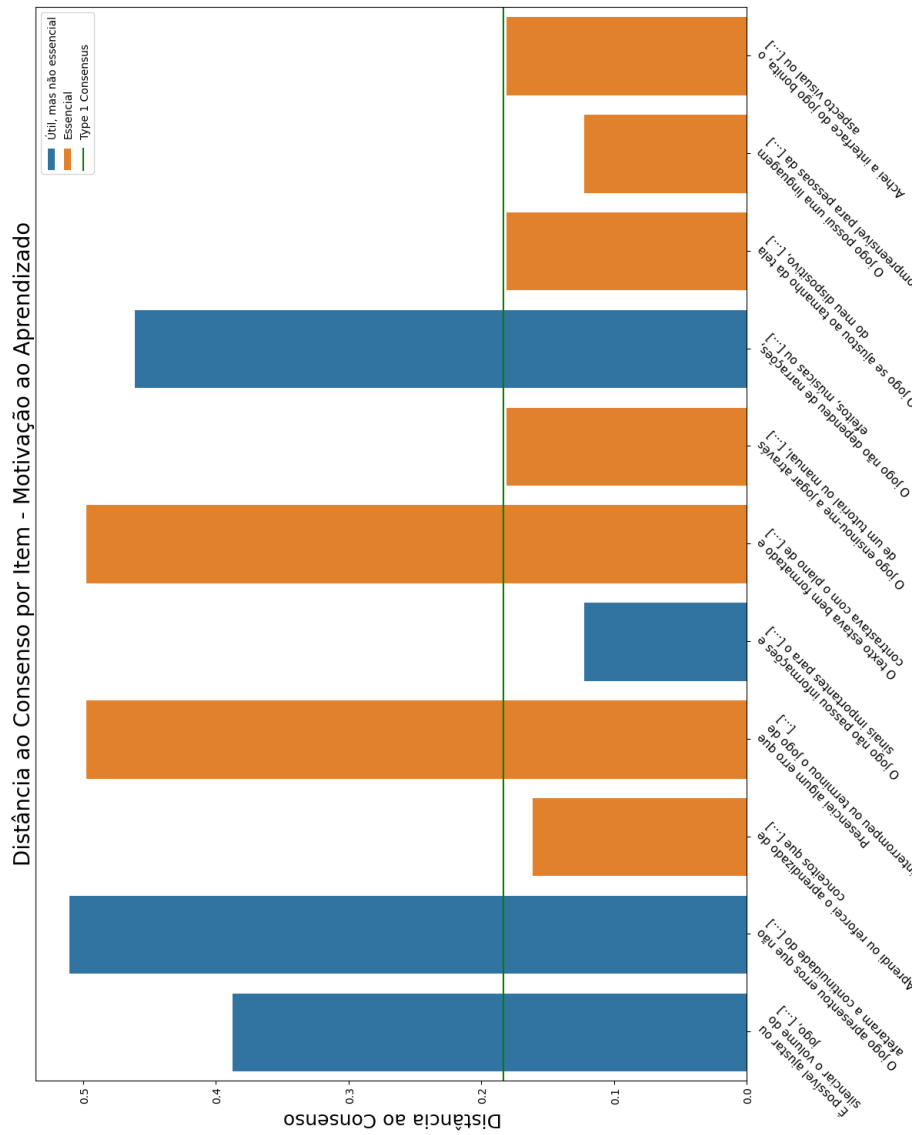
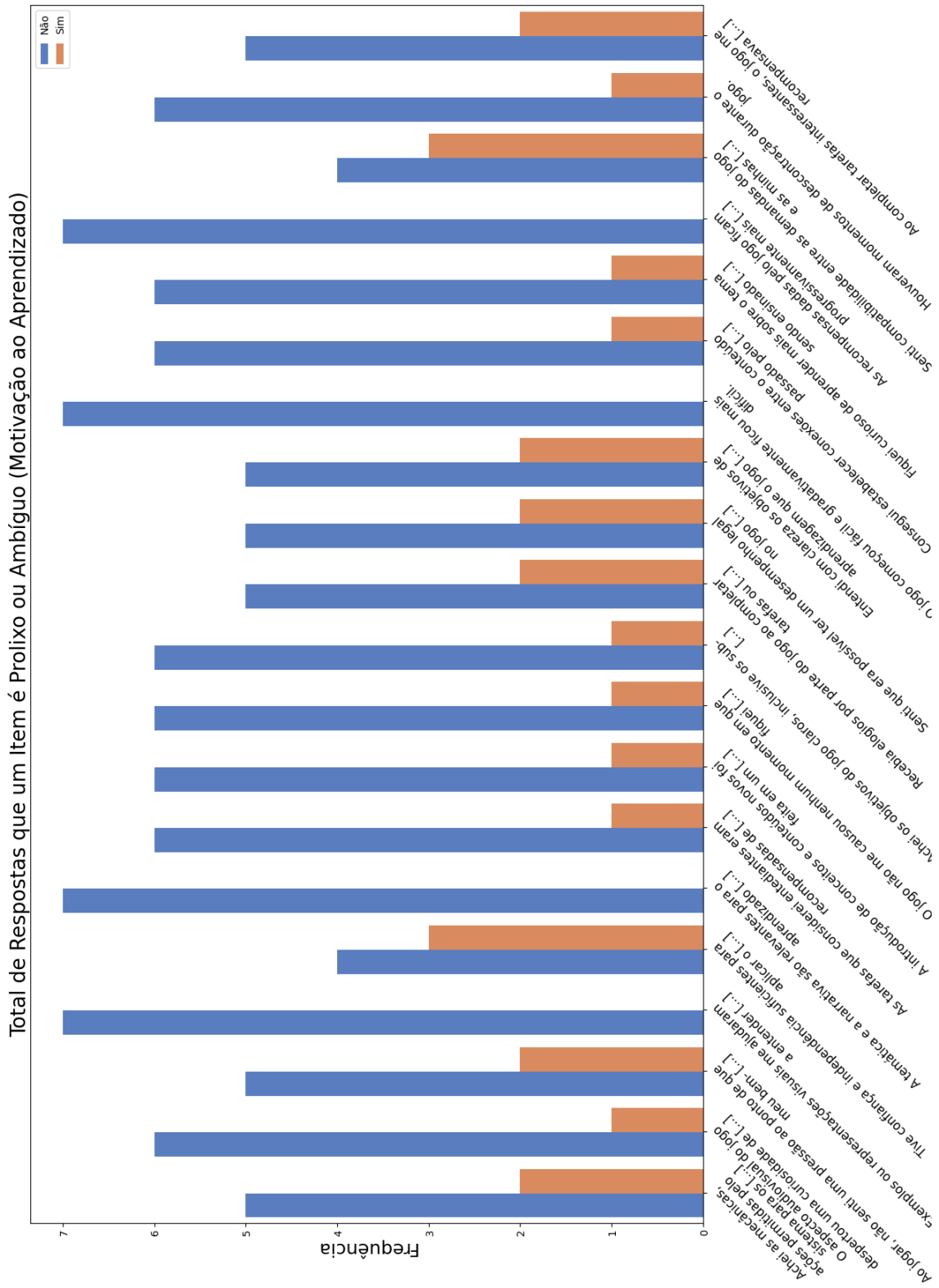


Figura D.14: Distância ao Consenso para a área de avaliação de *Motivação ao Aprendizado*



### D.1.3 Área de Avaliação III: Motivação ao Aprendizado (Gráficos)

Figura D.15: Ambiguidade dos itens da área de avaliação de *Motivação ao Aprendizado*.



## Área de Avaliação I: Jogabilidade (Tabelas)

Tabela D.1: Grau de certeza médio do consenso sobre a ambiguidade ou prolixidade na área de jogabilidade

Critério	Item	Distância ao Consenso	Grau de Certeza Médio	Considerou ambíguo ou prolixo?
Interações Cooperativas	Gostei das interações colaborativas, quando jogadores atuam em grupo para alcançar um objetivo compartilhado, que tive com outros jogadores.	0,6	0,58	Não
Velocidade das Interações com o Sistema	O jogo respondeu às minhas ações com agilidade e velocidade.	0,0	1,00	Não
Variedade de Estratégias para Ganhar o Jogo	Pude alternar as estratégias para tentar ganhar o jogo, não me senti limitado a uma única maneira de jogar.	0,4	0,68	Não
Contexto Social	Tive conversas com colegas sobre o jogo e achei tais conversas produtivas para o meu aprendizado.	0,2	0,82	Não
Engajamento com o Personagem	Ao longo do jogo senti uma identificação com o papel que estava desempenhando.	0,0	1,00	Não
Sensação de Distorção de Tempo	Senti o tempo passar em um ritmo diferente enquanto jogava, parecia estar passando mais devagar ou mais rápido.	0,8	0,52	Não
Compartilhamento	A minha equipe compartilhou recursos e informações de forma ideal para conquistar os nossos objetivos.	0,2	0,82	Não
Competência	Senti competência em relação ao meu desempenho no jogo, tive as habilidades necessárias para jogar bem.	0,8	0,52	Não
Presença Física	Durante o jogo, eu estava mais atento e comprometido com o ambiente do jogo do que com o seu entorno.	0,2	0,82	Não
Co-presença	Apesar de estar em um ambiente virtual, tive a sensação de estar compartilhando a minha experiência de forma vívida com outros.	0,2	0,82	Não
Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,0	1,00	Não
Fiero	Senti alegria durante a minha experiência com o jogo.	0,0	1,00	Não
Relaxamento	Senti pouca ou nenhuma tensão durante o jogo.	0,6	0,58	Não
Engajamento com o Mundo do Jogo	Esqueci do mundo ao meu redor enquanto jogava, era como se eu estivesse presente no mundo do jogo.	0,4	0,68	Não
Interações Competitivas	Os momentos em que competi com outros jogadores pelo mesmo objetivo foram positivos para a minha experiência.	0,4	0,68	Não
Imprevisibilidade nas Interações com o Sistema	Foi fácil prever as respostas do jogo às minhas ações, dificilmente fiquei surpreso.	0,2	0,82	Não
Variedade nas Ações do Agente	O jogo me oferecia a todos os instantes uma quantidade desejável de ações alternativas para o meu personagem performar.	0,6	0,58	Não
Ausência de Medo	Estive consciente de que minha experiência no jogo não envolvia riscos verdadeiros, e, portanto, não senti medo ao jogar.	0,4	0,68	Não

Tabela D.2: Distância ao consenso sobre a validade de conteúdo na área de jogabilidade

Critério	Item	Distância ao Consenso	Grau de Certeza Médio	Alternativa Preferida	Valor p (Quantidade de Permutações)
Engajamento com o Personagem	Ao longo do jogo senti uma identificação com o papel que estava desempenhando.	0,73	0,54	Útil, mas não essencial	0,925 (167)
Fiero	Senti alegria durante a minha experiência com o jogo.	0,43	0,67	Útil, mas não essencial	0,475 (37)
Imprevisibilidade nas Interações com o Sistema	Foi fácil prever as respostas do jogo às minhas ações, dificilmente fiquei surpreso.	0,65	0,56	Útil, mas não essencial	0,925 (101)
Ausência de Medo	Estive consciente de que minha experiência no jogo não envolvia riscos verdadeiros, e, portanto, não senti medo ao jogar.	0,82	0,52	Útil, mas não essencial	0,989 (390)
Sensação de Distorção de Tempo	Senti o tempo passar em um ritmo diferente enquanto jogava, parecia estar passando mais devagar ou mais rápido.	0,43	0,66	Essencial	0,484 (39)
Variedade de Estratégias para Ganhar o Jogo	Pude alternar as estratégias para tentar ganhar o jogo, não me senti limitado a uma única maneira de jogar.	0,18	0,83	Útil, mas não essencial	0 (19)
Relaxamento	Senti pouca ou nenhuma tensão durante o jogo.	0,43	0,66	Útil, mas não essencial	0,484 (39)

Interações Cooperativa	Gostei das interações colaborativas, quando jogadores atuam em grupo para alcançar um objetivo compartilhado, que tive com outros jogadores.	0,43	0,66	Útil, mas não essencial	0,459 (39)
Contexto Social	Tive conversas com colegas sobre o jogo e achei tais conversas produtivas para o meu aprendizado.	0,73	0,54	Útil, mas não essencial	0,925 (167)
Co-presença	Apesar de estar em um ambiente virtual, tive a sensação de estar compartilhando a minha experiência de forma vívida com outros.	0,51	0,62	Útil, mas não essencial	0,728 (52)
Competência	Senti competência em relação ao meu desempenho no jogo, tive as habilidades necessárias para jogar bem.	0,43	0,66	Essencial	0,510 (39)
Engajamento com o Mundo do Jogo	Esqueci do mundo ao meu redor enquanto jogava, era como se eu estivesse presente no mundo do jogo.	0,18	0,83	Útil, mas não essencial	0 (19)
Variedade das Ações do Agente	O jogo me oferecia a todos os instantes uma quantidade desejável de ações alternativas para o meu personagem performar.	0,43	0,66	Útil, mas não essencial	0,510 (39)
Velocidade das Interações com o Sistema	O jogo respondeu às minhas ações com agilidade e velocidade.	0,43	0,67	Essencial	0,580 (37)
Interações Competitivas	Os momentos em que competi com outros jogadores pelo mesmo objetivo foram positivos para a minha experiência.	0,51	0,62	Útil, mas não essencial	0,651 (52)
Animação	Senti entusiasmo durante a minha experiência com o jogo.			Essencial	0,0547 (18)
Compartilhamento	A minha equipe compartilhou recursos e informações de forma ideal para conquistar os nossos objetivos.	0,51	0,62	Útil, mas não essencial	0,690 (52)
Presença Física	Durante o jogo, eu estava mais atento e comprometido com o ambiente do jogo do que com o seu entorno.	0,38	0,69	Essencial	0,336 (32)

## Área de Avaliação II: Experiência de Usuário (Tabelas)

Tabela D.3: Grau de certeza médio do consenso sobre a ambiguidade ou prolixidade na área de experiência de usuário

Critério	Item	Distância ao Consenso	Grau de Certeza Médio	Considerou ambíguo ou prolixo?
Acessibilidade para Pessoas com Deficiência Auditiva	O jogo não dependeu de narrações, efeitos, músicas ou outros efeitos sonoros para passar informações importantes ao seu andamento.	0,4	0,68	Não
Acessibilidade para Pessoas com Daltonismo	O jogo não passou informações e sinais importantes para o andamento do jogo exclusivamente por cor.	0,8	0,52	Não
Reconhecimento de Adequabilidade	Aprendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.	0,0	1,00	Não
Nitidez da Imagem	O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	0,0	1,00	Não
Ausência de Erros Disruptivos	Presenciei algum erro que interrompeu ou terminou o jogo de forma inesperada.	0,0	1,00	Não
Textos Bem Formatos	O texto estava bem formatado e contrastava com o plano de fundo, facilitando a sua leitura.	0,4	0,68	Não
Linguagem Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	0,8	0,52	Não
Volume Ajustável	É possível ajustar ou silenciar o volume do jogo, diferenciando entre os volumes dos efeitos sonoros, trilhas sonoras e narração.	0,8	0,52	Não
Proteção de Erros	O jogo apresentou erros que não afetaram a continuidade do jogo, mas levaram a comportamentos inesperados pelo sistema.	0,4	0,68	Não
Capacidade de Aprendizagem	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo um domínio dos controles, antes que se tornasse desafiador.	0,4	0,68	Não
Atratividade	Achei a interface do jogo bonita, o aspecto visual ou sonoro cativou-me de uma forma positiva.	0,8	0,52	Não

Tabela D.4: Distância ao consenso sobre a validade de conteúdo na área de experiência de usuário

Critério	Item	Distância ao Consenso	Grau de Certeza Médio	Alternativa Preferida	Valor p (Quantidade de Permutações)
Volume Ajustável	É possível ajustar ou silenciar o volume do jogo, diferenciando entre os volumes dos efeitos sonoros, trilhas sonoras e narração.	0,46	0,65	Útil, mas não essencial	0,445 (33)

Proteção de Erros	O jogo apresentou erros que não afetaram a continuidade do jogo, mas levaram a comportamentos inesperados pelo sistema.	0,59	0,58	Útil, mas não essencial	0,651 (52)
Reconhecimento de Adequabilidade	Apreendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.	0,18	0,84	Essencial	0,0547 (18)
Ausência de Erros Disruptivos	Presenciei algum erro que interrompeu ou terminou o jogo de forma inesperada.	0,56	0,60	Essencial	0,523 (50)
Acessibilidade para Pessoas com Daltonismo	O jogo não passou informações e sinais importantes para o andamento do jogo exclusivamente por cor.	0,15	0,86	Útil, mas não essencial	0 (17)
Textos Bem Formatados	O texto estava bem formatado e contrastava com o plano de fundo, facilitando a sua leitura.	0,59	0,58	Útil, mas não essencial	0,482 (50)
Capacidade de Aprendizado	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo um domínio dos controles, antes que se tornasse desafiador.	0,18	0,84	Essencial	0 (19)
Acessibilidade para Pessoas com Deficiência Auditiva	O jogo não dependeu de narrações, efeitos, músicas ou outros efeitos sonoros para passar informações importantes ao seu andamento.	0,46	0,65	Útil, mas não essencial	0,555 (43)
Nitidez da Imagem	O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	0,18	0,84	Essencial	0,0522 (19)
Linguagem Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	0,15	0,86	Essencial	0 (17)
Atratividade	Achei a interface do jogo bonita, o aspecto visual ou sonoro cativou-me de uma forma positiva.	0,18	0,84	Útil, mas não essencial	0 (19)

## Área de Avaliação III: Motivação ao Aprendizado (Tabelas)

Tabela D.5: Grau de certeza médio do consenso sobre a ambiguidade ou prolixidade na área de motivação ao aprendizado.

Critério	Item	Distância ao Consenso	Grau de Certeza Médio	Considerou ambíguo ou prolixo?
Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,57	0,59	Não
Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre o jogo.	0,29	0,76	Não
Evitar Influências Negativas	Ao jogar, não senti uma pressão ao ponto de que meu bem-estar e desempenho no jogo fossem prejudicados.	0,57	0,59	Não
Concretude	Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	0,0	1,0	Não
Autoconfiança	Tive confiança e independência suficientes para aplicar o conhecimento adquirido pelo jogo em situações realísticas.	0,86	0,51	Não
Temática Relevante	A temática e a narrativa são relevantes para o aprendizado oferecido pelo jogo.	0,0	1,0	Não
Recompensas para Tarefas Entediantes	As tarefas que considerei entediantes eram recompensadas de forma evidente e previsível.	0,29	0,76	Não
Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,29	0,76	Não
Conforto	O jogo não me causou nenhum momento em que fiquei fisicamente desconfortável.	0,29	0,76	Não
Objetivo do Jogo	Achei os objetivos do jogo claros, inclusive os sub-objetivos ou a otimização do desempenho das tarefas.	0,29	0,76	Não
Elogios a Resultados Positivos	Recebia elogios por parte do jogo ao completar tarefas ou ter um bom desempenho no jogo.	0,57	0,59	Não
Moderação de Expectativas	Senti que era possível ter um desempenho legal no jogo mediante o meu esforço.	0,57	0,59	Não
Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados para avaliar o meu desempenho.	0,57	0,59	Não
Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.			
Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,29	0,76	Não
Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,29	0,76	Não
Intervalos de Reforços Positivos	As recompensas dadas pelo jogo ficam progressivamente mais difíceis de conquistar.	0,0	1,0	Não
Demanda de Habilidade	Senti compatibilidade entre as demandas do jogo e as minhas habilidades para superá-las.	0,86	0,51	Não



Humor	Houve momentos de descontração durante o jogo.	0,29	0,76	Não
Recompensas para Tarefas Interessantes	Ao completar tarefas interessantes, o jogo me recompensava inesperadamente, variando de alguma forma na recompensa que era dada.	0,57	0,59	Não

Tabela D.6: Distância ao consenso sobre a validade de conteúdo na área de motivação ao aprendizado

Critério	Item	Distância ao Consenso	Grau de Certeza Média	Alternativa Preferida	Valor p (Quantidade de Permutações)
Humor	Houve momentos de descontração durante o jogo.	0,51	0,62	Útil, mas não essencial	0,651 (52)
Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,12	0,88	Essencial	0 (17)
Demanda de Habilidade	Senti compatibilidade entre as demandas do jogo e as minhas habilidades para superá-las.	0,50	0,63	Útil, mas não essencial	0,644 (50)
Intervalo de Reforços Positivos	As recompensas dadas pelo jogo ficam progressivamente mais difíceis de conquistar.	0,16	0,85	Útil, mas não essencial	0,0547 (18)
Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre o jogo.	0,18	0,84	Essencial	0,0522 (19)
Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,07	0,93	Essencial	0 (15)
Recompensas para Tarefas Entediantes	As tarefas que considerei entediantes eram recompensadas de forma evidente e previsível.	0,51	0,62	Útil, mas não essencial	0,613 (52)
Autoconfiança	Tive confiança e independência suficientes para aplicar o conhecimento adquirido pelo jogo em situações realísticas.	0,18	0,84	Útil, mas não essencial	0 (19)
Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,16	0,85	Essencial	0,0546 (18)
Objetivo do Jogo	Achei os objetivos do jogo claros, inclusive os sub-objetivos ou a otimização do desempenho das tarefas.	0,50	0,63	Essencial	0,704 (50)
Concretude	Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	0,16	0,85	Essencial	0 (18)
Temática Relevante	A temática e a narrativa são relevantes para o aprendizado oferecido pelo jogo.	0,16	0,85	Essencial	0(18)
Elogios a Resultados Positivos	Recebia elogios por parte do jogo ao completar tarefas ou ter um bom desempenho no jogo.	0,76	0,53	Essencial	0,954(216)
Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,16	0,85	Essencial	0 (18)
Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,12	0,88	Essencial	0(17)
Recompensas para Tarefas Interessantes	Ao completar tarefas interessantes, o jogo me recompensava inesperadamente, variando de alguma forma na recompensa que era dada.	0,51	0,62	Útil, mas não essencial	0,555 (52)
Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados para avaliar o meu desempenho.	0,12	0,88	Essencial	0(17)
Conforto	O jogo não me causou nenhum momento em que fiquei fisicamente desconfortável.	0,39	0,69	Útil, mas não essencial	0,445 (34)
Moderação de Expectativas	Senti que era possível ter um desempenho legal no jogo mediante o meu esforço.	0,16	0,85	Útil, mas não essencial	0,0547 (18)
Evitar Influências Negativas	Ao jogar, não senti uma pressão ao ponto de que meu bem-estar e desempenho no jogo fossem prejudicados.	0,46	0,65	Útil, mas não essencial	0,578 (43)

# Apêndice E

## Dados Demográficos

Figura E.1: Distribuição da área de formação dos respondentes.

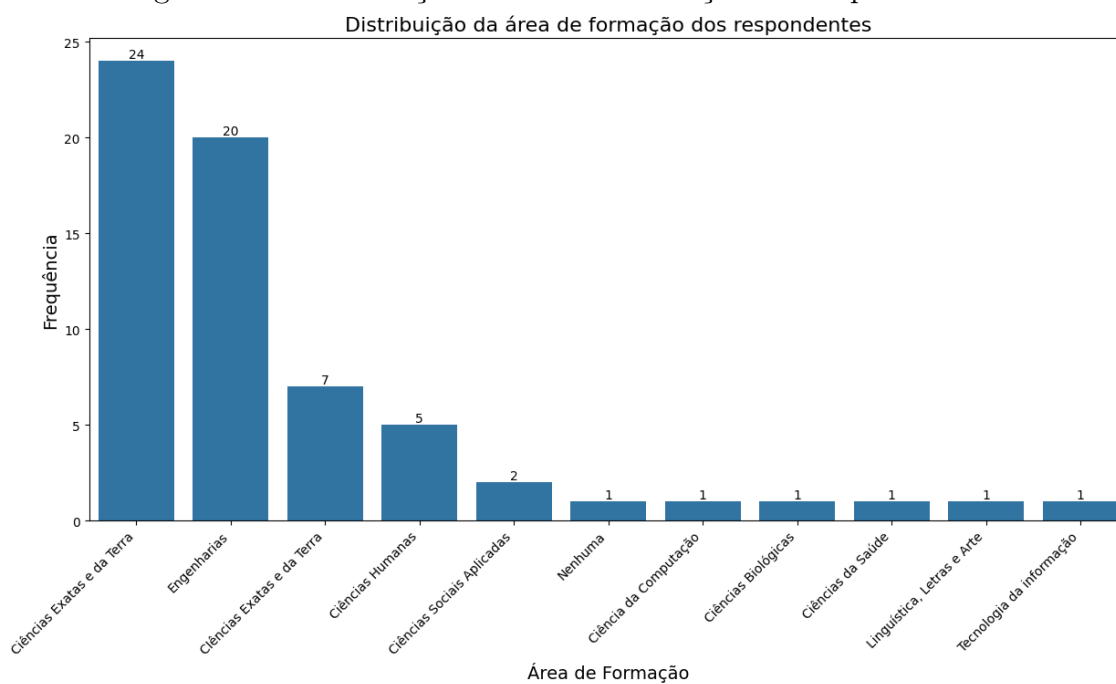


Figura E.2: Distribuição da maior titulação acadêmica dos respondentes.

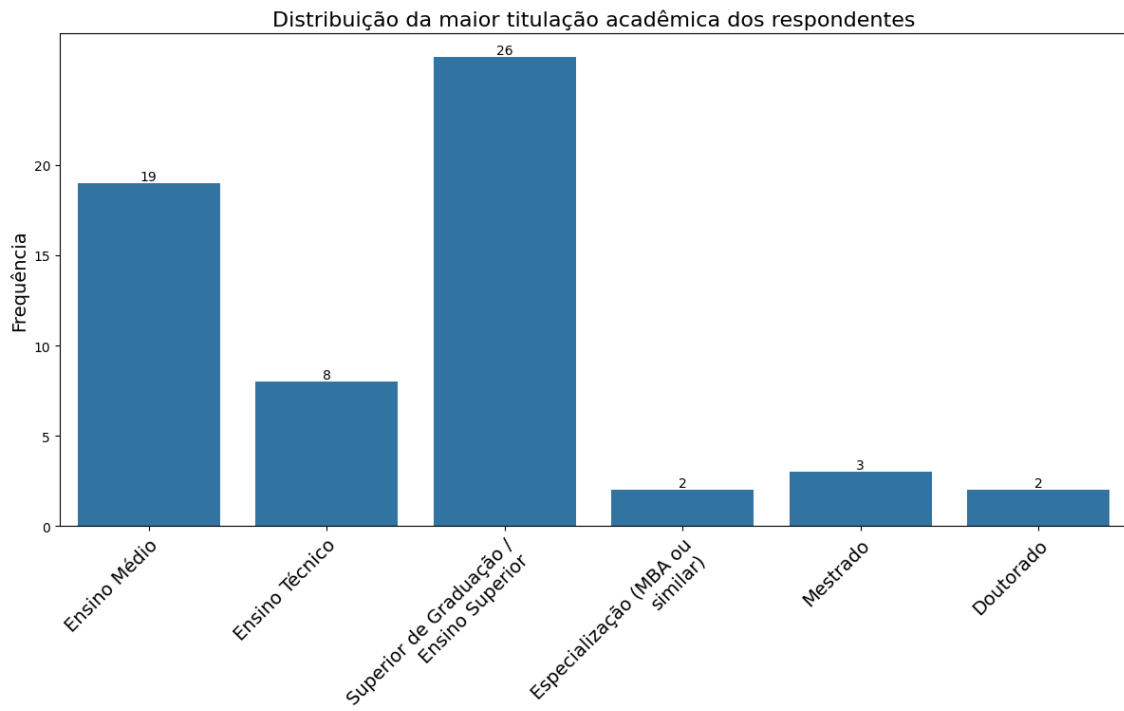


Figura E.3: Distribuição do ano de nascimento dos respondentes.

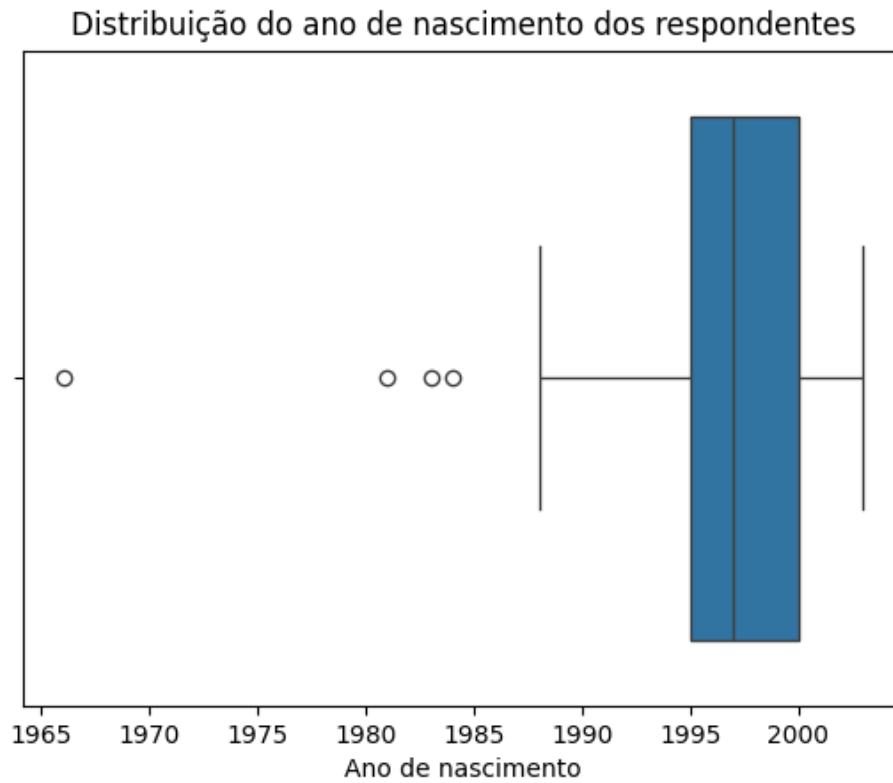


Figura E.4: Distribuição de faixa etária dos respondentes.

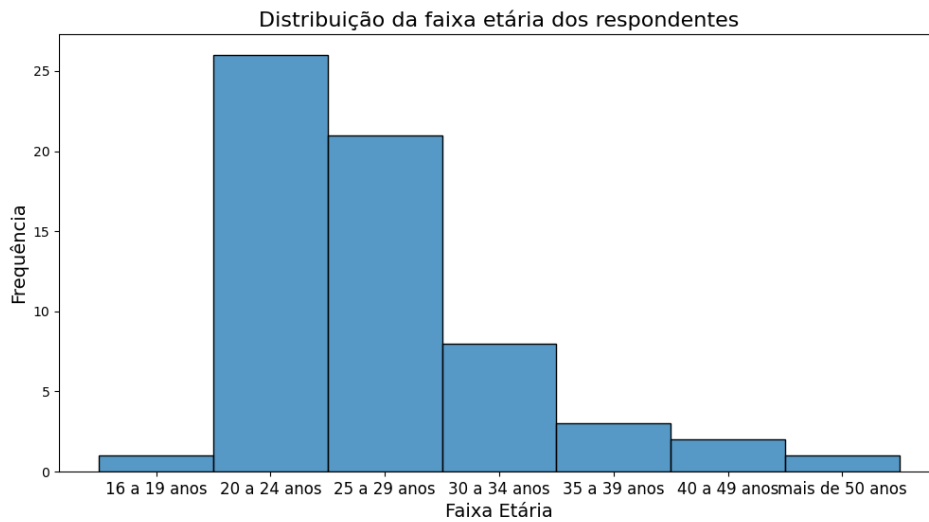


Figura E.5: Distribuição da renda mensal familiar dos respondentes.

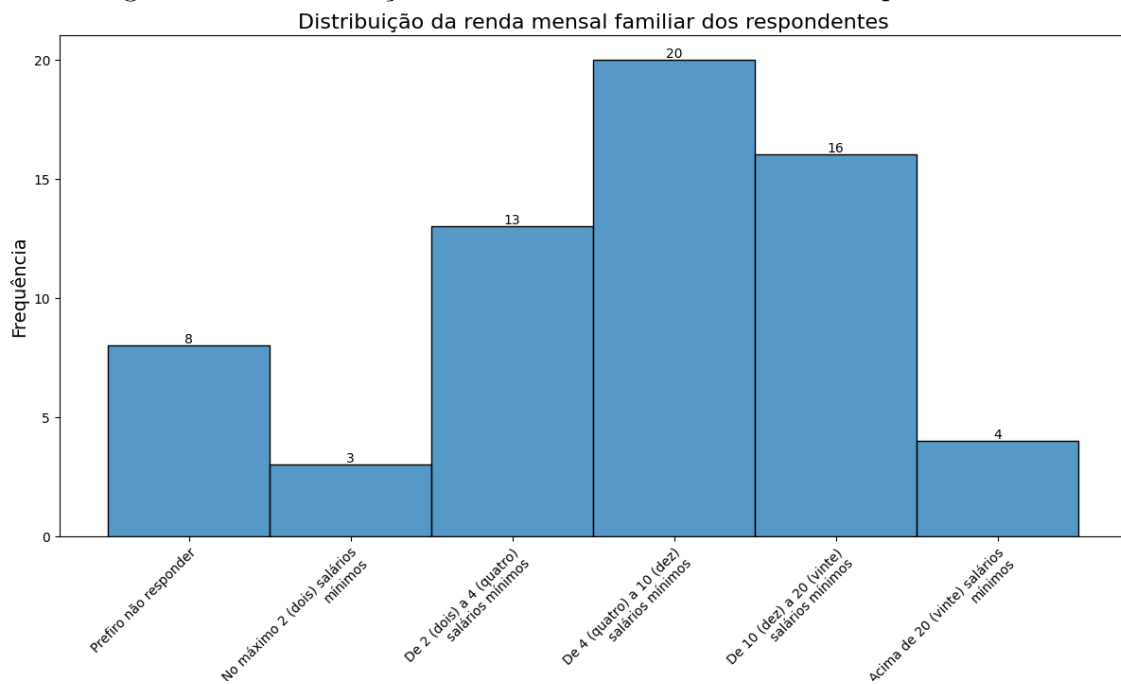
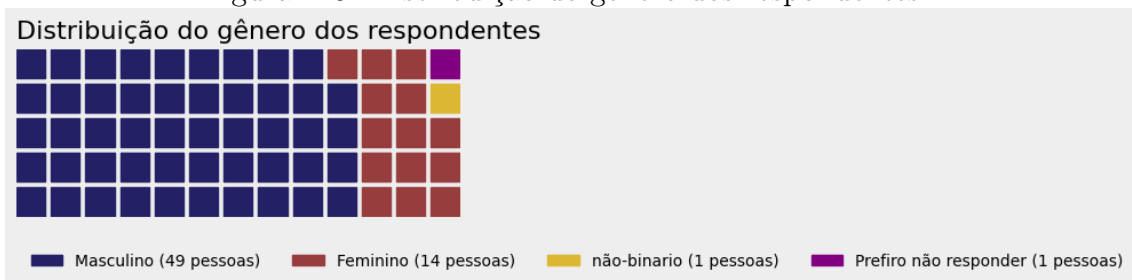


Figura E.6: Distribuição do gênero dos respondentes.



## E.1 Resultados do Teste Kruskal-Wallis Aplicado aos Dados Demográficos

Tabela E.1: Resultado do teste Kruskal-Wallis para o dado de Área de Formação (graus de liberdade = 1)

Critério	Item	Estatística H do teste Kruskal-Wallis	Valor p	Diferenças estatisticamente relevantes
Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,813	0,367	Não
Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	1,76	0,185	Não
Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,0123	0,912	Não
Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,00262	0,959	Não
Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados para avaliar o meu desempenho.	0,0659	0,797	Não
Concretude	Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	8,27	0,0403	Sim
Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,0565	0,812	Não
Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	2,86	0,0910	Não
Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,168	0,682	Não
Linguagem Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	0,171	0,679	Não
Inteligibilidade	Aprendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.	0,0289	0,865	Não
Capacidade de Aprendizado	O jogo ensinou me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,137	0,711	Não
Nitidez da Imagem	O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	4,06	0,0438	Sim

Tabela E.2: Resultado do teste Kruskal-Wallis para o dado de Faixa Etária (graus de liberdade = 2)

Critério	Item	Estatística H do teste Kruskal-Wallis	Valor p	Diferenças estatisticamente relevantes
Animação	Senti entusiasmo durante a minha experiência com o jogo.	2,82	0,244	Não
Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	2,04	0,36	Não
Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	3,48	0,175	Não
Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	5,04	0,0805	Não
Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados para avaliar o meu desempenho.	0,798	0,671	Não
Concretude	Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	10,0	0,00673	Sim
Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	2,04	0,360	Não
Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,600	0,740	Não
Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,0264	0,987	Não
Linguagem Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	0,692	0,708	Não
Inteligibilidade	Aprendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.	2,47	0,290	Não
Capacidade de Aprendizado	O jogo ensinou me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	1,49	0,474	Não
Nitidez da Imagem	O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	6,46	0,0395	Sim

Tabela E.3: Resultado do teste Kruskal-Wallis para o dado de Gênero (graus de liberdade = 1)

Critério	Item	Estatística H do teste Kruskal-Wallis	Valor p	Diferenças estatisticamente relevantes
Animação	Senti entusiasmo durante a minha experiência com o jogo.	1,41	0,236	Não
Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	1,54	0,215	Não
Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,599	0,439	Não
Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	1,18	0,277	Não
Requisitos de Aprendizado Concretude	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados pra avaliar o meu desempenho. Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	3,83	0,0504	Não
Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,0739	0,786	Não
Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	1,25	0,263	Não
Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	3,58	0,0584	Não
Linguagem Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	1,47	0,225	Não
Inteligibilidade	Aprendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.	1,43	0,232	Não
Capacidade de Aprendizado	O jogo ensinou me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,572	0,449	Não
Nitidez da Imagem	O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	0,0155	0,901	Não

Tabela E.4: Resultado do teste Kruskal-Wallis para o dado de maior titulação acadêmica (graus de liberdade = 3)

Critério	Item	Estatística H do teste Kruskal-Wallis	Valor p	Diferenças estatisticamente relevantes
Animação	Senti entusiasmo durante a minha experiência com o jogo.	1,47	0,690	Não
Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	1,19	0,754	Não
Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	4,03	0,259	Não
Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,344	0,952	Não
Requisitos de Aprendizado Concretude	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados pra avaliar o meu desempenho. Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	8,14	0,0432	Sim
Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil			Não
Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,531	0,912	Não
Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,527	0,913	Não
Linguagem Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	2,86	0,414	Não
Inteligibilidade	Aprendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.	5,97	0,113	Não
Capacidade de Aprendizado	O jogo ensinou me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.			Não
Nitidez da Imagem	O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	3,82	0,282	Não

Tabela E.5: Resultado do teste Kruskal-Wallis para o dado de raça/cor (graus de liberdade = 1)

Critério	Item	Estadística H do teste Kruskal-Wallis	Valor p	Diferenças estatística- camente relevantes
Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,0510	0,821	Não
Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,209	0,648	Não
Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,002	0,964	Não
Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,00331	0,985	Não
Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados pra avaliar o meu desempenho.	0,399	0,528	Não
Concretude	Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	0,00127	0,971	Não
Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,164	0,685	Não
Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,447	0,504	Não
Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.			Não
Linguagem Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	0,746	0,388	Não
Inteligibilidade	Aprendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.			Não
Capacidade de Aprendizado	O jogo ensinou me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	1,03	0,311	Não
Nitidez da Imagem	O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	0,167	0,682	Não

## Apêndice F

### Hábitos de Jogo dos Respondentes

Figura F.1: Distribuição da frequência que os respondentes costumam jogar.

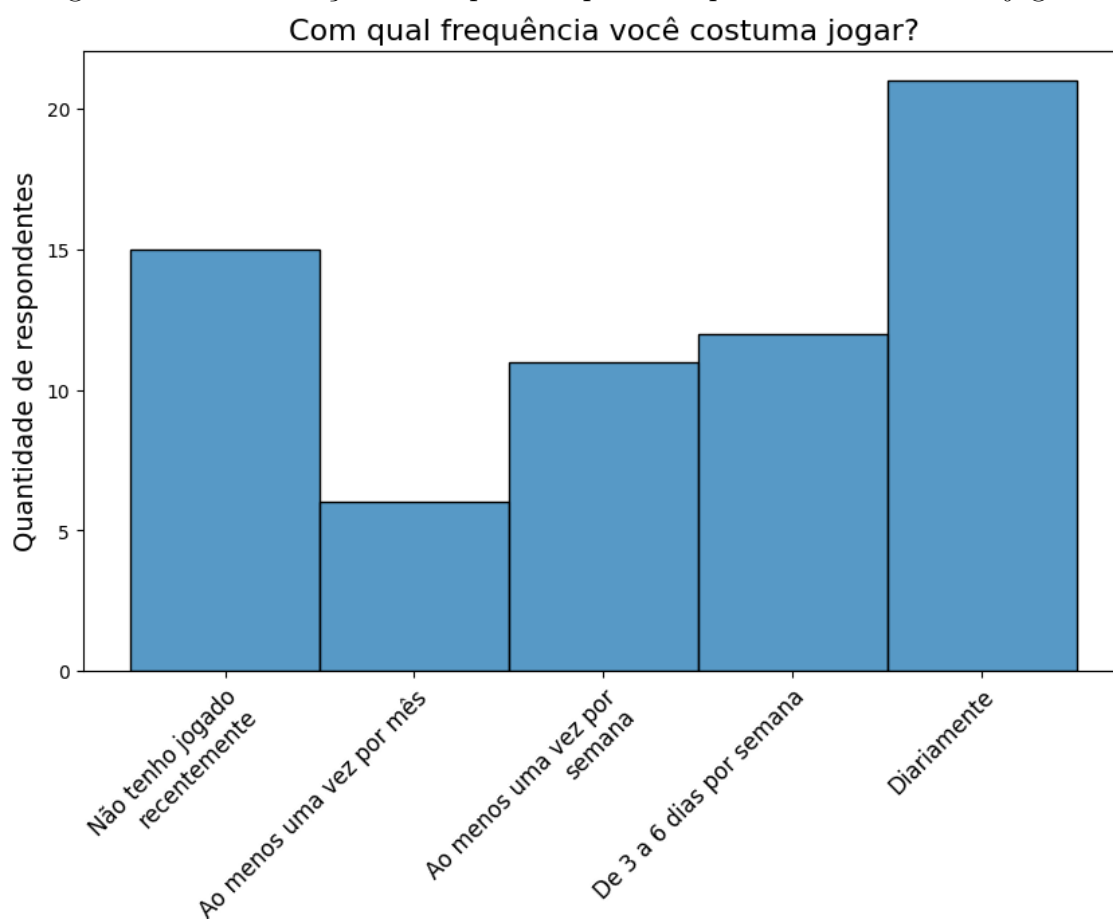




Figura F.2: Distribuição da resposta ao questionamento se jogos educacionais são divertidos.

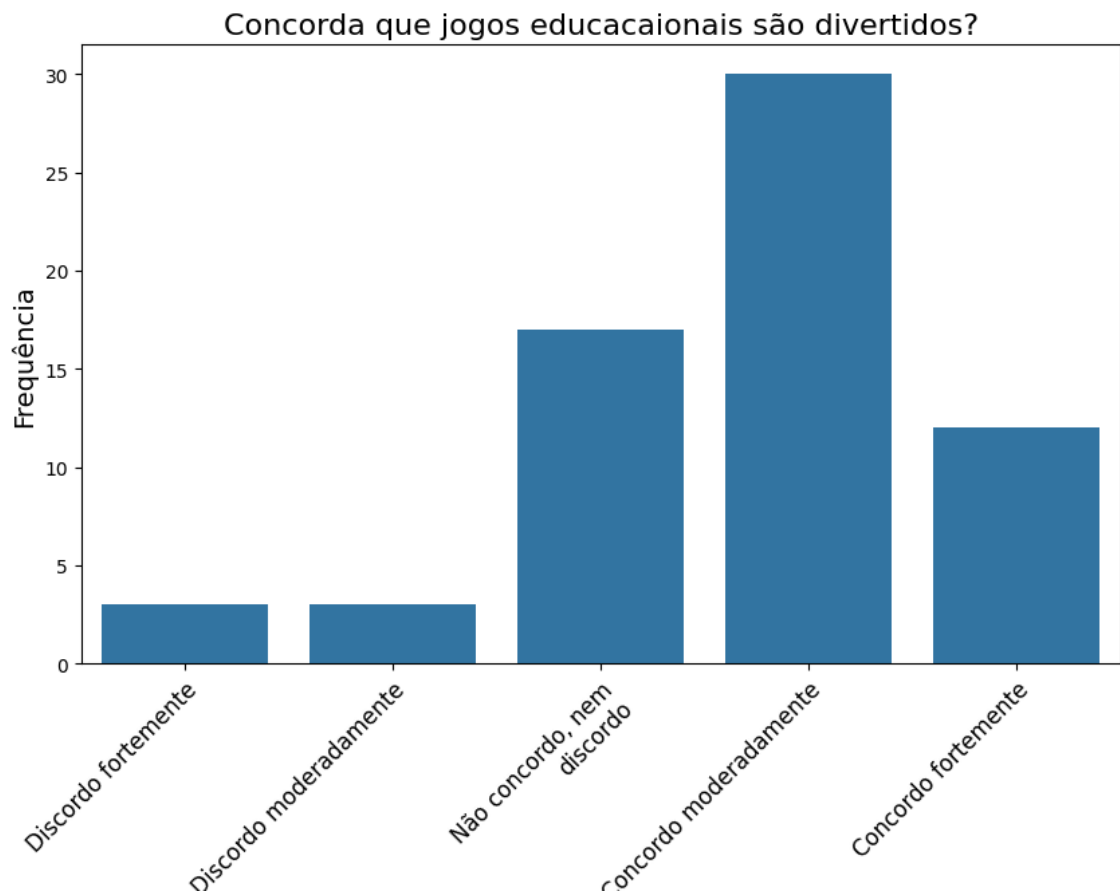


Figura F.3: Distribuição de respondentes que já haviam jogado jogos educacionais.

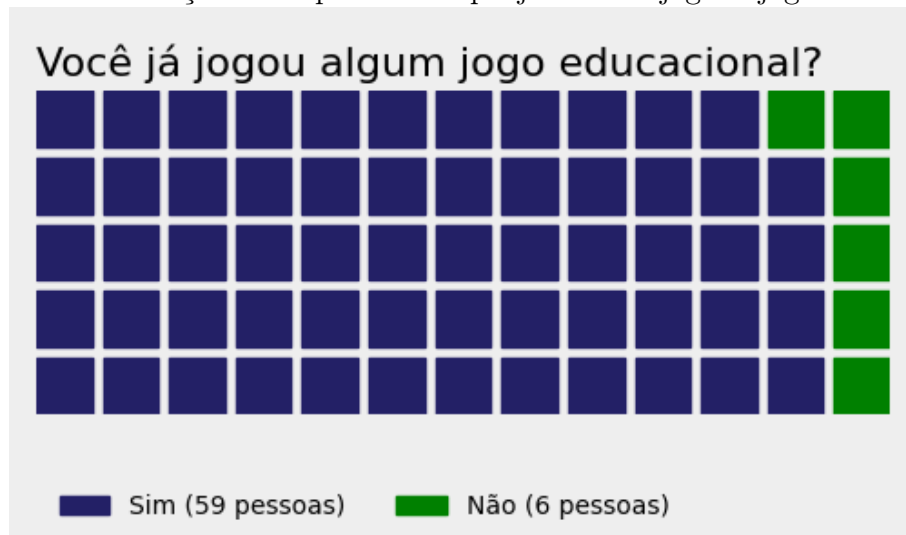
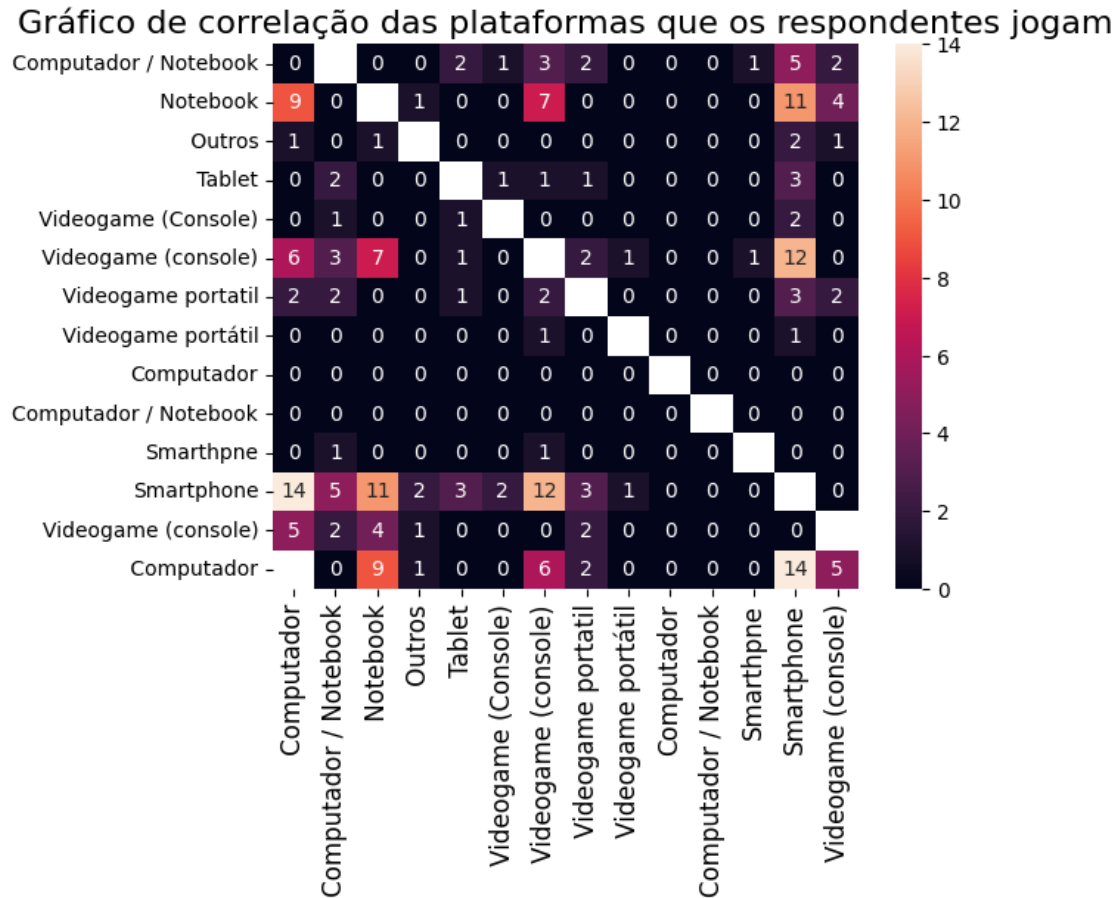


Figura F.4: Gráfico de correlação das plataformas na qual os respondentes costumam jogar.



## F.1 Resultados do Teste Kruskal-Wallis Aplicado aos Hábitos de Jogo

Tabela F.1: Resultado do teste Kruskal-Wallis para a pergunta sobre jogos educacionais serem divertidos (graus de liberdade = 2)

Critério	Item	Estatística H do teste Kruskal-Wallis	Valor p	Diferenças estatisticamente relevantes
Animação	Senti entusiasmo durante a minha experiência com o jogo.	6,92	0,0314	Sim
Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	1,27	0,530	Não
Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	4,14	0,126	Não
Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	5,46	0,0652	Não
Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados para avaliar o meu desempenho.	6,93	0,0313	Sim
Concretude	Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	2,15	0,341	Não
Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	7,52	0,0233	Sim
Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	1,34	0,512	Não

Curiosidade Sensorial		O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	9,78	0,00753	Sim
Linguagem Compreensível	Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	1,58	0,453	Não
Inteligibilidade		Aprendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.	7,64	0,0219	Sim
Capacidade de Aprendizado	de	O jogo ensinou me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	3,65	0,161	Não
Nitidez da Imagem		O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	0,953	0,621	Não

Tabela F.2: Resultado do teste Kruskal-Wallis para a pergunta se o respondente já jogou um jogo educacional (graus de liberdade = 1)

Critério	Item	Item	Estatística H do teste Kruskal-Wallis	Valor p	Diferenças estatisticamente relevantes
Animação		Senti entusiasmo durante a minha experiência com o jogo.	2,88	0,0895	Não
Mecânicas Relevantes	Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,00131	0,971	Não
Experiência		Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	5,25	0,0219	Sim
Ritmo do Jogo		A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,493	0,482	Não
Requisitos de Aprendizado	de	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados pra avaliar o meu desempenho.	1,81	0,178	Não
Concretude		Exemplos ou representações visuais me ajudaram a entender conceitos sendo ensinados pelo jogo.	1,22	0,270	Não
Dificuldade Crescente	Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,537	0,464	Não
Curiosidade Cognitiva	Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,106	0,745	Não
Curiosidade Sensorial	Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	1,83	0,177	Não
Linguagem Compreensível	Compreensível	O jogo possui uma linguagem compreensível para pessoas da minha área, houve poucos ou nenhum termo que eu não entendi.	4,23	0,0397	Sim
Inteligibilidade		Aprendi ou reforcei o aprendizado de conceitos que considero importantes para o meu ensino.	2,03	0,154	Não
Capacidade de Aprendizado	de	O jogo ensinou me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,566	0,452	Não
Nitidez da Imagem		O jogo se ajustou ao tamanho da tela do meu dispositivo, mantendo a nitidez da imagem.	0,000829	0,977	Não

# Apêndice G

## Análise Fatorial

### G.1 Análise Fatorial Exploratória

#### G.1.1 Overview

#### G.1.2 Primeira Rodada

Tabela G.1: Matriz fatorial da pesquisa no cenário de dois fatores selecionados.

Fator	Critério	Item	Carga Fatorial
F0	Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,31
	Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,50
	Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,78
	Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,47
	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,32
F1	Capacidade de Aprendizado	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,37
	Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,53
	Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,36
	Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados pra avaliar o meu desempenho.	0,88

Tabela G.2: Matriz fatorial da pesquisa no cenário de três fatores selecionados.

Fator	Critério	Item	Carga Fatorial
F0	Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,31
	Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,50
	Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,78
	Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,47
	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,32
F1	Capacidade de Aprendizado	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,37
	Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,53
	Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,36
	Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados para avaliar o meu desempenho.	0,88

F2	Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,71
	Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,35
	Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,34
	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,56

### G.1.3 Segunda Rodada

Tabela G.3: Matriz fatorial da pesquisa no cenário de dois fatores selecionados, decorrente de uma segunda rodada de análise fatorial exploratória no qual a primeira também teve dois fatores selecionados.

Fator	Critério	Item	Carga Fatorial
F0	Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,77
	Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,40
	Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,73
	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,55
F1	Aprendibilidade	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,35
	Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,73
	Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados pra avaliar o meu desempenho.	0,52
	Aprendibilidade	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,67

Tabela G.4: Matriz fatorial da pesquisa no cenário de dois fatores selecionados, decorrente de uma segunda rodada de análise fatorial exploratória no qual a primeira rodada teve três fatores selecionados.

Fator	Critério	Item	Carga Fatorial
F0	Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,82
	Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,54
	Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,61
	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,39
F1	Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,83
	Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,39
	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,50
	Aprendibilidade	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,33

Tabela G.5: Cargas fatoriais significativas para os fatores selecionados pela análise paralela na segunda rodada, no cenário no qual a primeira análise teve três fatores selecionados.

Fator	Critério	Item	Carga Fatorial
F0	Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,82
	Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,54
	Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,61
	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,39
F1	Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,83
	Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,39
	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,50
	Capacidade de Aprendizado	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,33

F2	Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,31
	Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,58
F3	Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados pra avaliar o meu desempenho.	0,73
	Capacidade de Aprendizado	O jogo ensinou me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,33
	Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,51
	Capacidade de Aprendizado	O jogo ensinou me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,75

Tabela G.6: Matriz fatorial escolhida como a ideal nesta pesquisa, rotulada e interpretada.

Fator	Critério	Item	Carga Fatorial
Jogabilidade	Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,82
	Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,54
	Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,61
Diversão	Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,83
	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,50
Scaffolding	Mecânicas Relevantes	Achei as mecânicas, ações permitidas pelo sistema para os jogadores, relevantes para os conceitos sendo ensinados.	0,31
	Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,58
	Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados pra avaliar o meu desempenho.	0,73
Capacidade de Aprendizado	-/-	O jogo ensinou me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,75

## G.1.4 Terceira Rodada

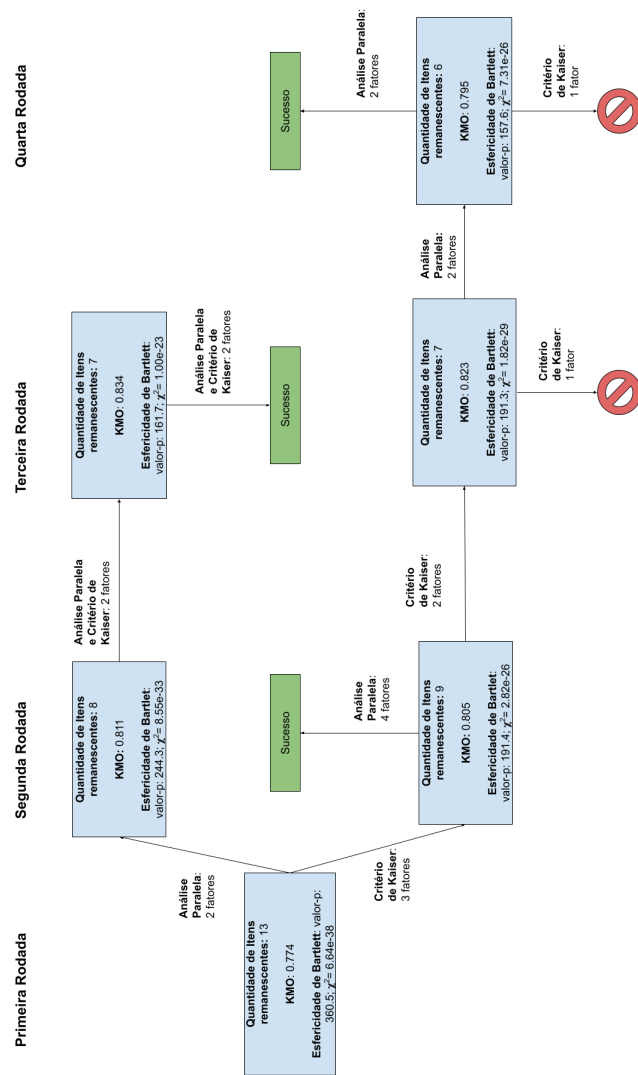
Tabela G.7: Matriz fatorial da pesquisa em um cenário de dois fatores selecionados, no qual uma matriz de dois fatores foi selecionado também na rodada inicial de análise fatorial.

Fator	Critério	Item	Carga Fatorial
F0	Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,75
	Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,46
	Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,72
F1	Curiosidade Sensorial	O aspecto audiovisual do jogo despertou uma curiosidade de querer aprender e descobrir mais sobre a temática do jogo.	0,57
	Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,76
	Requisitos de Aprendizado	Entendi com clareza os objetivos de aprendizagem que o jogo espera e os critérios que foram usados pra avaliar o meu desempenho.	0,64
	Aprendibilidade	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,42

Tabela G.8: Matriz fatorial da pesquisa em um cenário de dois fatores selecionados, no qual uma matriz de três fatores foi selecionada na rodada inicial de análise fatorial.

Fator	Critério	Item	Carga Fatorial
F0	Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,63
	Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,68
	Aprendibilidade	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,68
F1	Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,69
	Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,31
	Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,60
	Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,32

Figura G.1: Visão geral das rodadas de análise fatorial exploratória.



## G.1.5 Quarta Rodada

Tabela G.9: Matriz fatorial decorrente de uma quarta rodada de análise fatorial.

	Itens	Fatores	Carga Fatorial
F0	Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,73
	Experiência	Consegui estabelecer conexões entre o conteúdo passado pelo jogo com conceitos ou experiências anteriores.	0,71
	Aprendibilidade	O jogo ensinou-me a jogar através de um tutorial ou manual, permitindo o domínio dos controles, antes que se tornasse desafiador.	0,62
F1	Ritmo do Jogo	A introdução de conceitos e conteúdos novos foi feita em um ritmo adequado, de forma que não demandou esforço excessivo nem foi entediante.	0,77
	Dificuldade Crescente	O jogo começou fácil e gradativamente ficou mais difícil.	0,37
	Curiosidade Cognitiva	Fiquei curioso de aprender mais sobre o tema sendo ensinado por conta de algum aspecto do conteúdo do jogo.	0,61
	Animação	Senti entusiasmo durante a minha experiência com o jogo.	0,45