



PODER EXECUTIVO
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA



*Um Assistente para Apoio ao Processo de Identificação e Controle de
Ameaças à Validade de Experimentos Controlados em Engenharia de
Software*

Amadeu Anderlin Neto

Manaus

2013

Amadeu Anderlin Neto

***Um Assistente para Apoio ao Processo de Identificação e Controle de
Ameaças à Validade de Experimentos Controlados em Engenharia de
Software***

Prévia da dissertação a ser apresentada ao Programa de Pós-Graduação em Informática da Universidade Federal do Amazonas como requisito parcial para obtenção do grau de Mestre em Informática.

Área de concentração: Engenharia de *Software* Experimental.

Orientadora:

Prof^a. Dr^a. Tayana Uchôa Conte

Manaus

2013

Proposta de dissertação apresentada sob o título “Um Assistente para Apoio ao Processo de Identificação e Controle de Ameaças à Validade de Experimentos Controlados em Engenharia de *Software*”, a ser defendida por Amadeu Anderlin Neto em setembro de 2013, em Manaus, Amazonas, pela banca examinadora constituída pelos seguintes professores:

Prof^ª. D.Sc. Tayana Uchôa Conte
Instituto de Computação – UFAM
Orientadora

Prof. D.Sc. Leandro Silva Galvão de Carvalho
Instituto de Computação – UFAM

A definir

Manaus
2013

Sumário

<i>Lista de Figuras</i>	4
<i>Lista de Tabelas</i>	6
1. Introdução	7
1.1. Definição do Problema.....	7
1.2. Justificativa.....	8
1.3. Objetivos	8
1.4. Metodologia.....	8
1.5. Organização do Texto.....	10
2. Conceitos de Engenharia de <i>Software</i> Experimental	11
2.1. Estudos Qualitativos e Quantitativos	11
2.2. Estudos Primários e Secundários.....	12
2.3. Estratégias Experimentais	12
Para redução de escopo desta pesquisa, serão consideradas apenas causas de ameaças à validade de experimentos controlados.	14
2.4. Tipos de Experimentos.....	14
2.5. Tipos de Validade.....	14
2.6. Trabalhos Relacionados	16
2.6.1. Revisão Inicial da Literatura.....	16
3. Revisão Inicial da Literatura.....	20
3.1. Causas de Ameaças à Validade Interna Pré-Experimento.....	20
3.2. Causas de Ameaças à Validade Interna Em Experimento	25
3.3. Causas de Ameaças à Validade Externa	28
3.4. Relacionamentos entre Causas e Tratamentos Identificados.....	30
4. Revisão Complementar da Literatura.....	44
4.1. Resultados da RSL.....	45
4.1.1. Ameaças à Validade Interna e Ações de Controle.....	46
4.1.2. Ameaças à Validade Externa e Ações de Controle.....	62
4.1.3. Ameaças à Validade de Constructo e Ações de Controle	67
4.1.4. Ameaças à Validade de Conclusão e Ações de Controle.....	70
5. Assistente de Identificação de Causas e Tratamentos de Ameaças à Validade.....	74
<i>Referências Bibliográficas</i>	78

Lista de Figuras

<i>Figura 1.1.</i> Visão geral da metodologia experimental adotada, baseada nas propostas de Shull et al. (2001) e Mafra et al. (2006).....	9
<i>Figura 2.1.</i> Princípios de experimentos (Wöhlin et al., 2000).....	15
<i>Figura 3.1.</i> Relacionamentos identificados para tratamentos da causa C1.....	36
<i>Figura 3.2.</i> Relacionamentos identificados para tratamentos da causa C3.....	37
<i>Figura 3.3.</i> Relacionamentos para tratamento das causas C10, C17, C18, C19 e C20.....	37
<i>Figura 3.4.</i> Relacionamentos identificados para tratamentos das causas C4, C9 e C13.....	38
<i>Figura 3.5.</i> Relacionamentos identificados para tratamentos das causas C7 e C8.....	38
<i>Figura 3.6.</i> Relacionamentos identificados para tratamentos das causas C21, C22 e C27.....	39
<i>Figura 3.7.</i> Relacionamentos identificados para tratamentos das causas C25, C26 e C32.....	40
<i>Figura 3.8.</i> Relacionamentos identificados para tratamentos das causas C5, C23, C24 e C33.....	41
<i>Figura 3.9.</i> Relacionamentos identificados para tratamentos das causas C28, C29, C30 e C31.....	42
<i>Figura 3.10.</i> Relacionamentos identificados para tratamentos das causas de ameaças à validade externa.....	43
<i>Figura 4.1.</i> Ameaça INT-A01 e suas ações de controle.....	47
<i>Figura 4.2.</i> Ameaça INT-A02 e suas ações de controle.....	48
<i>Figura 4.3.</i> Ameaça INT-A03 e suas ações de controle.....	49
<i>Figura 4.4.</i> Ameaça INT-A04 e suas ações de controle.....	50
<i>Figura 4.5.</i> Ameaça INT-A05 e suas ações de controle.....	50
<i>Figura 4.6.</i> Ameaça INT-A06 e suas ações de controle.....	51
<i>Figura 4.7.</i> Ameaça INT-A07 e suas ações de controle.....	52
<i>Figura 4.8.</i> Ameaça INT-A08 e suas ações de controle.....	52
<i>Figura 4.9.</i> Ameaça INT-A10 e suas ações de controle.....	53
<i>Figura 4.10.</i> Ameaças INT-A11 e INT-A12 e suas ações de controle.....	54
<i>Figura 4.11.</i> Ameaça INT-A14 e suas ações de controle.....	55
<i>Figura 4.12.</i> Ameaça INT-A16 e suas ações de controle.....	56
<i>Figura 4.13.</i> Ameaças INT-A18, INT-A19 e INT-A20 e suas ações de controle.....	57
<i>Figura 4.14.</i> Ameaças INT-A23, INT-A24 e INT-A25 e suas ações de controle.....	58
<i>Figura 4.15.</i> Ameaças INT-A26 e INT-A27 e suas ações de controle.....	59
<i>Figura 4.16.</i> Ameaças INT-A28, INT-A29, INT-A30 e INT-A31 e suas ações de controle.....	60
<i>Figura 4.17.</i> Ameaças INT-A33 e INT-A34 e suas ações de controle.....	61
<i>Figura 4.18.</i> Ameaças INT-A35 e INT-A36 e suas ações de controle.....	62
<i>Figura 4.19.</i> Ameaça EXT-A01 e suas ações de controle.....	63

Figura 4.20. Ameaça EXT-A02 e suas ações de controle.....	64
Figura 4.21. Ameaças EXT-A03 e EXT-A04 e suas ações de controle.....	65
Figura 4.22. Ameaças EXT-A05 e EXT-A06 e suas ações de controle.....	65
Figura 4.23. Ameaças EXT-A07, EXT-A08 e EXT-A09 e suas ações de controle.....	66
Figura 4.24. Ameaças COT-A01 e COT-A02 e suas ações de controle.....	67
Figura 4.25. Ameaça COT-A03 e suas ações de controle.....	68
Figura 4.26. Ameaça COT-A04 e suas ações de controle.....	68
Figura 4.27. Ameaças COT-A05 e COT-A06 e suas ações de controle.....	69
Figura 4.28. Ameaças COT-A07, COT-A08 e COT-A09 e suas ações de controle.....	70
Figura 4.29. Ameaça COC-A01 e suas ações de controle.....	71
Figura 4.30. Ameaças COC-A02 e COC-A03 e suas ações de controle.....	72
Figura 4.31. Ameaças COC-A05, COC-A06 e COC-A08 e suas ações de controle.....	73
Figura 4.32. Ameaça COC-A07 e suas ações de controle.....	74
Figura 5.1. Tela apresentando parte da lista de causas de ameaças em experimento e tratamentos.....	75
Figura 5.2. Tela apresentando a mensagem de alerta ao pesquisador.....	76
Figura 5.3. Tela apresentando o relatório final.....	76
Figura 5.4. Diagrama de atividades do assistente.....	77

Lista de Tabelas

<i>Tabela 2.1. Comparação entre as estratégias experimentais (Wöhlin et al., 2000).....</i>	<i>13</i>
<i>Tabela 2.2. Tipos de validade x artigos.....</i>	<i>18</i>
<i>Tabela 2.3. Parte das causas e tratamentos retirados da literatura.....</i>	<i>18</i>
<i>Tabela 3.1. Lista de causas de ameaças à validade interna pré-experimento.....</i>	<i>21</i>
<i>Tabela 3.2. Lista de causas de ameaças à validade interna em experimento.....</i>	<i>26</i>
<i>Tabela 3.3. Lista de causas de ameaças à validade externa.....</i>	<i>29</i>
<i>Tabela 3.4. Lista de tratamentos das causas de ameaças à validade interna.....</i>	<i>30</i>
<i>Tabela 3.5. Lista de tratamentos das causas de ameaças à validade externa.....</i>	<i>34</i>
<i>Tabela 4.1. Resultados da RSL: Jan. 2003 – Dez. 2011.....</i>	<i>44</i>
<i>Tabela 4.2. Critérios de inclusão e exclusão.....</i>	<i>45</i>
<i>Tabela 4.3. Número de artigos que reportam ameaças à validade em experimentos controlados....</i>	<i>45</i>
<i>Tabela 4.4. Tipo de validade por ano.....</i>	<i>46</i>

1. Introdução

Estudos experimentais têm se tornado uma importante ferramenta para avaliação e validação de pesquisas em ES (Sjöberg *et al.*, 2005). Travassos *et al.* (2002) discutem que novas tecnologias não devem ser apenas sugeridas, publicadas ou apresentadas para venda, mas também devem ser comparadas com as tecnologias existentes através de experimentação, que é o centro do processo científico.

Travassos *et al.* (2008) afirmam que a experimentação em ES visa desenvolver uma base de evidência para compreensão e intervenção científica nos processos envolvidos no desenvolvimento de tecnologias de *software*. Evidências relativas ao desempenho da tecnologia, como benefícios, limitações, custos e riscos auxiliam na transferência segura da tecnologia para a indústria. Wazlawick (2008) reforça que o pesquisador provoca mudanças no ambiente a ser pesquisado para observar se cada intervenção produz os resultados esperados.

Por essa razão, o planejamento de todo experimento deve ser feito com bastante rigor. Além disso, o pesquisador deve atentar para os detalhes do experimento, pois quanto maior a intervenção, mais ameaças à validade podem ser introduzidas. De acordo com Wöhlin *et al.* (2000), ameaças à validade são potenciais riscos que podem ocorrer durante o planejamento e execução de estudos experimentais.

1.1. Definição do Problema

Grande parte dos estudos experimentais apresenta algum tipo de ameaça à validade. Portanto, é essencial que as ameaças sejam identificadas e tratadas durante o planejamento e execução do experimento. É necessário catalogar as causas de ameaças à validade que podem comprometer a execução de experimentos em Engenharia de *Software* Experimental (ESE).

Em vista disso, a questão de pesquisa a ser tratada neste trabalho é a seguinte: *Como é possível aprimorar o processo de identificação de causas e tratamentos de ameaças à validade de experimentos controlados na área de Engenharia de Software?*

Uma vez que a catalogação das causas e tratamentos de ameaças à validade esteja consolidada, será desenvolvido um assistente que irá auxiliar os pesquisadores em ES a identificar as ameaças à validade dos experimentos controlados. Além disso, o assistente irá sugerir tratamentos que foram descritos em artigos que relatavam experimentos controlados, ou seja, tratamentos que foram aplicados na prática por outros pesquisadores.

1.2. Justificativa

Para que um experimento seja validado pela comunidade científica, é necessário que todos os fatores que levaram ao resultado final estejam explícitos. A validade não pode ser medida diretamente, mas as ameaças à validade do experimento podem ser verificadas e tratadas (Shadish *et al.*, 2002). Se as ameaças à validade não forem apresentadas, o resultado do experimento pode ser rejeitado.

Uma questão fundamental a respeito de um experimento é: *Quão válidos são os resultados?* (Wöhlin *et al.*, 2000). A questão da validade deve ser considerada desde a fase de planejamento, pois quanto mais claras são as ameaças à validade do experimento, mais confiáveis são os resultados.

1.3. Objetivos

O presente trabalho tem como objetivo geral *o aprimoramento do processo de identificação e tratamento de ameaças à validade de experimentos controlados na área de ES.*

Como objetivos específicos, temos:

- Analisar e extrair causas de ameaças à validade e seus possíveis tratamentos da literatura;
- Elaborar um modelo conceitual com relacionamentos entre causas e tratamentos;
- Desenvolver um assistente que listará uma série de causas de ameaças à validade e seus possíveis tratamentos encontrados na literatura;
- Avaliar o modelo conceitual e o assistente através de estudos experimentais.

1.4. Metodologia

Kitchenham e Charters (2007) afirmam que a ES deve se basear em evidências, de forma análoga ao que é feito na Medicina, provendo meios pelos quais melhores evidências provenientes da pesquisa possam ser integradas com experiência prática e valores humanos no processo de tomada de decisão considerando o desenvolvimento e a manutenção de *software*.

Para apoiar a definição e o aprimoramento desta pesquisa, será utilizada uma adaptação da abordagem baseada em evidência para definição de novas tecnologias de *software* apresentada em Mafra *et al.* (2006). Esta abordagem é uma extensão da metodologia proposta por Shull *et al.* (2001) para introdução de tecnologias de *software* para a indústria de forma segura. A abordagem é baseada na condução de estudos experimentais como forma de determinar o que funciona ou não durante a aplicação da tecnologia proposta.

Mafrá *et al.* (2006) identificaram a necessidade de executar estudos secundários antes dos estudos primários sugeridos por Shull *et al.* (2001), para que a definição da nova tecnologia pudesse estar baseada em evidências da literatura. Assim, Mafrá *et al.* (2006) propõem a adoção de estudos secundários como atividades iniciais. Por estudo secundário, entende-se a execução de um estudo que visa identificar, avaliar e interpretar todos os resultados relevantes em um determinado tópico de pesquisa, fenômeno de interesse ou questão de pesquisa. Uma revisão sistemática é um tipo de estudo secundário (Kitchenham e Charters, 2007; Biolchini *et al.*, 2007).

A abordagem metodológica adotada possui seis etapas, conforme é apresentado na Figura 1.1. Na primeira parte é realizada a *definição inicial da tecnologia*, com a execução de três atividades preliminares: uma revisão da literatura inicial, um estudo preliminar e uma revisão complementar da literatura. Os resultados destas atividades serão utilizados como apoio para a definição da proposta.

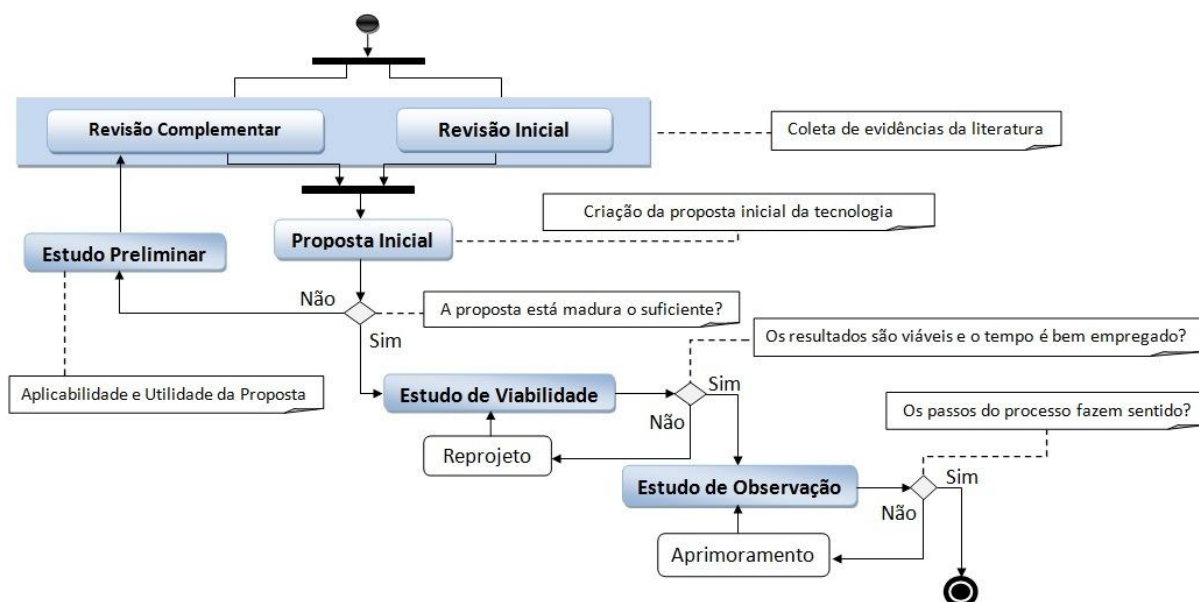


Figura 1.1. Visão geral da metodologia experimental adotada, baseada nas propostas de Shull *et al.* (2001) e Mafrá *et al.* (2006).

Através das etapas desta abordagem, pode-se acompanhar o desenvolvimento do assistente:

1. Revisão inicial da literatura: com o objetivo de identificar causas de ameaças à validade e possíveis tratamentos em experimentos controlados na área de ES. Os procedimentos utilizados para extração de dados e o resultado parcial desta revisão serão apresentados na Seção 6 do Capítulo 2;
2. Estudo preliminar: para obtenção de indícios para continuidade da pesquisa, através de um pequeno estudo experimental. Além disso, os resultados deste estudo serão utilizados como

auxílio para a definição inicial do assistente;

3. Revisão complementar da literatura: execução de uma revisão sistemática com o objetivo de complementar a base de conhecimento construída na revisão inicial da literatura. Serão buscados trabalhos mais recentes com o objetivo de identificar novas causas, tratamentos e relacionamentos de ameaças à validade;
4. Proposta inicial: criação de uma versão inicial do assistente, utilizando a base de conhecimento definida nos passos anteriores.
5. Estudo de viabilidade: para caracterizar a técnica proposta e verificar sua viabilidade de uso;
6. Estudo de observação: para aprofundar a compreensão de como o assistente é aplicado, possibilitando seu aprimoramento.

O presente trabalho encontra-se no final da etapa de revisão inicial da literatura. Posteriormente, será realizado o estudo preliminar com a participação de quatro alunos de pós-graduação que conduziram experimentos controlados recentemente. A revisão complementar da literatura será executada seguindo o protocolo disponibilizado por Kitchenham e Charters (2007). Pretende-se realizar em conjunto os estudos de viabilidade e observação. Este estudo será constituído por dois grupos: um realizará o processo de identificação de ameaças à validade e tratamentos sem o auxílio do assistente e outro com o auxílio do assistente.

1.5. Organização do Texto

O restante deste trabalho está organizado em cinco capítulos. O Capítulo 2 descreve conceitos de ESE para uma melhor compreensão do presente trabalho, os trabalhos relacionados e o processo de análise e extração de dados da revisão inicial da literatura. No Capítulo 3, são apresentadas as causas e tratamentos de ameaças à validade interna e externa identificados na revisão inicial da literatura. Além disso, os relacionamentos identificados também são mostrados. O funcionamento do assistente é descrito no Capítulo 4. Por fim, o Capítulo 5 apresenta o cronograma que esta pesquisa segue.

2. Conceitos de Engenharia de *Software* Experimental

Neste capítulo serão apresentados conceitos importantes da área de ESE para que haja uma melhor compreensão de todo o trabalho. Serão dadas as seguintes definições: (1) estudos qualitativos e quantitativos; (2) estudos primários e secundários; (3) estratégias experimentais; (4) experimentos *in vivo*, *in vitro*, *in virtuo* e *in silico*; e (5) tipos de validade. Além disso, serão mostrados alguns trabalhos relacionados a esta pesquisa e os procedimentos de análise e extração de dados na revisão inicial da literatura.

2.1. Estudos Qualitativos e Quantitativos

De acordo com a forma com que os dados são medidos, os estudos experimentais podem ser classificados em duas categorias: estudos qualitativos e estudos quantitativos.

Os estudos qualitativos estão relacionados à pesquisa sobre os objetos quando os resultados são apresentados em termos naturais (Denzin e Lincoln, 1994). Além disso, possuem caráter subjetivo e são orientados à descoberta. Métodos qualitativos vêm das ciências sociais, em oposição aos métodos quantitativos que derivam das ciências naturais. De um modo geral, as diversas visões da pesquisa qualitativa convergem no sentido de adotar uma abordagem interpretativa para o assunto em análise (subjetivismo) e uma análise do processo em estudo que não é medido em termos de quantidade, intensidade e frequência estatística.

Por outro lado, os estudos quantitativos estão relacionados com a quantificação de um relacionamento, ou com a comparação de um ou mais grupos (Creswell, 1994). Normalmente, os estudos quantitativos são conduzidos na forma de experimentos controlados, em laboratório, possuem caráter objetivo e são orientados à verificação. Estudos quantitativos são apropriados quando temos por objetivo testar o efeito de algumas manipulações ou atividades, utilizando dados quantitativos para promover a comparação e a análise estatística.

O uso de uma estratégia ou outra implica diferentes perspectivas sobre a interação do homem e seu comportamento com o objeto em estudo. A estratégia quantitativa implica a aceitação de que existe uma verdade objetiva distanciada do mundo real e que esta pode ser revelada pelo uso do método científico na medição das relações entre diferentes variáveis. A estratégia qualitativa emerge de um paradigma interpretativo e fenomenológico com ênfase em uma abordagem construtivista onde não existe uma diferenciação tão clara entre a objetividade e a realidade. O uso desta estratégia explora a construção social da realidade, aproxima a relação entre o pesquisador e o objeto de estudo e o próprio desenvolvimento da pesquisa modela seu processo.

As estratégias qualitativa e quantitativa não são competitivas entre si, mas sim complementares. A combinação entre estes métodos visando o aumento da qualidade e aplicabilidades dos estudos.

2.2. Estudos Primários e Secundários

De acordo com Conte (2009), estudos primários são conduzidos visando à caracterização de uma tecnologia em uso dentro de um contexto específico. Nessa categoria encontram-se os estudos experimentais (Wöhlin *et al.*, 2000): experimentos, estudos de caso e *surveys*. Por outro lado, estudos secundários são conduzidos visando à identificação, avaliação e interpretação dos resultados relevantes de um determinado tópico de pesquisa, fenômeno de interesse ou questão de pesquisa (Conte, 2009). Revisão sistemática é um tipo de estudo secundário (Kitchenham e Charters, 2007). Resultados obtidos por diversos estudos primários correlatos atuam como fonte de informação a ser investigada por estudos secundários (Mafra *et al.*, 2006).

2.3. Estratégias Experimentais

A seleção da estratégia experimental a ser aplicada depende dos objetivos do estudo, dos recursos disponíveis, do acesso aos indivíduos, do controle sobre as variáveis de interesse, das habilidades do investigador e dos resultados esperados.

Travassos *et al.* (2002) afirmam que as características principais usadas para diferenciar essas estratégias são o controle de execução, o controle de medição, o custo de investigação e a facilidade de repetição. Além disso, o risco está sendo considerado como uma característica importante.

A seguir, são apresentadas as principais estratégias experimentais descritas na literatura (Wöhlin *et al.*, 2000):

- *Survey* – é baseado em retrospectiva para identificar as características/opiniões de uma ampla população de indivíduos. Os dados são coletados através de questionários ou mesmo entrevistas estruturadas. Os *surveys* podem ser descritivos, explanatórios e exploratórios (Babbie, 1990). *Surveys* descritivos podem ser conduzidos para permitir que afirmações possam ser feitas para uma determinada população. Os *surveys* explanatórios visam à explicação de determinado fenômeno sobre uma população. Por fim, *surveys* exploratórios são utilizados como um pré-teste para uma investigação mais aprofundada, visando descobrir questões de grande importância.
- Estudo de caso – investiga um fenômeno contemporâneo dentro de seu contexto da vida

real, especialmente quando as fronteiras entre o fenômeno e o contexto não estão claramente evidentes. Os dados são colhidos pelo pesquisador a partir da observação da atividade de interesse.

- Experimento controlado – investigação de uma hipótese testável, onde uma ou mais variáveis independentes são manipuladas para medir seus efeitos sobre uma ou mais variáveis dependentes. Para a condução de um experimento controlado é necessária uma hipótese clara, que guiará a execução do experimento. Os experimentos controlados são apropriados para investigar diferentes aspectos, dentre os quais podemos citar: confirmação de teoria, exploração de relacionamentos e avaliação da precisão de modelos.

Estratégias de pesquisa viáveis usam vários métodos, escolhidos de tal forma que os pontos fracos de cada método são abordados através da utilização de métodos complementares (Creswell, 2002). Apesar de existir diversas estratégias experimentais, todas apresentam falhas conhecidas. Cada uma fornece elementos de prova, limitados e qualificados sobre os fenômenos estudados. No entanto, cada uma é falha de forma diferente (McGrath, 1995).

A Tabela 2.1 apresenta a comparação entre as estratégias experimentais. Os *surveys* não oferecem controle nenhum sobre a execução ou medição sobre as variáveis, sendo impossível manipulá-las. Por outro lado, o nível de controle dos estudos de caso é médio, mas, ao contrário dos *surveys*, possuem o controle sobre a medição das variáveis. O maior problema do estudo de caso, segundo Travassos *et al.* (2002), é a possibilidade de fatores de confusão, ou seja, é difícil diferenciar o efeito proveniente de um fator do efeito proveniente de outro. Os experimentos controlados possuem controle sobre execução, medição sobre as variáveis e investigação, além da facilidade de repetição, porém são muito mais caros de serem realizados.

Tabela 2.1. Comparação entre as estratégias experimentais (Wöhlin *et al.*, 2000).

Fator	<i>Survey</i>	Estudo de Caso	Experimento Controlado
Controle da execução	Nenhum	Nenhum	Tem
Controle da medição	Nenhum	Tem	Tem
Controle da investigação	Baixo	Médio	Alto
Facilidade de repetição	Alta	Baixa	Alta
Custo	Baixo	Médio	Alto

Para redução de escopo desta pesquisa, serão consideradas apenas causas de ameaças à validade de experimentos controlados.

2.4. Tipos de Experimentos

Ambientes de experimentação em ES buscam combinar processos de experimentação e de empacotamento de experimentos, padrões de representação de dados, gerência de conhecimento, caracterização de ferramentas, serviços, modelos computacionais de qualidade e modelos necessários para cada uma dos quatro tipos de experimentos (Travassos e Barros, 2003):

- *In vivo*: esses experimentos envolvem pessoas em seus ambientes, como, por exemplo, em organizações de desenvolvimento de *software* durante a execução de um processo de desenvolvimento real;
- *In vitro*: tais estudos são executados em um ambiente controlado, como um laboratório ou uma comunidade controlada;
- *In virtuo*: esses experimentos envolvem a interação entre participantes e modelos computacionais da realidade. O comportamento do ambiente com o qual os participantes interagem é descrito como um modelo e representado por um programa computacional;
- *In silico*: em tais estudos, tanto os participantes quanto o mundo real são descritos como modelos computacionais. O ambiente é completamente composto por modelos numéricos para os quais a interação humana não é permitida.

2.5. Tipos de Validade

De acordo com Wöhlin *et al.* (2000), as ameaças à validade de um estudo podem ser classificadas em quatro categorias:

1. A validade de constructo considera os relacionamentos entre a teoria e a observação, ou seja, se o tratamento reflete corretamente a construção da causa e se o resultado reflete corretamente a construção do efeito. Exemplo: período em que um aluno está no curso de Ciência da Computação para representar a experiência em programação (baixa validade de constructo). Para obter uma alta validade de constructo, pode-se utilizar a quantidade de meses de prática.
2. A validade de conclusão diz respeito à relação entre o tratamento e o resultado, em termos de significância estatística. É relacionada à habilidade de chegar a uma conclusão correta a respeito dos relacionamentos entre o tratamento e o resultado do experimento. Exemplo: a

escolha do teste estatístico, a escolha do tamanho da amostra, a confiabilidade das medidas e a confiabilidade da implementação dos tratamentos.

3. A validade interna observa se o relacionamento entre o tratamento e o resultado é causal, sem influência de outro fator que não é controlado ou medido. Em outras palavras, deve-se verificar se o tratamento causa o resultado (efeito). Exemplo: participantes que abandonam o experimento; o desempenho do participante pode ser afetado negativamente através de cansaço.
4. A validade externa diz respeito à generalização dos resultados. Deve-se verificar se existe um relacionamento causal entre a causa e o efeito, e se o resultado pode ser generalizado para fora do escopo do estudo. Além disso, verifica se existe uma relação entre o tratamento e o resultado. Exemplo: a população escolhida pode não representar a que queremos generalizar; o material experimental pode não representar a prática industrial.

Este documento apresentará apenas causas e tratamentos de ameaças à validade interna e externa. As demais serão abordadas ao longo da pesquisa de mestrado.

A Figura 2.1 indica onde os tipos de validade podem sofrer ameaças. Ao verificar a validade de constructo, pode-se fazer as seguintes perguntas: A causa está bem relacionada com o tratamento? O efeito está bem relacionado com o resultado? Existe coerência entre as métricas e a intenção do pesquisador?

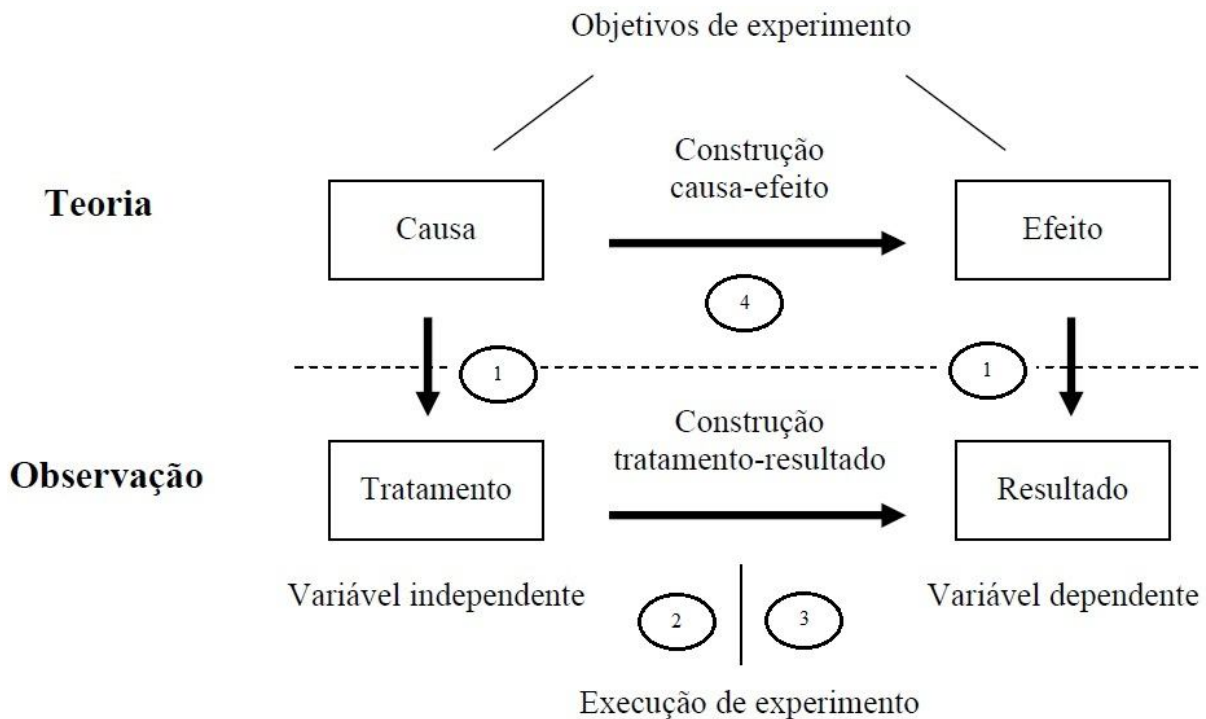


Figura 2.1. Princípios de experimentos (Wöhlin et al., 2000).

Para a validade conclusão, as seguintes perguntas podem ser formuladas: Existe uma relação estatística justificável de causa e efeito com determinada significância? O estudo experimental é repetível? O processo de análise dos resultados usado é válido?

No caso da validade externa, a pergunta é a seguinte: É possível generalizar a conclusão para fora do contexto de realização do experimento? Para a validade interna, temos: As relações causais estudadas são coerentes com o modelo? Existe efetivamente uma relação causal? As respostas para estas perguntas são um guia para manter a validade do experimento.

2.6. Trabalhos Relacionados

Inicialmente, temos como trabalho relacionado o *survey* apresentado em Sjøberg *et al.* (2005). Este *survey* relata como os experimentos controlados em ES são conduzidos e como informações relevantes são reportadas. Um total de 103 artigos foi selecionado, os quais foram publicados em nove *Journals* e três conferências da área de ES. Os tópicos dos experimentos e seus participantes foram caracterizados (número de participantes, alunos *versus* profissionais, recrutamento e as recompensas de participação), tarefas (tipo de tarefa, duração, tipo e tamanho da aplicação) e ambientes (localização, desenvolvimento de ferramentas). Além disso, o *survey* reporta como as validades interna e externa são atendidas. Foram listados nove tipos de ameaças à validade interna e oito tipos de ameaças à validade externa.

Em Feldt e Magazinius (2010), outro *survey* é realizado com o propósito de investigar como as ameaças à validade são analisadas e mitigadas em estudos na área de ESE. Com isso, são apresentados *guidelines* de como realizar as análises e mitigações para que a validade dos resultados seja maximizada. Foram analisados 43 artigos publicados na conferência *Empirical Software Engineering and Measurement* (ESEM) do ano de 2009. Em todos os artigos foram identificadas um total de 136 ameaças à validade dentre todos os tipos e 69 tratamentos.

Barros e Dias Neto (2011) apresentam uma lista de ameaças à validade em Engenharia de Software Baseada em Evidência (ESBE). Esta lista contém quatro ameaças à validade interna, quatro ameaças à validade de conclusão, três ameaças à validade externa e três ameaças à validade de constructo. Foram analisados 23 artigos publicados no *Symposium on Search-Based Software Engineering* (SSBSE) dos anos de 2009 e 2010.

2.6.1. Revisão Inicial da Literatura

Como base para revisão inicial da literatura, foi utilizada a Revisão Sistemática (RS) executada por Kampenes *et al.* (2009), cujo objetivo era investigar a qualidade de experimentos publicados em ES.

De acordo com Conte (2009), RSs são baseadas em uma estratégia de pesquisa bem definida, que visa detectar o máximo possível de material bibliográfico relevante. É definido um protocolo de revisão que especifica a questão central de pesquisa e os métodos que serão utilizados para executar a revisão. Uma RS propõe uma avaliação justa do tópico de pesquisa à medida que utiliza uma metodologia de revisão rigorosa, confiável e passível de auditoria (Kitchenham e Charters, 2007).

Kampenes *et al.* (2009) analisaram 103 artigos publicados em nove *Journals* e três conferências da área de ES. Ao final, concluíram que a forma com que os experimentos eram descritos apresentava baixa qualidade. Uma série de melhorias foi proposta, dentre elas está “*uma melhor descrição de ameaças à validade, com foco para as ameaças atuais. Ameaças que não são problema para o experimento podem ser omitidas*”.

O objetivo ao utilizar a RS de Kampenes *et al.* (2009) foi de identificar os trabalhos que reportavam experimentos controlados e extrair as causas de ameaças à validade e seus possíveis tratamentos. Desse modo, foi construída uma base inicial para elaboração de um modelo conceitual que apresenta relacionamentos entre as causas e os tratamentos de ameaças à validade.

Os 103 artigos apresentados na RS de Kampenes *et al.* (2009) foram lidos e analisados com o objetivo de identificar quais apresentavam ameaças à validade. Vale ressaltar que existiam experimentos que eram reportados em mais de um artigo. Desse modo as duplicatas foram retiradas de forma que o artigo mais atual era selecionado. Além disso, três artigos da lista dos 103 não estavam disponíveis para *download*. Ao final da leitura e análise de todos os artigos listados na RS de Kampenes *et al.* (2009), 46 artigos tiveram ameaças à validade identificadas, totalizando 55 experimentos. A Tabela 2.2 apresenta os tipos de validade que foram reportados pelos autores e seus respectivos artigos.

Em relação aos tipos de validade, foram identificadas 33 causas de ameaças à validade interna, o que representa 72% de todas as causas identificadas, e 53 tratamentos. Para as ameaças à validade externa foram encontradas 6 causas, o que representa 13% das causas identificadas, e 13 tratamentos. Finalizando, 5 causas foram identificadas para ameaças à validade de constructo, representando 11% do total, e 3 causas para ameaças à validade de conclusão, 6% do total. Exemplos dos dados extraídos dos artigos selecionados são mostrados na Tabela 2.3.

Tabela 2.2. Tipos de validade x artigos.

	Artigos																								
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
Interna	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Externa	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x
Constructo					x	x	x								x										
Conclusão															x										

	Artigos																								
	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
Interna	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Externa	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Constructo					x										x			x		x	x				
Conclusão															x			x		x	x				

Tabela 2.3. Parte das causas e tratamentos retirados de [EP9] e [EP10].

Autores	Título	Causa	Tratamento
Dunsmore et al. (2002)	Further investigations into the development and evaluation of reading techniques for object-oriented code inspection.	Nível de habilidade dos participantes desequilibrado; Experimento com longa duração; Participantes adquirem aprendizado durante o experimento.	Questionário para caracterização da habilidade e Divisão dos participantes em blocos de nível de habilidade; Informar o contexto da pesquisa para envolver os participantes; Utilização de um tratamento apenas.
Freimut et al. (2001)	Investigating the impact of reading techniques on the accuracy on different defect content estimation techniques.	Nível de habilidade dos participantes desequilibrado; Forma de atribuição dos participantes aos tratamentos; Utilização do tratamento é feita de forma diferente da ideal pelos participantes	Divisão dos participantes em blocos de nível de habilidade e formação de grupos aleatória; Atribuição aleatória aos tratamentos; Análise dos dados para possível descarte.

O fato de poucas causas de ameaças à validade externa, de constructo e de conclusão terem sido identificadas nos fornece indícios para realização de uma revisão complementar da literatura. Isso também deve ser feito pelo fato de que a RS de Kampenes *et al.* (2009) compreende trabalhos publicados entre os anos de 1993 a 2002, necessitando de uma busca mais atual, pois houve um crescimento significativo de publicações que reportam experimentos controlados.

3. Revisão Inicial da Literatura

Um dos objetivos específicos do presente trabalho é a identificação de causas e tratamentos de ameaças à validade de experimentos controlados. O tipo de ameaça não é o foco do trabalho, mas sim o que pode causar uma ameaça à validade. Desse modo, conforme apresentado na Seção 6 do Capítulo 2, foram identificadas 33 causas de ameaças à validade interna e 53 tratamentos aplicados e 6 causas de ameaças à validade externa e 13 tratamentos.

As causas, os tratamentos e os relacionamentos identificados eram validados em reuniões entre dois pesquisadores. Havia discussões sobre o motivo pelo qual uma causa, tratamento ou relacionamento era listado e, ao estabelecer um consenso, o item em questão era mantido ou descartado. Os relacionamentos identificados ocorrem entre causas e tratamentos, onde ao selecionar o tratamento X para minimizar a causa Y, outra causa Z pode ser potencializada.

Durante as reuniões, também ficou decidido que as causas das ameaças à validade interna seriam separadas em dois grupos: (1) causas que podem influenciar antes do experimento ser executado e (2) durante a execução do experimento. Nas próximas seções serão apresentadas as causas e tratamentos das ameaças à validade e os relacionamentos identificados.

3.1. *Causas de Ameaças à Validade Interna Pré-Experimento*

As causas de ameaças à validade interna que podem ocorrer antes do experimento e as evidências em que foram encontradas são listadas na Tabela 3.1. A causa C1 está relacionada à experiência do participante do estudo. Isto quer dizer que o tempo de prática do participante em determinado assunto deve ser analisado. O tempo de prática de uma pessoa em programação orientada a objetos, por exemplo, num experimento que queira avaliar um novo compilador para uma linguagem que atende a este paradigma.

A causa C2 está relacionada ao perfil dos participantes selecionados para execução do experimento. Existem procedimentos que apresentam melhores resultados se uma pessoa com certo perfil utilizá-lo. Como exemplo, temos um experimento que irá comparar dois tipos de projetos de *software*. O primeiro é indicado a projetistas novatos e possui resultados que comprovam seu alto desempenho. Por outro lado, o segundo requer maiores habilidades. Sendo os participantes do experimento alunos de graduação, a utilização do primeiro tratamento se torna favorável ao perfil dos participantes selecionados.

A causa C3 trata da habilidade dos participantes, pois pode existir certo desequilíbrio entre as habilidades de cada um. Um exemplo é a execução de um experimento com os tratamentos A e

B, sendo que os participantes que utilizam o tratamento A possuem melhores habilidades que os do tratamento B. Sendo assim, os resultados do tratamento A possuem grandes chances de serem melhores que os do tratamento B.

Tabela 3.1. Lista de causas de ameaças à validade interna pré-experimento.

Causa	Evidência(s)
C1. Tempo de experiência do participante do estudo em um determinado assunto/tema.	EP1, EP7, EP20, EP28, EP29, EP32, EP36, EP38.
C2. Perfil dos participantes favorece o uso do tratamento.	EP1.
C3. Nível de habilidade dos participantes desequilibrado.	EP1, EP4, EP6, EP9, EP10, EP12, EP16, EP17, EP20, EP21, EP23, EP24, EP25, EP26, EP27, EP28, EP29, EP30, EP31, EP32, EP33, EP35, EP36, EP38, EP39, EP40, EP41, EP42, EP43, EP44, EP45, EP46.
C4. Material experimental pode conter problemas relacionados à escrita.	EP1, EP25, EP46.
C5. Participantes podem ter familiaridade com os instrumentos utilizados.	EP2.
C6. Forma de atribuição dos participantes aos tratamentos.	EP2, EP4, EP10, EP20, EP24, EP27, EP30, EP31, EP32, EP33, EP35, EP37, EP38, EP40, EP45.
C7. Diferenças entre os instrumentos de dados.	EP5, EP6, EP7, EP12, EP16, EP20, EP21, EP23, EP24, EP25, EP29, EP30, EP31, EP32, EP33, EP34, EP38, EP39, EP41, EP42.
C8. Diferenças entre as tarefas experimentais.	EP5, EP6, EP25, EP28.
C9. Tarefas/instrumentos mal especificados para aplicação do tratamento.	EP5, EP6.
C10. Receio dos participantes em realizar o experimento.	EP7.
C11. Realização de uma atividade piloto.	EP8.

Causa	Evidência(s)
C12. Treinamento insuficiente para nivelamento de habilidade dos participantes.	EP11, EP25.
C13. Elaboração do material experimental feita por não especialistas.	EP14, EP25.
C14. Diferenças no ambiente de realização do experimento.	EP15.
C15. Aplicação de procedimentos experimentais diferentes entre os participantes antes do experimento.	EP15.
C16. Participantes a favor ou contra a utilização do tratamento.	EP22.
C17. Replicação do experimento mal executada.	EP25, EP31, EP32.
C18. Tempo estabelecido para término do experimento.	EP25, EP28, EP35, EP43.
C19. Falta de comprometimento do participante.	EP38, EP42.
C20. Idioma do material experimental diferente do idioma dos participantes.	EP42.

A causa C4 apresenta a importância de um material experimental bem elaborado. Problemas relacionados à escrita podem estar presentes no material e podem acabar causando certa dificuldade de compreensão por parte dos participantes. Erros de escrita, omissão de informação, ambiguidade e inconsistência são problemas que devem ser evitados.

Pode existir uma familiaridade dos participantes com os instrumentos utilizados, a causa C5 trata desse aspecto. Durante a elaboração dos instrumentos que serão utilizados na execução do experimento, deve-se atentar ao fato de que os participantes podem estar familiarizados com determinado instrumento, pois faz uso deste no ambiente acadêmico ou profissional. Dessa forma, o participante ao utilizar o documento com o qual está familiarizado, pode ter desempenho melhor em comparação ao uso do documento em que não há familiaridade.

A causa C6 diz respeito à forma de atribuição dos participantes aos tratamentos. Durante a execução do experimento, quais participantes utilizarão o tratamento X e quais utilizarão o tratamento Y? Voltando ao exemplo do compilador para uma linguagem orientada a objetos, quais

participantes irão desenvolver programas que serão analisados pelo novo compilador? E quais serão analisados pelo compilador antigo? Nesse caso, os tratamentos a serem utilizados devem ser atribuídos aos participantes de forma que não influenciem os resultados.

As diferenças entre os instrumentos de dados são retratadas como a causa C7. Por instrumentos de dados entende-se todo instrumento que, quando utilizado pelos participantes, forneça dados para cálculo das variáveis do experimento. Por exemplo, temos um experimento que deve avaliar uma nova técnica de inspeção de documentos. Deve-se utilizar os documentos de inspeção A e B, onde o participante deve encontrar defeitos através da utilização de uma técnica de inspeção. Uma parte dos participantes utiliza o documento A e a técnica X e outra parte utiliza o documento B e a técnica Y. A diferença entre os documentos A e B deve ser mínima para que os resultados do experimento não sejam influenciados pela utilização dos documentos, mas sim da técnica.

Utilizando raciocínio semelhante ao apresentado na causa C7, a causa C8 trata das diferenças entre as tarefas experimentais. A escolha do tipo, do tamanho e complexidade da tarefa é uma etapa crítica para o pesquisador responsável por um experimento, pois se forem atribuídas tarefas diferentes entre os participantes, estas devem apresentar certo grau de equivalência ou compensação. Como exemplo, temos a utilização da técnica de inspeção A e B e tarefas específicas a serem executadas para cada uma. No caso de diferença entre as tarefas, o desempenho dos participantes pode variar, causando influência sobre o resultado do experimento.

A causa C9 diz respeito às tarefas e instrumentos mal especificados. Pode ocorrer quando a tarefa ou instrumento é elaborado por uma pessoa não especializada no assunto abordado. Dessa forma, as tarefas podem não compreender o domínio do problema e, ao final do experimento, o resultado não apresenta conformidade com o objetivo do estudo. Por sua vez, os instrumentos podem conter falhas de especificação. Por esse motivo, as tarefas e os instrumentos utilizados podem apresentar alguma deficiência.

Uma vez que os participantes sejam funcionários de uma determinada empresa, existe certo receio que os impede de fornecer informações de forma exata. Isto acontece porque as informações fornecidas podem comprometê-los na organização. Ao saber que está sendo avaliado, o participante pode não se comportar da forma com que está acostumado. Nesses casos, os pesquisadores devem analisar cuidadosamente as informações obtidas, pois podem não refletir a realidade.

A causa C11 trata da realização de uma atividade piloto. Na maioria dos casos, as atividades piloto são executadas em amostras pequenas com o objetivo de avaliar o material experimental antes de aplicá-lo numa amostra com maior significância estatística. Além disso, pode haver uma

atividade piloto com os mesmos participantes, a fim de nivelamento de conhecimento. Com isso, deve-se atentar ao aprendizado que a atividade possa fornecer aos participantes para que não comprometa a execução do experimento principal.

A causa C12 aborda os casos em que o treinamento para nivelamento das habilidades dos participantes pode falhar. Em alguns casos, o treinamento pode não conter a explicação completa de todos os tratamentos ou da teoria necessária para entendimento dos procedimentos experimentais. Além disso, podem existir participantes que não compreendam todo o procedimento experimental que deve ser seguido. Dessa forma, a execução do experimento pode ser comprometida.

A causa C13 diz respeito à elaboração do material experimental, pois podem existir casos em que a construção do material seja feita por uma pessoa que não possui habilidades especializadas no tema abordado pelo experimento. Com isso o material experimental pode apresentar problemas, como por exemplo, tarefas inadequadas para o experimento.

Experimentos que apresentam diferenças entre os ambientes de realização do experimento são compreendidos pela causa C14. Ao tentar uma comparação entre estudantes e profissionais da indústria, por exemplo, o ambiente em sala de aula se torna bem diferente do ambiente de trabalho. Mesmo tentando um cenário fiel ao encontrado na indústria, podem existir fatores capazes de influenciar o resultado do experimento.

A causa C15 está relacionada à utilização de diferentes procedimentos experimentais entre os participantes do experimento. Isso pode ocorrer durante um treinamento, por exemplo, onde para uma parte dos participantes é fornecido um documento escrito que contém explicação do experimento e para outra parte é realizada uma apresentação, onde dúvidas podem ser esclarecidas. Desse modo, a parte dos participantes que recebeu apenas o documento escrito pode apresentar dúvidas durante a execução do experimento, consumindo tempo do experimento, influenciando o resultado.

A causa C16 trata dos participantes que são contrários ou favoráveis à utilização do tratamento utilizado no experimento. Por exemplo, quando uma organização participa de um experimento e faz com que os funcionários utilizem um novo tratamento. Podem existir participantes que estejam familiarizados com o tratamento atual da empresa e executam as atividades do experimento de forma displicente. Isso ocorre porque o participante deseja mostrar que a utilização do novo tratamento não trará benefícios, se comparado com a utilização do tratamento atual da empresa. Dessa forma, os resultados podem ser interpretados incorretamente.

A causa C17 trata da replicação de experimentos. Na literatura existem estudos que afirmam que para uma maior validade externa, ou seja, para generalização dos resultados, deve haver

replicação do estudo. O fato de replicar um experimento se baseia no princípio de seguir o procedimento experimental apresentado pelo estudo original. Desse modo, o pesquisador que pretende replicar experimentos deve atentar aos detalhes do procedimento experimental do estudo principal, pois, se houverem muitas mudanças, o estudo pode acabar abandonando o objetivo do experimento original.

Ao estabelecer tempo para execução e conclusão do experimento o pesquisador acaba gerando a causa C18. Nesse caso, o tempo estabelecido pode não ser suficiente para o cumprimento de todas as tarefas experimentais. Dessa forma, o participante pode não concluir todas as tarefas ou acelerar a conclusão, podendo cometer erros durante a execução do experimento. Com isso, os dados obtidos podem ser prejudicados.

A causa C19 está relacionada à falta de comprometimento do participante. Isso acontece pelo fato do participante não se dedicar ao experimento, não se importar com a execução correta, não estar motivado para desempenhar as atividades de forma precisa. Desse modo, os dados fornecidos por este tipo de participante podem apresentar diferenças significativas frente aos demais.

Quando o idioma do material experimental não é o mesmo idioma dos participantes, a compreensão do texto pode ser comprometida, o que retrata a causa C20. Esse fato pode ocorrer quando a realização de um experimento é feita em um determinado país com o idioma de outro. Um exemplo é a replicação de um estudo experimental com o objetivo de avaliar técnicas de inspeção, onde o estudo original foi realizado nos Estados Unidos e a replicação na Suécia.

3.2. Causas de Ameaças à Validade Interna Em Experimento

A Tabela 3.2 apresenta as causas de ameaças à validade interna que podem ocorrer durante o experimento e as evidências em que foram encontradas. A causa C21 diz respeito à comunicação entre os participantes durante os intervalos entre as execuções. Como exemplo, temos um experimento com intervalo de um dia. Nesse tempo, os participantes podem conversar sobre algo relacionado ao experimento, podendo causar interferência nos resultados.

A causa C22 trata da longa duração dos experimentos. O participante pode ficar cansado, por exemplo, durante um experimento com cinco horas de duração sem intervalos. Além disso, caso exista mais de uma tarefa a serem realizadas, estas podem deixar o participante entediado caso sejam muito parecidas.

A causa C23 representa as diferenças na forma de medição dos dados. Isso ocorre quando a responsabilidade pela medição fica a cargo de pessoas, por exemplo. Caso apenas uma pessoa

realize a medição, os dados são obtidos por apenas um ponto de vista, podendo apresentar viés na interpretação.

Tabela 3.2. Lista de causas de ameaças à validade interna em experimento.

Causa	Evidência(s)
C21. Comunicação entre os participantes durante os intervalos entre as execuções.	EP2, EP3, EP17, EP18, EP23, EP25.
C22. Experimento com longa duração (cansaço, tédio).	EP4, EP7, EP9, EP13, EP15, EP18, EP27, EP37, EP39, EP43, EP46.
C23. Diferenças na forma de medição dos dados.	EP2, EP15, EP26, EP27.
C24. Utilização do tratamento é feita de forma diferente da ideal pelos participantes.	EP2, EP4, EP10, EP19, EP25, EP46.
C25. Participantes adquirem aprendizado durante o experimento.	EP3, EP5, EP6, EP7, EP9, EP16, EP18, EP20, EP22, EP23, EP24, EP25, EP27, EP29, EP30, EP31, EP32, EP33, EP37, EP38, EP39, EP41, EP42, EP45, EP46.
C26. Utilização dos mesmos instrumentos e tratamentos em mais de uma execução pelos participantes.	EP3.
C27. Troca de informação entre os participantes durante o experimento.	EP4, EP7.
C28. Coleta de dados realizada de forma incorreta por parte de alguns participantes.	EP5.
C29. Participantes não comparecem ao experimento.	EP6, EP14, EP28, EP36, EP37, EP43, EP46.
C30. Falta de precisão dos dados.	EP7, EP26, EP33, EP36.
C31. Participantes não completam o experimento.	EP15, EP18, EP35.
C32. Ordem de utilização dos instrumentos.	EP15, EP23, EP24, EP25, EP27, EP30.
C33. Comunicação entre os membros do mesmo grupo.	EP25.

A utilização do tratamento realizada de forma diferente da ideal por parte dos participantes se torna a causa C24. A preocupação dos pesquisadores para que o participante utilize os procedimentos experimentais de forma correta é retratada nesta causa. Existem participantes que são atribuídos aos tratamentos e, ao utilizá-los, acabam o fazendo de forma incorreta, causando divergência entre os dados.

A causa C25 diz respeito à aquisição de aprendizado durante o experimento por parte dos participantes. Ao apresentar mais de um tratamento, ou instrumento, ou tarefa, a ordem de execução/utilização durante o experimento pode causar aprendizado aos participantes. Por exemplo, ao utilizar o instrumento A e depois o instrumento B, pode ocorrer um aprendizado durante a utilização do instrumento A que pode melhorar o desempenho do participante durante a utilização do instrumento B. Esse aprendizado pode ser indesejado pelo pesquisador responsável pelo experimento.

A causa C26 trata da utilização dos mesmos instrumentos ou tratamentos pelos participantes em mais de uma execução. Com isso, ao utilizar um mesmo tratamento, o participante pode ficar entediado e apresentar desinteresse pela continuidade do experimento, por exemplo.

A troca de informação entre os participantes durante o experimento é identificada como a causa C27. Esta troca de informação deve ser observada com bastante rigor, pois informações relevantes sobre execução e utilização de tratamentos e/ou instrumentos podem ser disseminadas entre os participantes durante o experimento.

A causa C28 diz respeito à realização da coleta de dados de forma incorreta por parte dos participantes. Podem existir participantes que não seguem as instruções fornecidas e a coleta de dados pode ser feita de forma indesejada. Além disso, ao avaliar uma inspeção de documentos em busca de defeitos, por exemplo, o participante pode identificá-los através de leitura inicial, sem estar utilizando o tratamento.

A causa C29 retrata o não comparecimento dos participantes no dia da execução do experimento. Após, o pesquisador ter realizado o balanceamento dos grupos que utilizarão os tratamentos, pode ocorrer a desistência de algum participante. Dessa forma, o pesquisador deve tomar providências cabíveis, pois pode existir um desbalanceamento entre os grupos formados, por exemplo.

A falta de precisão nos dados fornecidos pelos participantes foi identificada como a causa C30. Isso ocorre quando a responsabilidade da medição dos dados fica a cargo do próprio participante. Porém, os participantes podem não ser precisos ao informarem determinado resultado. Por exemplo, tempo de inspeção de um documento. Durante a inspeção, o participante pode realizar

outras tarefas ou fazer intervalos e continuar cronometrando o tempo. Ao final, o tempo total de inspeção não irá refletir a realidade da aplicação.

A causa C31 trata dos participantes que não completam a execução do experimento. Isso ocorre por desistência do participante com o decorrer do experimento ou por falta de dados que deixaram de ser coletados, por exemplo.

A ordem de utilização dos instrumentos é identificada como a causa C32. Ao estabelecer uma ordem para que os participantes utilizem os instrumentos, o pesquisador deve atentar para o fato de não causar aprendizado ou, até mesmo, tédio. A ordem de utilização pode afetar o desempenho dos participantes tanto positiva quanto negativamente, fazendo com que o resultado do experimento seja modificado.

A causa C33 trata da comunicação entre membros de um mesmo grupo. Caso haja divisão dos participantes em grupos, deve-se atentar ao fato de que um grupo pode desempenhar as tarefas de uma melhor forma, pois estes produziram uma melhor capacidade de comunicação, facilitando o desempenho do grupo.

3.3. Causas de Ameaças à Validade Externa

As causas de ameaças à validade externa e as evidências em que foram encontradas são listadas na Tabela 3.3. A causa C34 está relacionada à representatividade dos instrumentos experimentais. Os instrumentos utilizados em experimentos executados em ambiente acadêmico pode não representar os instrumentos utilizados na indústria em termos de complexidade, tamanho, dentre outros. Além disso, no caso de instrumentos próprios da indústria, estes podem não representar todos os instrumentos do contexto abordado pelo experimento.

A representatividade da amostra é identificada pela causa C35. É difícil para os pesquisadores obter uma amostra representativa em estudos experimentais na área de ES. Grande parte dos estudos utiliza como participantes alunos de graduação e pós-graduação, sendo grande parte não representante dos profissionais da área.

A causa C36 diz respeito ao ambiente experimental. Em alguns casos, o ambiente pode proporcionar uma maior concentração, ou pode atrapalhar o desempenho do participante, ou seja, as condições apresentadas no ambiente podem influenciar positiva ou negativamente os resultados. Ao medir o tempo de resposta de uma transação de bando de dados, por exemplo, a infraestrutura da rede pode causar diferenças nos resultados.

A representatividade do processo executado é identificada pela causa C37. O processo utilizado em alguns experimentos pode não representar os que são aplicados pela indústria. Devido

ao tempo de execução do experimento, algumas fases dos processos podem ser descartadas, podendo causar alguma influência nos resultados. Por outro lado, por decisão dos pesquisadores, a utilização de todo o processo é descartada, sendo aplicadas somente as fases que interessam ao experimento.

Tabela 3.3. Lista de causas de ameaças à validade externa.

Causa	Evidência(s)
C34. Representatividade dos documentos.	EP3, EP4, EP5, EP6, EP7, EP8, EP9, EP10, EP12, EP14, EP15, EP16, EP18, EP19, EP20, EP22, EP23, EP24, EP25, EP28, EP29, EP30, EP31, EP32, EP33, EP34, EP35, EP36, EP38, EP39, EP40, EP41, EP42, EP43, EP44, EP45.
C35. Representatividade da amostra.	EP1, EP2, EP3, EP4, EP5, EP6, EP7, EP8, EP9, EP10, EP12, EP13, EP14, EP15, EP16, EP18, EP19, EP20, EP21, EP23, EP24, EP25, EP28, EP29, EP30, EP31, EP32, EP33, EP34, EP35, EP36, EP37, EP38, EP39, EP40, EP41, EP42, EP43, EP44, EP45.
C36. Ambiente experimental.	EP2, EP14, EP17, EP21, EP26, EP34, EP35, EP37, EP46.
C37. Representatividade do processo executado.	EP3, EP9, EP10, EP16, EP19, EP20, EP22, EP23, EP24, EP25, EP29, EP30, EP31, EP32, EP33, EP36, EP38, EP39, EP42, EP43.
C38. Problemas introduzidos.	EP9, EP39.
C39. Representatividade das tarefas.	EP1, EP7, EP13, EP15, EP17, EP35, EP36.

A causa C38 trata dos problemas que são introduzidos nos artefatos para obtenção de dados. Por exemplo, inserção de defeitos em documento que deve ser inspecionado por um participante utilizando uma nova técnica. Os defeitos introduzidos podem não representar os que ocorrem na indústria em termos de complexidade.

A causa C39 está relacionada à representatividade das tarefas experimentais. As tarefas utilizadas nos experimentos executados em ambiente acadêmico pode não representar as tarefas aplicadas na indústria em termos de complexidade, tamanho, dentre outros.

3.4. *Relacionamentos entre Causas e Tratamentos Identificados*

A Tabela 3.4 apresenta os tratamentos para causas de ameaças à validade interna identificados e as causas em que foram aplicados. Além disso, as evidências em que os tratamentos foram encontrados são listadas.

Conforme pode ser observado, não foram identificados tratamentos para as causas C2, C11, C14, C16 e C26. Além disso, podemos verificar que alguns tratamentos podem ser aplicados a mais de uma causa: tratamento T10 aplicado às causas C5, C24 e C25; tratamento T13 aplicado às causas C7 e C32; e tratamento T26 aplicado às causas C21 e C27.

Tabela 3.4. Lista de tratamentos das causas de ameaças à validade interna.

Tratamento	Causa(s)	Evidência(s)
T1. Caracterizar a experiência dos participantes.	C1.	EP1, EP7, EP20, EP29, EP32, EP36, EP38.
T2. Analisar as atividades realizadas pelos participantes anteriormente.	C3.	EP30, EP39.
T3. Caracteriza a habilidade dos participantes.	C3.	EP1, EP4, EP9, EP16, EP20, EP21, EP23, EP28, EP29, EP32, EP36, EP41, EP42, EP43, EP44, EP45, EP46.
T4. Aplicar sessões de treinamento para nivelar habilidade.	C3.	EP2, EP4, EP31, EP32, EP35, EP40.
T5. Dividir os participantes em blocos a partir do nível de habilidade.	C3.	EP4, EP9, EP10, EP16, EP20, EP21, EP23, EP28, EP29, EP30, EP32, EP36, EP38, EP44, EP46.
T6. Formar grupos aleatoriamente.	C3.	EP4, EP6, EP10, EP12, EP16, EP20, EP21, EP23, EP24, EP25, EP27, EP28, EP29, EP30, EP32, EP33, EP35, EP36, EP38, EP40, EP41, EP42, EP44, EP46.
T7. Elaborar material com maior clareza.	C4.	EP1, EP46.
T8. Realizar atividade piloto para avaliar o material.	C4.	EP1.

Tratamento	Causa(s)	Evidência(s)
T9. Elaborar instrumentos diferentes para minimizar o aprendizado.	C5.	EP2.
T10. Aplicar sessões de treinamento para equilibrar o conhecimento.	C5, C24, C25.	EP2, EP3, EP16, EP18, EP20, EP30.
T11. Atribuir aleatoriamente os participantes aos tratamentos.	C6.	EP2, EP4, EP10, EP20, EP24, EP27, EP30, EP31, EP32, EP33, EP35, EP37, EP38, EP40, EP45.
T12. Elaborar instrumentos com alto nível de equivalência.	C7.	EP5, EP6, EP7, EP12, EP16, EP20, EP30, EP33, EP38, EP41.
T13. Utilizar os mesmos instrumentos para todos os participantes.	C7, C32.	EP15, EP16, EP21, EP23, EP24, EP25, EP27, EP29, EP30, EP31, EP32, EP42.
T14. Analisar a variância dos resultados para observar o efeito causado.	C7.	EP39.
T15. Elaborar tarefas com alto nível de equivalência.	C8.	EP5, EP6, EP25.
T16. Estabelecer uma forma de compensação entre as tarefas.	C8.	EP25.
T17. Obter artefatos da literatura.	C9.	EP5, EP6.
T18. Informar o anonimato dos dados coletados.	C10.	EP7.
T19. Analisar os questionários pós-experimento para verificar o efeito causado.	C12.	EP25.
T20. Realizar a revisão do material por mais de uma pessoa.	C13.	EP46.
T21. Aplicar os mesmo procedimentos experimentais a todos os participantes, com o mínimo de diferenças.	C15.	EP15.

Tratamento	Causa(s)	Evidência(s)
T22. Utilizar os procedimentos experimentais do estudo original.	C17.	EP25, EP31, EP32.
T23. Não estabelecer tempo para término do experimento.	C18.	EP25.
T24. Informar que a aprovação em uma disciplina depende do desempenho do participante no experimento.	C19.	EP38.
T25. Disponibilizar dicionários para que os participantes possam consultar.	C20.	EP42.
T26. Solicitar que os participantes não conversem sobre o experimento.	C21, C27.	EP2, EP7, EP18.
T27. Não possibilitar acesso ao material experimental fora da execução.	C21.	EP3, EP23.
T28. Fazer com que os participantes utilizem os instrumentos na mesma ordem.	C21.	EP25.
T29. Disponibilizar intervalos durante as execuções.	C22.	EP2, EP18, EP43.
T30. Dividir o experimento em tarefas de conclusão rápida.	C22.	EP7, EP15, EP37, EP46.
T31. Informar o contexto da pesquisa para envolver os participantes.	C22.	EP9.
T32. Se houverem vários tratamentos, dividir o experimento em fases com um tratamento em cada.	C22.	EP13.
T33. Realizar o experimento em dias consecutivos.	C22.	EP27.
T34. Realizar a medição com mais de uma pessoa.	C23.	EP2.

Tratamento	Causa(s)	Evidência(s)
T35. Utilizar uma técnica específica para medição.	C23.	EP27.
T36. Estabelecer uma medida para controle de conformidade.	C24.	EP2.
T37. Analisar os dados para descarte.	C24.	EP4, EP10.
T38. Manipular a ordem de execução dos tratamentos.	C25.	EP7, EP16, EP24, EP27, EP29, EP30, EP31, EP32, EP33, EP37, EP39, EP42, EP46.
T39. Utilizar um tratamento apenas.	C25.	EP9, EP38, EP41.
T40. Utilizar tipos de tarefas diferentes.	C25.	EP22.
T41. Utilizar o mesmo tratamento para todos os participantes.	C25.	EP25.
T42. Coletar toda informação que possa comprometer o experimento.	C27.	EP4.
T43. Comparar as variáveis coletadas para verificar o efeito causado.	C28.	EP5.
T44. Analisar os dados dos participantes para verificar equilíbrio entre os grupos.	C29.	EP6, EP36.
T45. Informar que as atividades do experimento são equivalentes às realizadas fora dele.	C29.	EP37.
T46. Realizar apenas uma sessão do experimento.	C29.	EP43.
T47. Formar os grupos e atribuir os participantes aos tratamentos no dia da execução do experimento.	C29.	EP28.
T48. Solicitar um relatório de progresso do experimento.	C30.	EP26.

Tratamento	Causa(s)	Evidência(s)
T49. Elaborar uma porcentagem de erro para estabelecer critério de aceitação.	C30.	EP33.
T50. Analisar os dados para verificar a precisão.	C30.	EP36.
T51. Excluir os dados incompletos.	C31.	EP18.
T52. Utilizar os instrumentos em todas as ordens possíveis.	C32.	EP24, EP30.
T53. Formação de grupos diferentes em cada fase do experimento, se possível.	C33.	EP25.

Por outro lado os tratamentos para causas de ameaças à validade externa identificados e as causas em que foram aplicados são apresentados na Tabela 3.5. As evidências em que os tratamentos foram encontrados também são listadas.

Podemos verificar que alguns tratamentos podem ser aplicados a mais de uma causa: tratamento T54 aplicado às causas C34, C35, C37 e C38; e tratamento T61 aplicado às causas C35 e C36.

Tabela 3.5. Lista de tratamentos das causas de ameaças à validade externa.

Tratamento	Causa(s)	Evidência(s)
T54. Replicar o experimento.	C34, C35, C37, C38, C39.	EP1, EP5, EP7, EP16, EP30, EP33, EP39.
T55. Elaborar instrumentos que sejam realistas ao contexto.	C34.	EP4, EP9, EP10, EP12, EP14, EP15, EP18, EP19, EP29, EP41, EP43.
T56. Apresentar dados que provem que não há diferenças entre os instrumentos e/ou tarefas.	C34, C39.	EP35.
T57. Utilizar os mesmos instrumentos, no caso de replicação.	C34.	EP44.
T58. Realizar treinamento para uso do tratamento.	C35.	EP3.

Tratamento	Causa(s)	Evidência(s)
T59. Apresentar dados que provem que não há diferenças entre estudantes e profissionais da indústria.	C35.	EP4.
T60. Selecionar participantes com experiência industrial.	C35.	EP13, EP14, EP18, EP19, EP20, EP29, EP31.
T61. Simular um ambiente real.	C35, C36.	EP12, EP26, EP37.
T62. Realizar o estudo em ambiente industrial.	C36.	EP14.
T63. Utilizar um processo que seja aplicado na indústria.	C37.	EP3, EP10, EP16, EP23, EP25, EP29, EP30, EP38, EP39, EP42, EP43.
T64. Utilizar parte de um processo aplicado na indústria.	C37.	EP9, EP31, EP32.
T65. Utilizar processo equivalente a processos aplicados na indústria.	C37.	EP21.
T66. Buscar os problemas em diversas fontes da literatura.	C38.	EP9.

Foram identificados diferentes relacionamentos entre causas e tratamentos. Os relacionamentos serão apresentados graficamente para uma melhor visualização e entendimento. A Figura 3.1 apresenta os relacionamentos que envolvem a causa C1. Conforme pode ser observado, a causa C1 possui T1 como tratamento imediato, onde a caracterização da experiência do participante deve ser realizada. Contudo, a aplicação deste tratamento pode causar C15. Por sua vez, C15 é tratada através de T21, pois todos os procedimentos experimentais aplicados devem ser realizados igualmente para todos os participantes.

Como tratamentos opcionais, ou seja, tratamentos que podem ser aplicados para minimizar ainda mais a ameaça, temos: T3, T4, T5 e T6. O primeiro é realizado com o objetivo de caracterizar as habilidades do participante, visando um maior equilíbrio entre todos os participantes. O segundo é aplicado para que as habilidades dos participantes sejam niveladas. Assim como a aplicação do tratamento T1, o uso dos tratamentos T3 e T4 podem causar C15. Ainda com relação ao uso de T4, temos que este pode causar C12, que por sua vez é tratado por T19. Os efeitos dos treinamentos

realizados podem ser verificados através de análise dos questionários pós-experimento, onde o participante deve relatar possíveis dificuldades e facilidades proporcionadas pelo treinamento.

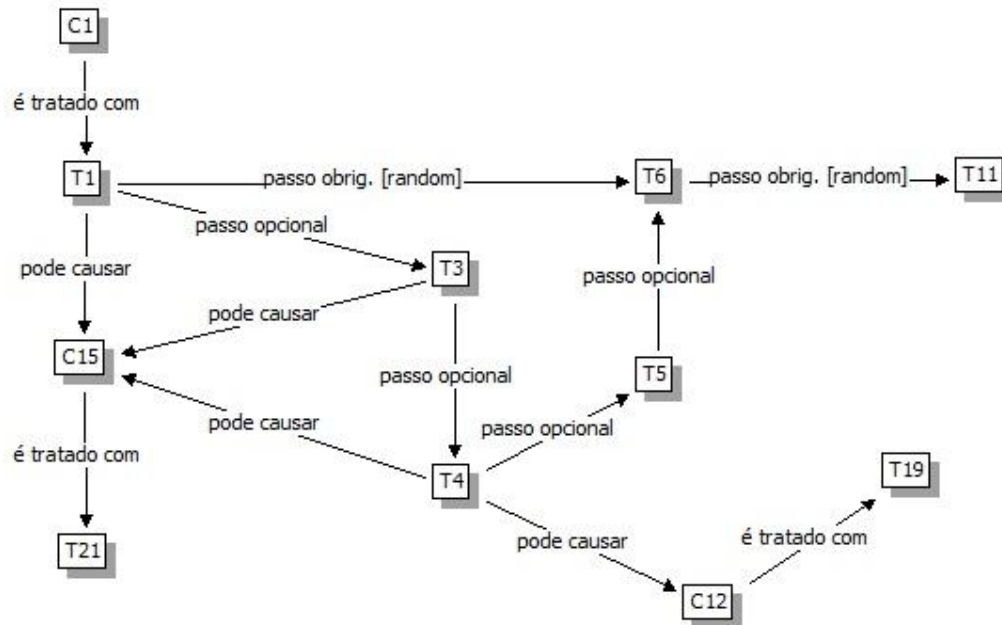


Figura 3.1. Relacionamentos identificados para tratamentos da causa C1.

Ainda como passo opcional, o tratamento T5 é aplicado com o objetivo de separar os participantes em grupos de baixa, média e alta habilidade. Dessa forma, T6 pode ser aplicado para formação dos grupos de forma aleatória. Como tratamento complementar à T6, temos a utilização de T11 com o objetivo de atribuir aleatoriamente os tratamentos a serem utilizados durante o experimento pelos participantes. Caso a amostra do estudo seja obtida de forma aleatória, devem ser aplicados como passos obrigatórios os tratamentos T6 seguido de T11. Por outro lado, se a amostra for obtida por conveniência, os passos opcionais podem ser aproveitados.

A Figura 3.2 apresenta os relacionamentos das causas C3. Analisando a figura, pode ser visto que os tratamentos T3, T4, T5, T6 e T11 estão relacionados à causa C3 como tratamentos imediatos, diferentemente da causa C1. Além disso, estes tratamentos podem ser utilizados de forma individual ou em conjunto para mitigar a causa em questão. O raciocínio empregado é o mesmo que foi apresentado para a Figura 3.1. Ainda com relação à causa C3, pode ser notado que T2 também é um tratamento imediato, que é a análise de atividades realizadas pelos participantes anteriormente, pois com base nos dados históricos é possível estabelecer um balanceamento entre os participantes. Além disso, podemos observar que como tratamento da causa C6 temos T11, que é a atribuição aleatória dos tratamentos aos participantes.

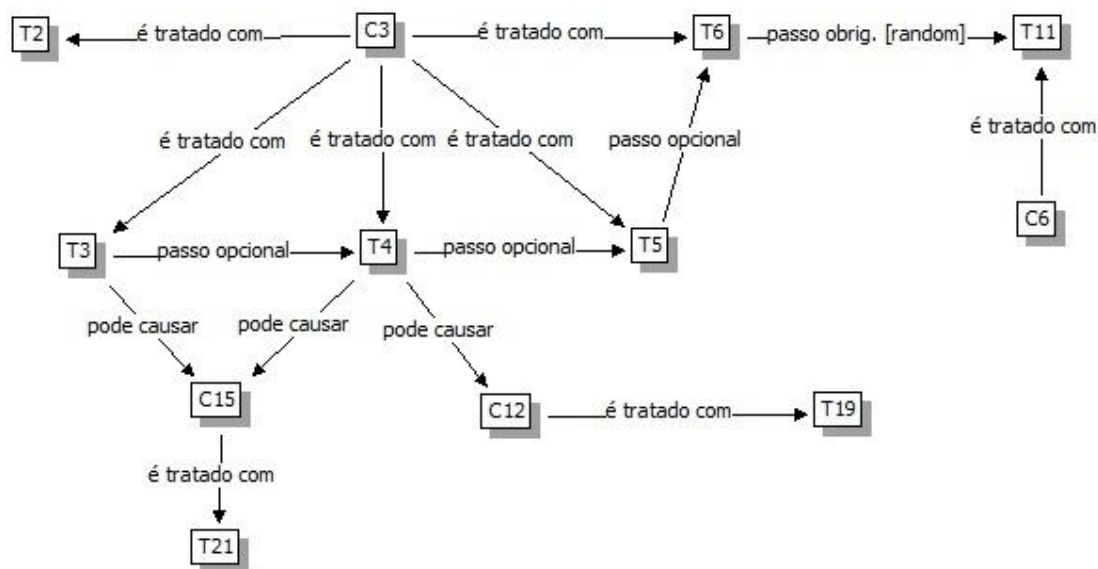


Figura 3.2. Relacionamentos identificados para tratamentos da causa C3.

Na Figura 3.3 são apresentados os relacionamentos para as causas C10, C17, C18, C19 e C20. Às causas C10, C18 e C19 são aplicados apenas um tratamento imediato T18, T23 e T24, respectivamente, não existindo nenhum outro relacionamento entre eles. O tratamento T18 é utilizado para que os participantes tenham conhecimento de que os dados que são coletados durante todo o experimento são mantidos em anonimato total. Dessa forma, pretende-se proporcionar um ambiente de confiança aos participantes. A utilização do tratamento T23 é feita para que não seja estabelecido tempo de execução do experimento, com o objetivo de deixar o participante sem pressa de concluir o estudo. Para obter certo comprometimento dos participantes, no caso destes serem alunos, informar que a aprovação em uma disciplina depende do desempenho no experimento.

As causas C17 e C20 também possuem apenas um tratamento imediato T22 e T25, respectivamente. Porém, existe uma relação entre as causas, onde o estudo original pode apresentar o material experimental em um idioma diferente dos participantes do estudo de replicação. Dessa forma, C20 pode ocorrer caso o experimento seja uma réplica.

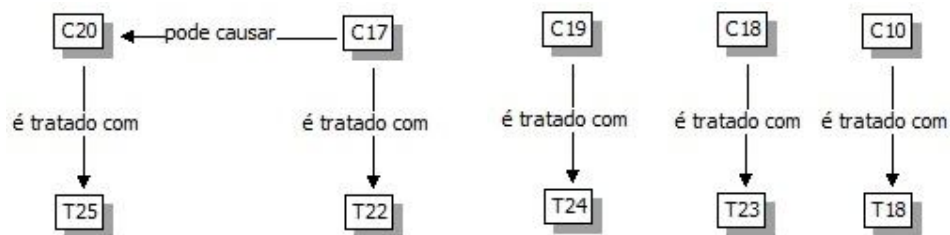


Figura 3.3. Relacionamentos para tratamento das causas C10, C17, C18, C19 e C20.

A Figura 3.4 apresenta os relacionamentos das causas C4, C9 e C13. O tratamento T17 é aplicado apenas a C9 de forma imediata, com o objetivo de obter tarefas e instrumentos da

literatura, evitando problemas de domínio do problema e de especificação. A causa C4 é tratada por T7 e T8. Ao aplicar T7, o pesquisador deve elaborar um material experimental bem claro e de fácil compreensão, para que as dúvidas dos participantes sejam mínimas. Por outro lado, a aplicação de T8 pode causar C11, pois a aplicação de uma atividade piloto para avaliação do material experimental pode prover aprendizado aos participantes. Para tratar a causa C13 temos a aplicação do tratamento T20, onde a revisão do material experimental deve ser realizada por mais de uma pessoa. Pode ser observado que C13 pode causar C4 e C9, pois a não utilização de especialista para elaboração do material experimental favorece uma má especificação do material experimental.

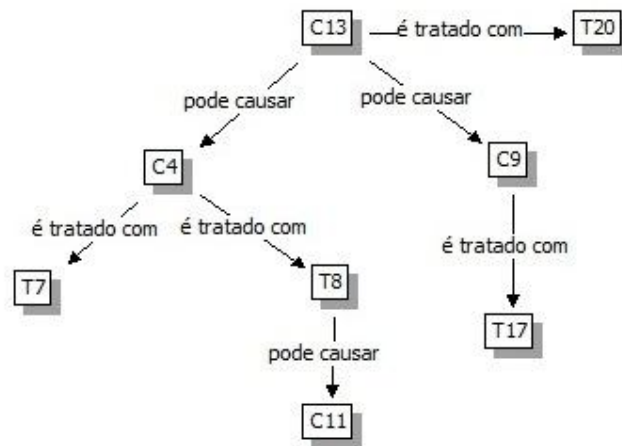


Figura 3.4. Relacionamentos identificados para tratamentos das causas C4, C9 e C13.

Os relacionamentos identificados para as causas C7 e C8 são apresentados na Figura 3.5. As causas em questão apresentam como tratamentos imediatos T12, T13 e T14, para C7, e T15 e T16, para C8. Ao elaborar instrumentos com um alto grau de equivalência, o pesquisador estará mitigando C7.

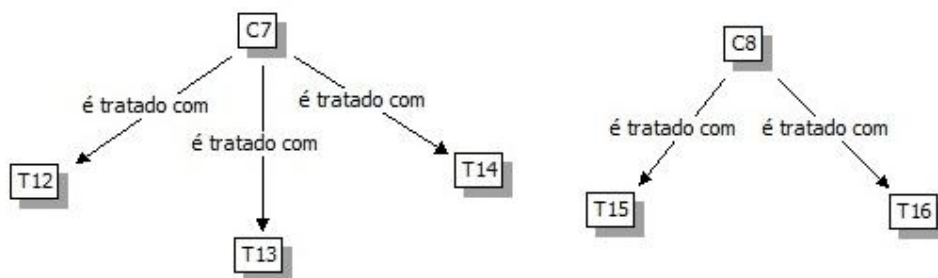


Figura 3.5. Relacionamentos identificados para tratamentos das causas C7 e C8.

Além disso, pode-se fazer com que todos os participantes utilizem os mesmos instrumentos, onde os resultados não sofrerão influência dos instrumentos. Pode-se também analisar a variância dos resultados de cada participante ao término do experimento para observar se houve viés.

A Figura 3.6 apresenta os tratamentos para as causas C21, C22 e C27. Os tratamentos para a causa C22 são cinco: T29, T30, T31, T32 e T33. Ao aplicar o tratamento T29, o pesquisador possibilita um intervalo aos participantes, para que façam alguma refeição, ou até mesmo retornarem à execução no dia seguinte. O tratamento T30 deve ser utilizado caso o experimento possa ser dividido em tarefas de conclusão rápida, para que as execuções não sejam cansativas. O pesquisador pode também informar aos participantes o contexto da pesquisa (T31), tentando motivá-los, mostrando a importância do experimento no contexto abordado. Caso o experimento possua mais que um tratamento, pode haver uma divisão em fases, onde um tratamento é apresentado e utilizado em cada uma das fases (T32). Por último, temos o tratamento T33, que tem por objetivo a execução do experimento em dias consecutivos. Conforme é mostrado na Figura 3.6, os tratamentos T29 e T33 causam C21. Isso ocorre porque existe um intervalo nesses tratamentos, então existe a possibilidade de troca de informação durante os intervalos.

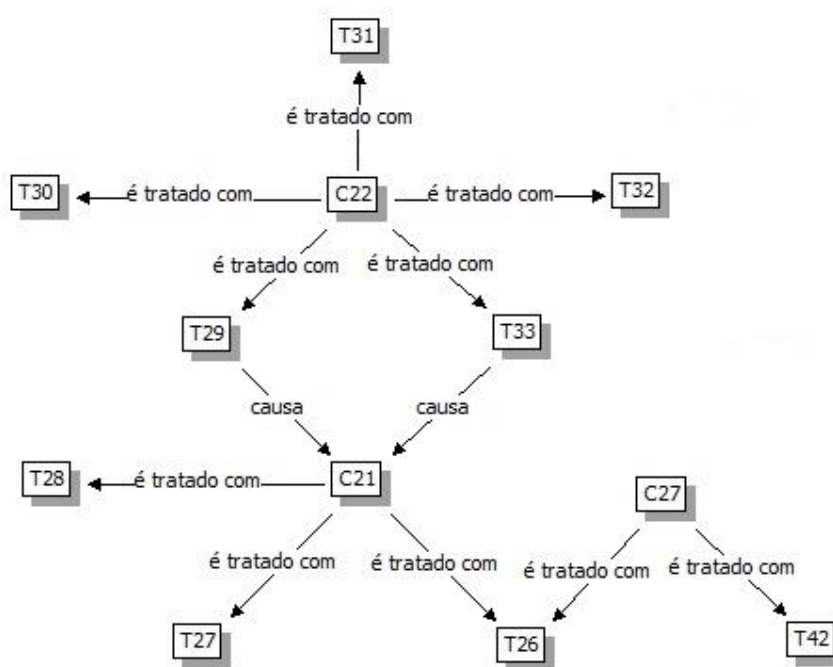


Figura 3.6. *Relacionamentos identificados para tratamentos das causas C21, C22 e C27.*

Para tratar C21, temos T26, T27 e T28. Ao utilizar o tratamento T28, os instrumentos devem ser utilizados na mesma ordem por todos os participantes. Desse modo, nenhuma informação sobre a utilização de diferentes instrumentos é gerada, fazendo com que os participantes tenham conhecimento dos instrumentos apenas na hora de sua utilização. O tratamento T27 tem como objetivo não permitir que os participantes tenham acesso ao material experimental fora do ambiente de execução do experimento. Com isso, nenhuma informação é gerada enquanto o experimento não é concluído. Por último temos o tratamento T26, onde os pesquisadores solicitam aos participantes que evitem conversar sobre o experimento, pois informações sobre utilização de tratamentos ou

instrumentos pode ser disseminada. A causa C27 pode ser tratada por T26 e por T42, cujo objetivo é coletar toda informação que possa comprometer a execução do experimento.

Na Figura 3.7 são apresentados os relacionamentos das causas C25, C26 e C32. Conforme pode ser visto, a causa C25 possui cinco tratamentos que podem ser aplicados. Como primeiro tratamento, temos a realização de sessões de treinamento para que haja um equilíbrio de conhecimento entre os participantes (T10). O tratamento T38 consiste da manipulação da ordem de execução dos tratamentos, ou seja, ao estabelecer determinada ordem, o pesquisador evita proporcionar aprendizado ao participante. Ao utilizar apenas um tratamento (T39) os participantes não irão adquirir aprendizado entre as execuções, podem adquirir apenas um aprendizado que é previsto pelo pesquisador. O tratamento T40 tem por objetivo utilizar tipos de tarefas diferentes, para que não haja influência de uma sobre a outra. O tratamento T41 consiste da utilização do mesmo tratamento por todos os participantes, pode até existir mais de um, porém todos os participantes utilizam. Dessa forma pretende-se mitigar a causa C25.

Conforme pode ser observado, C26 e C32 podem causar C25. Isso porque as duas primeiras geram certo aprendizado que pode ser adquirido pelo participante. Para a causa C26 não foi identificado nenhum tratamento. Por outro lado, a causa C32 possui como tratamentos o T13 e o T52. A utilização de T13 tem por objetivo a utilização dos mesmos instrumentos por todos os participantes. Já T52, sugere utilizar todos os instrumentos em todas as ordens possíveis.

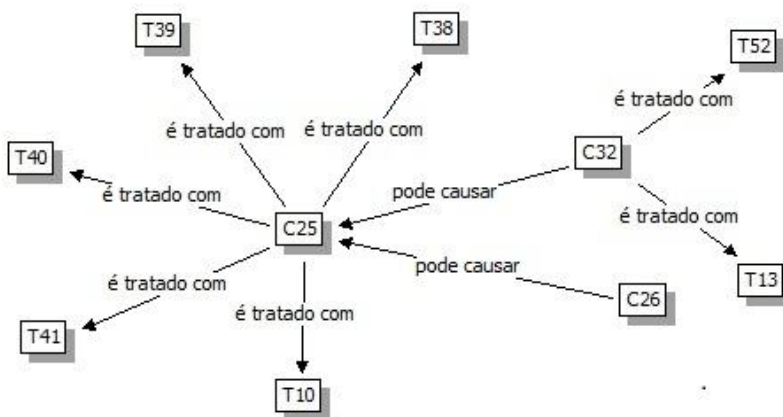


Figura 3.7. Relacionamentos identificados para tratamentos das causas C25, C26 e C32.

A Figura 3.8 apresenta os tratamentos para as causas C5, C23, C24 e C33. Como tratamentos para a causa C5 temos T9 e T10. O objetivo do tratamento T9 é a elaboração de instrumentos diferentes para que o aprendizado seja mínimo. O tratamento T10 consiste em realizar sessões de treinamento para que haja um equilíbrio de conhecimento entre os participantes. Os tratamentos para C24 são T10 (explicado anteriormente), T36 e T37. O tratamento T36 consiste na

elaboração de uma medida para controle de conformidade, estabelecendo critérios mínimos de utilização. Já o tratamento T37 sugere a análise dos dados coletados para descarte, caso seja necessário.

As causas C23 e C33 possuem dois e um tratamentos imediatos, respectivamente. Como tratamento para C23 temos a medição sendo realizada por mais de uma pessoa (T34), onde a percepção de uma pessoa pode ser confrontada com a percepção de outra(s). Tendo apenas uma pessoa responsável pela medição, os dados podem ser direcionados para um melhor resultado. Por outro lado, o tratamento T35 permite que a medição seja realizada e, posteriormente, analisada através de uma técnica específica. Finalmente, o tratamento T53 da causa C33 tem como objetivo a formação de diferentes grupos em cada fase do experimento, para que a comunicação entre os grupos seja distribuída de forma semelhante.

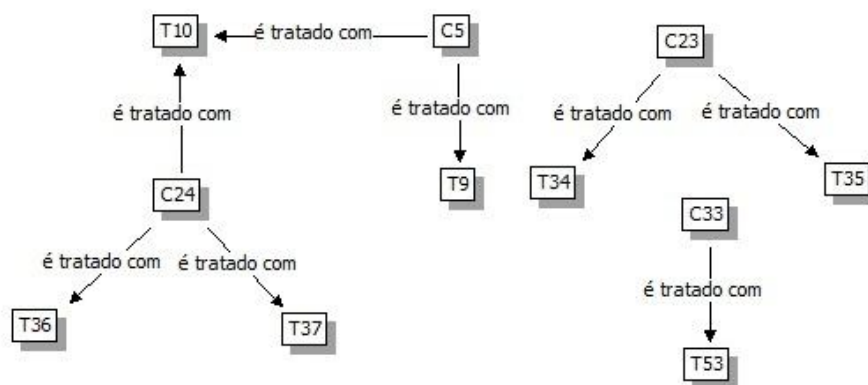


Figura 3.8. Relacionamentos identificados para tratamentos das causas C5, C23, C24 e C33.

A Figura 3.9 apresenta os tratamentos imediatos para as causas C28, C29, C30 e C31. A causa C29 apresenta quatro tratamentos: o primeiro verifica os dados dos participantes para definir o equilíbrio entre os grupos formados (T44); o segundo faz com que o pesquisador informe que as atividades a serem desenvolvidas no experimento são de conhecimento dos participantes (T45); o terceiro consiste na realização de apenas uma sessão do experimento, ou seja, apenas uma execução, fazendo com que os participantes não tenham que retornar em outra data (T46); e por fim, T47 que tem como objetivo a formação de grupos e atribuição dos tratamentos aos participantes apenas no dia de execução do experimento.

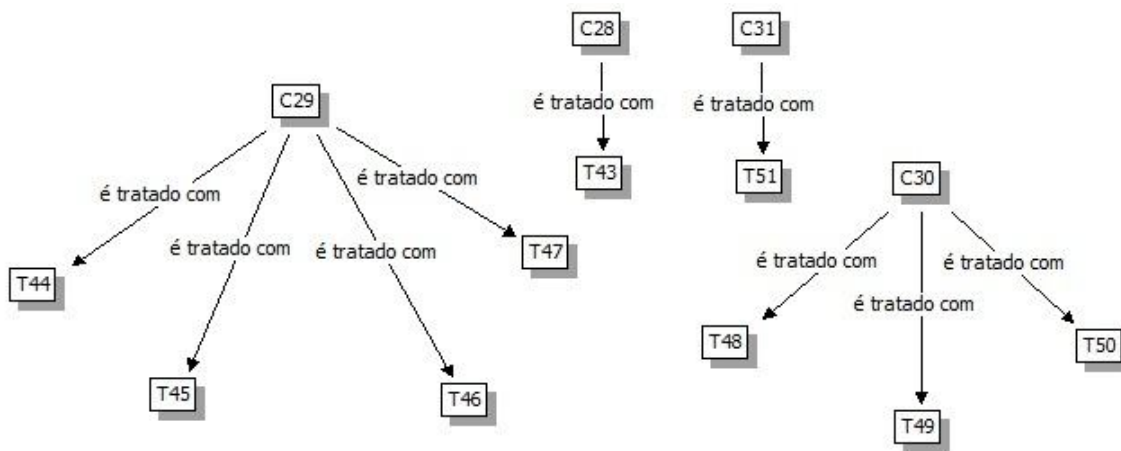


Figura 3.9. Relacionamentos identificados para tratamentos das causas C28, C29, C30 e C31.

A Figura 3.10 apresenta os relacionamentos identificados entre as causas e tratamentos de ameaças à validade externa. As causas C34 e C39 são tratadas por T54 e T56. Além disso, a primeira é tratada por T55 e T57. Muitos trabalhos afirmam que para uma maior validade externa, deve-se replicar o experimento original. Realizando um estudo com amostras diferentes, artefatos diferentes, processos da indústria e problemas que ocorrem no ambiente industrial, os resultados ganham poder de generalização. Os experimentos seriam mais realistas se fossem executados em sistemas reais, com profissionais de software que são representantes da população alvo da tecnologia, usando sua tecnologia usual de desenvolvimento em seu ambiente de funcionamento usual. Como tratamento para as causas C34 e C39, sugere-se a replicação do experimento (T54) com a utilização de instrumentos e/ou tarefas industriais, ou instrumentos e/ou tarefas provenientes de organizações diferentes, caso o estudo original tenha ocorrido na indústria. O tratamento T55 sugere que os instrumentos elaborados reflitam a realidade do contexto industrial. Caso não exista diferença na utilização de instrumentos e/ou tarefas elaborados pelo pesquisador (T56), deve-se apresentar provas de que tal fato realmente ocorre. No caso de replicação de um experimento, é indicado que os instrumentos sejam os mesmos do experimento original (T57).

Para a causa C35, os tratamentos aplicados são T54, T58, T59, T60 e T61. A utilização do tratamento T58 é feita ao se realizar sessões de treinamento para uso do tratamento. Os pesquisadores acreditam que com o passar do tempo, os participantes apresentarão habilidades de profissionais da indústria através do treinamento. Existem evidências na literatura que mostram que não há diferenças no desempenho de estudantes ao invés de profissionais em estudos experimentais. Dessa forma, devem ser apresentados dados que provem este fato (T59). O pesquisador pode também tratar a causa C35 com a seleção de participantes com experiência industrial (T60). Além disso, o pesquisador pode simular um ambiente real (T61) no laboratório, porém ainda assim

existirão fatores que poderão influenciar os resultados, por exemplo, o comportamento do participante sabendo que está sendo observado. A causa C36 possui dois tratamentos imediatos, T61 e T62. O tratamento T62 consiste da realização do estudo em ambiente industrial. Porém, o fator de comportamento apresentado em T61 também pode estar presente no experimento.

A causa C37 apresenta como tratamentos T54, T63, T64 e T65. O tratamento T63 sugere a utilização de um processo que seja utilizado em grande escala pela indústria. Desse modo, o processo executado pelo experimento reflete a realidade industrial. Por outro lado, o tratamento T64, quando aplicado, utiliza apenas parte do processo aplicado na indústria. Isso ocorre para que os participantes tenham tempo de concluir o experimento. Caso o processo aplicado na indústria seja seguido corretamente, pode faltar tempo para a conclusão do estudo. Finalmente, o tratamento T65 consiste em utilizar processos equivalentes aos processos aplicados na indústria. As formas de aplicação do processo a ser utilizado devem ser encontradas nos processos industriais.

Por fim, a causa C38 é tratada por T54 e T66, que consiste na buscar dos problemas em diversas fontes, para que estes possam ser válidos dentro do contexto do estudo experimental.

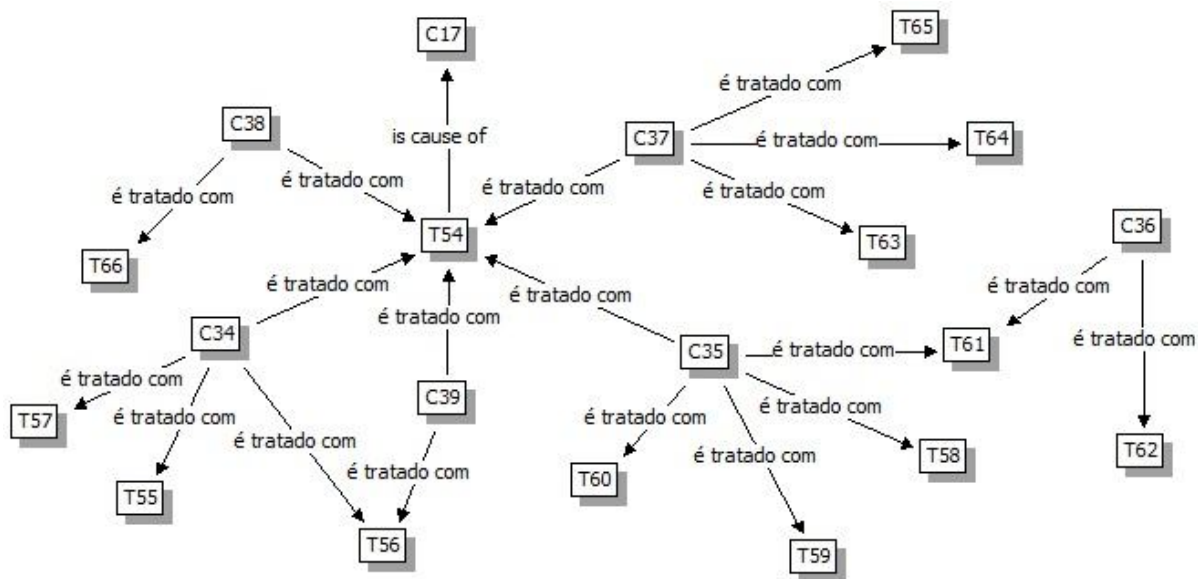


Figura 3.10. Relacionamentos identificados para tratamentos das causas de ameaças à validade externa.

Pode ser observado que o tratamento T54 causa C17, pois a replicação dos experimentos para aumento da validade externa faz com que ameaças à validade interna ocorram. O fato de adotar um tratamento que pode causar outra ameaça mostra que o pesquisador deve analisar cuidadosamente o experimento e verificar qual causa pode ter menor importância sobre o resultado final. A escolha deve ser feita com base nos resultados que o pesquisador espera conseguir.

4. Revisão Complementar da Literatura

Inicialmente, utilizamos como base a SLR de Kampenes et al. (2009) – que tinha como objetivo investigar a qualidade dos experimentos publicados em Engenharia de Software. Essa escolha foi feita pelo fato de, até onde vai nosso conhecimento, ser a SLR que abrangeu a maior quantidade de veículos de publicação de artigos na área de Engenharia de Software. No entanto, não identificamos dados suficientes para satisfazer nossas questões de pesquisa. Acreditamos que isso aconteceu porque Kampenes et al. (2009) selecionaram artigos publicados entre os anos de 1993 e 2002. Com isso, identificamos a necessidade de realizar uma nova SLR. De acordo com Carver et al. (2004), um número crescente de pesquisadores estavam incluindo resultados de experimentos em suas pesquisas.

Para conduzirmos a nossa SLR, utilizamos os mesmos procedimentos descritos por Kampenes et al. (2009), assim como os mesmos veículos de publicação. No entanto, nossas questões de pesquisa são diferentes. Por isso, adicionamos um novo critério de exclusão para atender às nossas questões de pesquisa: o artigo não descreve ameaças à validade e ações de controle.

Além disso, para os veículos de publicação apresentados a seguir, utilizamos máquinas de busca: ESEM, ICSE, IEEE Computer, IEEE Software, ISESE, IST, JSS, METRICS e TSE. Utilizamos máquinas de busca nestes veículos porque o número de artigos era muito grande para realizar busca manual. Assim, criamos a seguinte string de busca:

- controlled experiment OR randomized experiment OR randomised experiment OR quasi-experiment OR quasi experiment OR formal experiment OR empirical study OR empirical evaluation OR empirical analysis

Esta string de busca foi construída com base nos termos identificados nos artigos selecionados na SLR de Kampenes et al. (2009). A tabela 1 apresenta o resultado quantitativo da SLR.

Tabela 4.1. Resultados da RSL: Jan. 2003 – Dez. 2011.

Journal/Conference	Filter 1	Filter 2	Final	Percentual
Empirical Software Engineering (EMSE)	225	109	42	25,77%
Information and Software Technology (IST)	157	115	26	15,95%
Empirical Software Engineering and Measurement (ESEM)	117	77	22	13,50%
International Symposium on Empirical Software Engineering (ISESE)	92	67	20	12,27%
IEEE Transactions on Software Engineering (TSE)	92	80	20	12,27%
Journal of Systems and Software (JSS)	159	126	15	9,20%
IEEE International Symposium on Software Metrics (METRICS)	23	17	7	4,29%
ACM Transactions on Software Engineering Methodology (TOSEM)	139	35	4	2,45%
International Conference on Software Engineering (ICSE)	48	46	4	2,45%
Software Maintenance and Evolution (SME)	195	96	2	1,23%
Software: Practice and Experience (SP&E)	578	211	1	0,61%
IEEE Computer	8	6	0	0,00%
IEEE Software	15	5	0	0,00%
Total:	1848	990	163	100%

No primeiro filtro, foram lidos títulos e resumos dos artigos. Se um artigo satisfizesse um dos critérios de inclusão (ver tabela 2), o artigo era selecionado, caso contrário era descartado. No segundo filtro, foram lidas as seções do artigo que descreviam experimentos. Se pelo menos um critério de exclusão (ver tabela 2) fosse identificado nessa fase, o artigo era descartado. Os artigos que restaram foram utilizados para extração de dados.

Tabela 4.2. Critérios de inclusão e exclusão.

Inclusion criteria (a paper must be...)	Exclusion criteria (a paper must not be...)
<ul style="list-style-type: none"> - The use of at least two treatment conditions. - Subjects or teams as experimental units. - Performance of a Software Engineering task. - The study had to be a cause-effect investigation, i.e., the use of a treatment had to precede the measure of an outcome. 	<ul style="list-style-type: none"> - Editorials, prefaces, article summaries, interviews, news, reviews, correspondence, discussions, comments, reader's letters, and summaries of tutorials, workshops, panels, and poster sessions. - Correlation studies - studies that are based solely on calculations using existing data (e.g., from data mining), and evaluations of simulated teams based on data for individuals were excluded. - Multiple case studies - studies that use projects or companies as treatment groups, in which data is collected at several levels (treatment defined, but no experimental unit defined). - Papers that, at the outset, would not provide sufficient data for our analyses (e.g., summaries of research programs). - Usability experiments, because those are part of another discipline (human computer interaction). - Papers that do not describe threats to validity.

Os dados extraídos foram: (a) ameaças à validade e (b) ações de controle para as ameaças identificadas. Após essa etapa, foi construída uma lista de ameaças à validade e de ações de controle, as quais foram extraídas dos artigos selecionados na revisão de extensão. Finalmente, foram identificados relacionamentos entre as ameaças e as ações de controle. Todo o processo foi revisado por outro pesquisador. Foram realizadas reuniões para que possíveis divergências fossem discutidas e resolvidas.

4.1. Resultados da RSL

Experimentos controlados são reportados em 163 artigos (8,8%) de 1848 artigos checados por esta SLR, ver tabela 4.3. EMSE, ESEM, ISESE e METRICS, onde o foco é especialmente em engenharia de software experimental, reportaram mais da metade dos experimentos selecionados, cerca de 55,83%. A média do percentual dos experimentos controlados no decorrer dos anos variou entre 4,0% e 11,7%, ver tabela 4.3.

Tabela 4.3. Número de artigos que reportam ameaças à validade em experimentos controlados.

Journal	2003	2004	2005	2006	Year 2007	2008	2009	2010	2011	Total
EMSE	2	3	8	7	4	4	7	2	5	42 (18.7% of 225)
IST	2	6	2	0	1	1	10	1	3	26 (16.6% of 157)
ESEM	-	-	-	-	6	2	2	4	7	21 (18.0% of 117)
TSE	4	1	3	2	1	4	0	0	5	20 (21.7% of 92)
ISESE	2	2	3	10	-	-	-	-	-	17 (18.5% of 92)
JSS	4	1	2	2	1	3	2	0	0	15 (9.4% of 159)
METRICS	2	5	0	-	-	-	-	-	-	7 (30.4% of 23)
TOSEM	0	0	0	1	1	0	0	1	1	4 (2.9% of 139)
ICSE	0	0	1	0	0	0	1	1	1	4 (8.3% of 48)
SME	0	0	1	0	1	0	0	0	0	2 (1.0% of 195)
SP&E	0	0	0	0	0	0	1	0	0	1 (0.2% of 578)
IEEE Comp	0	0	0	0	0	0	0	0	0	0 (0% of 8)
IEEE SW	0	0	0	0	0	0	0	0	0	0 (0% of 15)
Total	16	18	20	22	15	14	23	9	22	

Com relação aos tipos de validade reportados nos artigos (ver tabela 4.4), podemos observar que a maior parte das ocorrências se deu em torno da validade interna e externa. De acordo com Wohlin et al. (2012), a validade interna tem maior prioridade, ou seja, o pesquisador deve identificar e controlar o maior número de ameaças interna possíveis.

Tabela 4.4. Tipo de validade por ano.

Validity Type	Year									
	2003	2004	2005	2006	2007	2008	2009	2010	2011	Total
Internal	13	15	16	16	14	11	20	7	17	129 (81.1% of 159)
External	13	15	15	16	14	10	20	6	18	127 (79.9% of 159)
Construct	8	9	6	9	10	7	19	6	11	85 (53.5% of 159)
Conclusion	6	5	3	5	6	4	14	5	6	54 (34.0% of 159)
Unknown	2	2	4	5	1	3	3	1	4	25 (15.7% of 159)

É importante mencionar que a classificação apresentada na tabela 4.4 foi reportada pelos autores dos respectivos artigos. Por outro lado, nós classificamos os tipos de validade dos 163 artigos baseados na definição dos tipos de validade descritas em Wohlin et al. (2012). A seguir, serão apresentadas as ameaças à validade e as ações de controle identificadas na SLR.

Identificamos um total de: (1) 36 ameaças à validade interna e 86 ações de controle; (2) 9 ameaças à validade externa e 23 ações de controle; (3) 10 ameaças à validade de constructo e 20 ações de controle; e (4) 8 ameaças à validade de conclusão e 18 ações de controle.

4.1.1. Ameaças à Validade Interna e Ações de Controle

Nesta seção apresentaremos as ameaças à validade interna que foram identificadas através da SLR. Identificamos um total de 36 ameaças à validade interna. Além disso, descreveremos as ações de controle para cada ameaça. Identificamos um total de 86 ações de controle, também como resultado da SLR.

Para facilitar a identificação das ameaças à validade e das ações de controle durante o texto, definimos um código para cada uma delas. Por exemplo, INT-A03, onde:

- INT é o tipo de validade, nesse caso interna. Temos também: externa (EXT), constructo (COT) e conclusão (COC);
- A é a letra que identifica se é uma ameaça à validade ou ação de controle, nesse caso é uma ameaça. Temos também: ação de controle (C); e
- 03 é um número sequencial, nesse caso refere-se à terceira ameaça à validade interna identificada.

A primeira ameaça à validade interna é “*Diferenças entre os participantes relacionadas à experiência*”, INT-A01 (Figura 4.1). Esta ameaça ocorre a partir da seleção dos participantes. Geralmente, os participantes possuem níveis diferentes de experiência em um determinado tópico

de pesquisa. Identificamos cinco ações de controle para essa ameaça:

- INT-C01 – Caracterizar a experiência do participante através de questionário;
- INT-C05 – Atribuir os participantes aleatoriamente aos grupos;
- INT-C06 – Caracterizar a experiência do participante através de teste pré-experimento;
- INT-C32 – Atribuir aleatoriamente os tratamentos aos participantes; e
- INT-C60 – Agrupar os participantes de acordo com o nível de experiência.

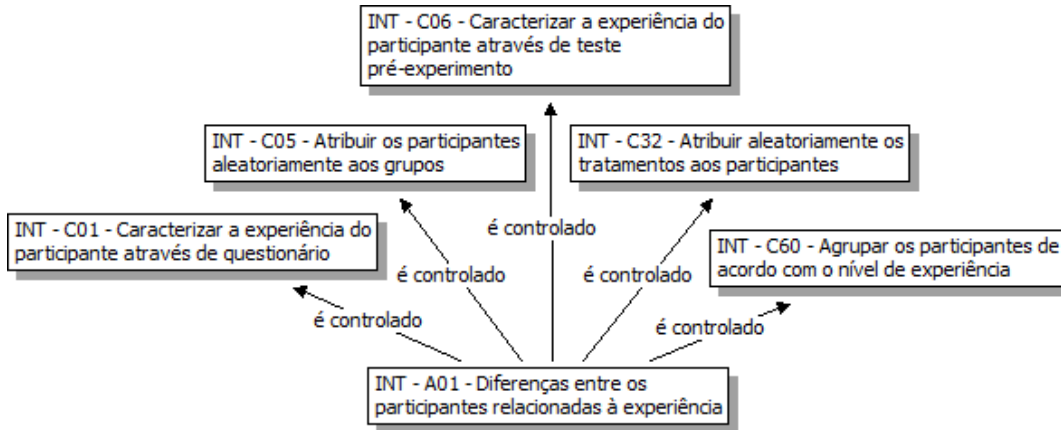


Figura 4.1. Ameaça INT-A01 e suas ações de controle.

As ações de controle podem ser aplicadas em conjunto ou separadas. As ações de controle INT-C01 e INT-C06 visam obter informações do participante para poder formar grupos balanceados (princípio de balanceamento). Por outro lado, as ações INT-C05 e INT-C32 utilizam o princípio de randomização. Por fim, a ação INT-C60 tem por objetivo agrupar os participantes de acordo com os níveis de experiência (princípio de blocagem).

A segunda ameaça é “*Familiaridade com o material experimental*”, INT-A02 (Figura 4.2). Se o participante possui pouca familiaridade com o material experimental, a execução do experimento pode estar comprometida. Para a ameaça INT-A02, assim como para a ameaça INT-A01, identificamos cinco ações de controle:

- INT-C11 – Aplicar atividade inicial para nivelar a familiaridade do participante com o material;
- INT-C33 – Realizar treinamento para prover familiaridade com o material experimental;
- INT-C45 – Explicar o material experimental para nivelar a familiaridade;
- INT-C56 – Estabelecer tempo longo de utilização do material; e
- INT-C70 – Utilizar material familiar a todos os participantes.

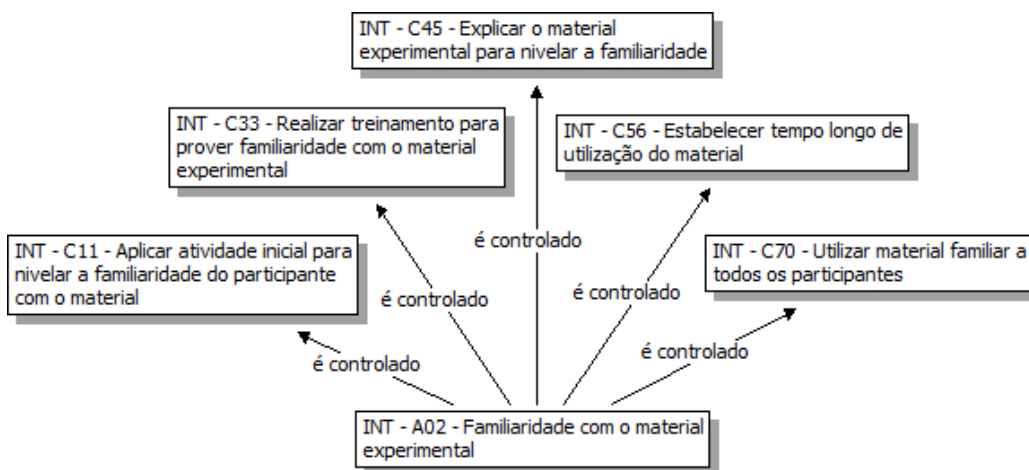


Figura 4.2. Ameaça INT-A02 e suas ações de controle.

As ações INT-C11 e INT-C45 visam estabelecer um nivelamento de familiaridade entre os participantes. Por outro lado, as ações INT-C33 e INT-C56 têm como objetivo prover familiaridade para os participantes. Por fim, a ação INT-C70 sugere que o material experimental seja familiar aos participantes.

O fato de o participante adquirir aprendizado durante o experimento é caracterizado pela ameaça INT-A03 (Figura 4.3). A aquisição de aprendizado pode ocorrer pela utilização de material experimental, aplicação dos tratamentos, dentre outros. Para esta ameaça identificamos doze ações de controle:

- INT-C13 – Estabelecer ordem de realização das tarefas;
- INT-C17 – Utilizar material experimental distinto nas diferentes execuções;
- INT-C18 – Estabelecer um design que equilibra o aprendizado;
- INT-C23 – Aplicar o tratamento logo após o treinamento;
- INT-C40 – Estabelecer ordem de utilização do material;
- INT-C44 – Estabelecer critérios para interação entre o pesquisador e o participante;
- INT-C49 – Elaborar tarefas com pequeno tempo de duração;
- INT-C54 – Permitir acesso ao material apenas durante o experimento;
- INT-C64 – Realizar apenas uma execução com um tratamento;
- INT-C76 – Elaborar tarefas independentes uma das outras;
- INT-C80 – Realizar estudo piloto para verificar o aprendizado fornecido pelo material; e
- INT-C86 – Realizar treinamento para minimizar o aprendizado.

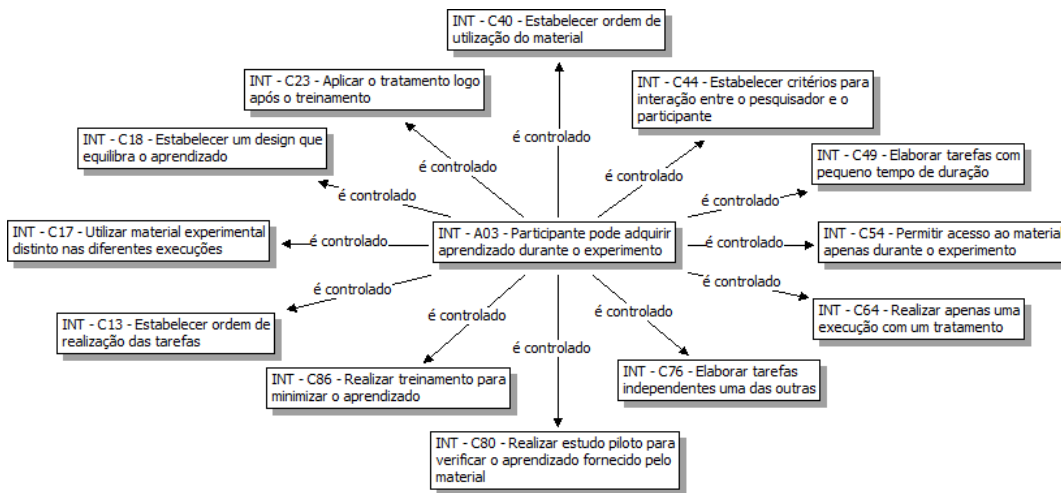


Figura 4.3. Ameaça INT-A03 e suas ações de controle.

As ações INT-C13, INT-C49 e INT-C76 visam minimizar o aprendizado fornecido pela realização das tarefas. Uma tarefa pode alterar o desempenho do participante. Do mesmo modo, as ações INT-C17, INT-C40, INT-C54 e INT-C80 visam minimizar o aprendizado fornecido pelo material experimental. As ações INT-C23 e INT-C64 têm como objetivo controlar o aprendizado fornecido pelo tratamento a ser aplicado. Por outro lado, a ação INT-C18 tem como objetivo minimizar o aprendizado que pode ocorrer através do design experimental. A ação INT-C44 visa controlar o aprendizado que pode ser fornecido pelo pesquisador, involuntariamente, quando há interação com os participantes. Por fim, a ação INT-C86 tem como objetivo estabelecer aprendizado no treinamento para que não ocorra durante o experimento. Assim, haverá um equilíbrio do aprendizado adquirido ao decorrer do experimento.

A Figura 4.4 apresenta a ameaça “*Diferenças no material experimental utilizado*”, INT-A04. Caso o material seja diferente entre os participantes, os resultados podem apresentar viés. Identificamos seis ações de controle para esta ameaça:

- INT-C03 – Utilizar o mesmo material experimental para todos os participantes;
- INT-C25 – Elaborar o material experimental o mais equivalente possível;
- INT-C31 – Revisar o material experimental (pessoas diferentes, se possível);
- INT-C50 – Selecionar especialistas para elaboração do material experimental;
- INT-C57 – Realizar estudo piloto para avaliar o material experimental; e
- INT-C59 – Selecionar material utilizado em estudos anteriores.

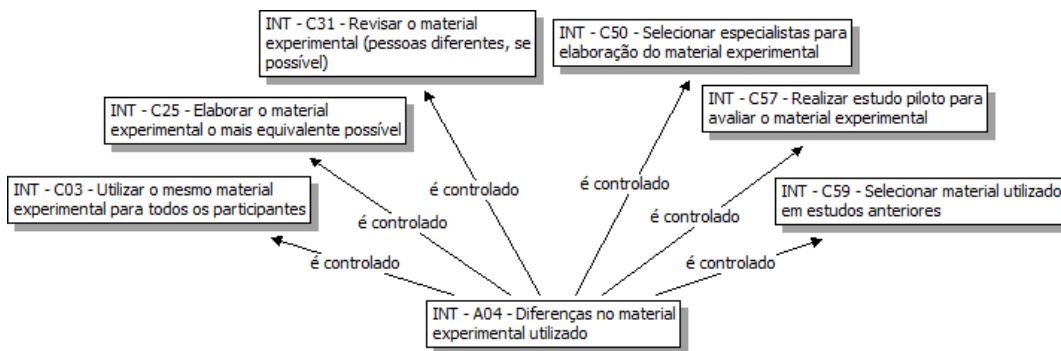


Figura 4.4. Ameaça INT-A04 e suas ações de controle.

A ação INT-C03 sugere que todos os participantes utilizem o mesmo material experimental. Dessa forma, caso o material experimental forneça viés, este será aplicado sobre todos os participantes. Caso seja necessário elaborar diferentes materiais, a ação INT-C25 propõe que as diferenças sejam pequenas, sem prejudicar o resultado do experimento. Além disso, o material deve ser revisado (INT-C31) e deve ser elaborado por especialistas (INT-C50). É desejável também a realização de estudo piloto para avaliar o material experimental (INT-C57). Por fim, a ação INT-C59 sugere a utilização de material experimental de estudos anteriores.

O problema de precisão nos dados coletados (INT-A05) é ilustrado na Figura 4.5. Os dados podem não ser precisos caso o participante seja responsável por fornecer os dados. Identificamos cinco ações de controle para esta ameaça:

- INT-C04 – Utilizar ferramenta para coleta de dados;
- INT-C16 – Atribuir recompensa ao desempenho do participante;
- INT-C24 – Confrontar dados coletados manualmente e automaticamente e excluir, caso haja inconsistência;
- INT-C58 – Solicitar que os participantes sejam precisos na coleta de dados; e
- INT-C65 – Confrontar dados coletados pelos participantes e por observadores e excluir, caso haja inconsistência.

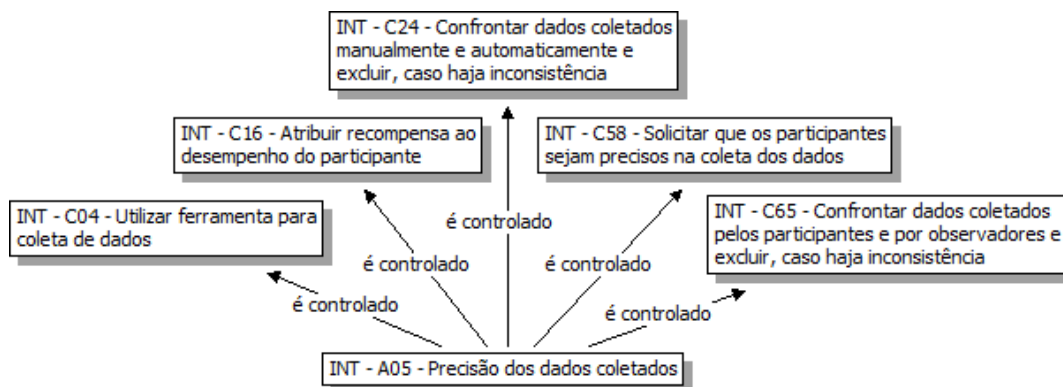


Figura 4.5. Ameaça INT-A05 e suas ações de controle.

Ao utilizar uma ferramenta para coleta de dados (INT-C04) a precisão é maior. A ação INT-C16 sugere fornecer recompensa ao desempenho do participante. Dessa forma o participante será mais cuidadoso ao informar os dados. As ações INT-C24 e INT-C65 têm como objetivo analisar os dados com base em duas perspectivas distintas. Dessa forma, se houver inconsistência, os dados devem ser descartados. Por fim, a ação INT-C58 sugere que os pesquisadores solicitem que o participante seja o mais preciso possível.

A Figura 4.6 apresenta a ameaça “*Diferenças entre os participantes relacionadas à habilidade*”. A ameaça INT-A06, assim como a INT-A01, ocorre a partir da seleção dos participantes. O nível de habilidade dos participantes possui variação. Como ações de controle para esta ameaça identificamos as seguintes:

- INT-C02 – Realizar treinamento para nivelar a habilidade dos participantes;
- INT-C05 – Atribuir os participantes aleatoriamente aos grupos;
- INT-C09 – Caracterizar a habilidade do participante através de questionário;
- INT-C10 – Agrupar os participantes de acordo com o nível de habilidade;
- INT-C32 – Atribuir aleatoriamente os tratamentos aos participantes;
- INT-C66 – Selecionar participantes utilizando como base o desempenho acadêmico; e
- INT-C81 – Caracterizar a habilidade do participante através de teste pré-experimento.

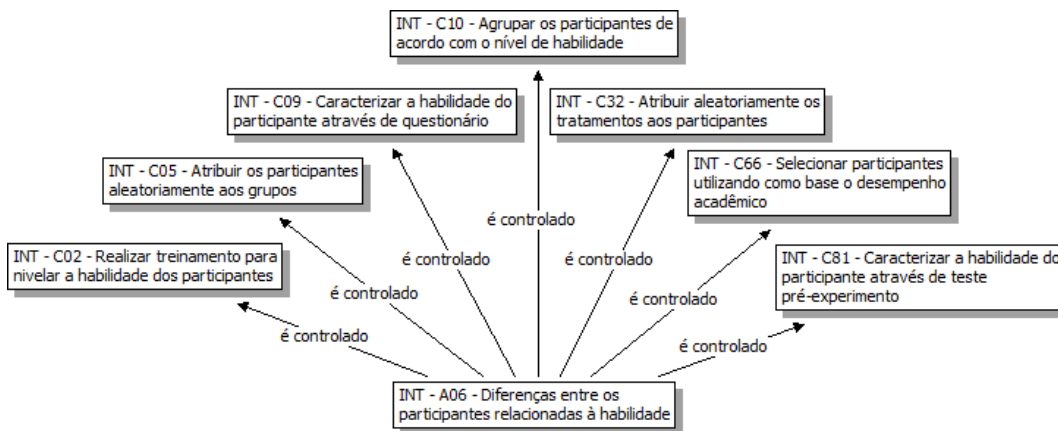


Figura 4.6. Ameaça INT-A06 e suas ações de controle.

As ações de controle INT-C09, INT-C66 e INT-C81 visam obter informações do participante para poder formar grupos balanceados (princípio de balanceamento). Por outro lado, as ações INT-C05 e INT-C32 utilizam o princípio de randomização. A ação INT-C10 tem por objetivo agrupar os participantes de acordo com os níveis de habilidade (princípio de blocagem). Por fim, a ação INT-C02 visa estabelecer equilíbrio entre os participantes.

A ameaça “*Participante pode abandonar o experimento*” (INT-A07) é apresentada na Figura 4.7. Para esta ameaça identificamos quatro ações de controle:

- INT-C07 – Atribuir os participantes aos grupos no dia do experimento;
- INT-C19 – Selecionar participantes voluntários;
- INT-C28 – Excluir dados incompletos ou incorretos; e
- INT-C42 – Atribuir recompensa para a participação no experimento.

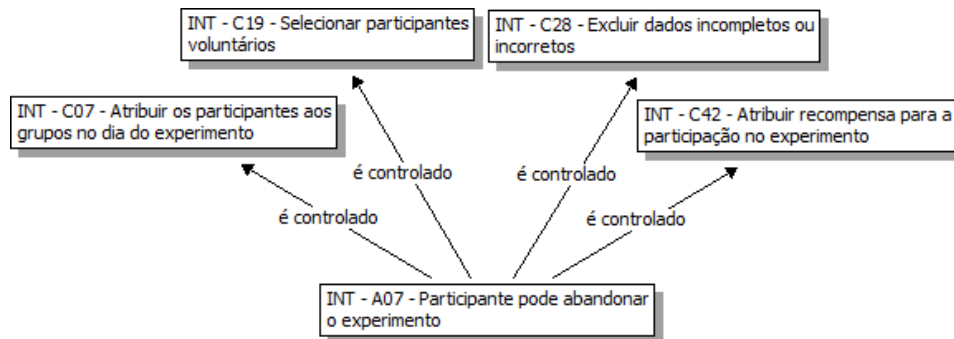


Figura 4.7. Ameaça INT-A07 e suas ações de controle.

A ação de controle INT-C07 visa criar os grupos experimentais no dia do experimento. Caso algum participante abandone o experimento, os grupos não serão prejudicados. A ação INT-C19 sugere a seleção de participantes voluntários. Assim, o risco de abandono é menor. Podemos também excluir os dados incompletos (INT-C28). Por fim, o responsável pelo experimento pode estabelecer recompensas para que os participantes realizem o experimento (INT-C42).

A Figura 4.8 apresenta a ameaça “*Comunicação entre os participantes durante o experimento*”, INT-A08. A troca de informação durante o experimento pode alterar o desempenho dos participantes. Identificamos cinco ações de controle para esta ameaça:

- INT-C35 – Distribuir os participantes com espaço suficiente entre eles;
- INT-C38 – Introduzir observadores no ambiente experimental;
- INT-C51 – Solicitar que os participantes não conversem durante o experimento;
- INT-C52 – Não permitir o compartilhamento de material durante o experimento; e
- INT-C84 – Excluir dados de participantes que conversarem.

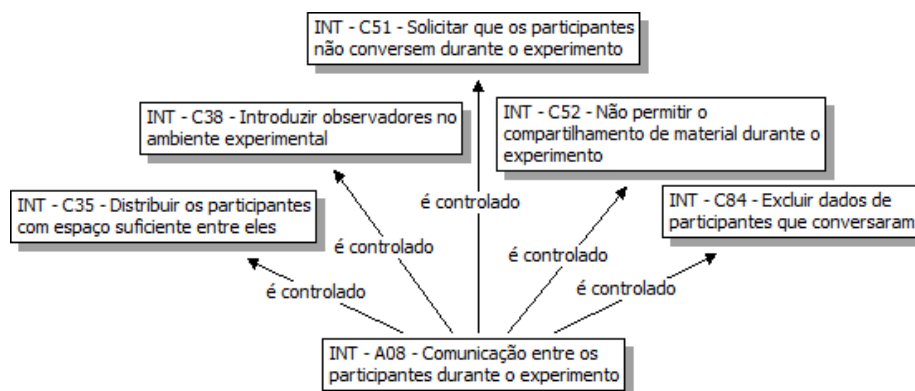


Figura 4.8. Ameaça INT-A08 e suas ações de controle.

Ao distribuir os participantes com espaço suficiente entre eles, o responsável pelo experimento evita o compartilhamento de informação (INT-C35). Outra ação de controle para esta ameaça é introduzir observadores no ambiente experimental (INT-C38). O pesquisador pode ainda solicitar que os participantes não conversem durante o experimento (INT-C51) e não permitir o compartilhamento de material (INT-C52). Por fim, a ação INT-C84 sugere que os dados dos participantes que trocaram informações sejam descartados.

Não identificamos ações de controle para a ameaça “*Atribuição de recompensa pela participação no experimento e não pelo desempenho*” (INT-A09). Ao estabelecer recompensa pela participação, os participantes podem não executar o experimento de forma séria e correta. Desse modo, o resultado pode não refletir a realidade.

A ameaça INT-A10 trata das diferentes motivações dos participantes. Assim, é uma ameaça à validade em potencial nos experimentos. Para esta ameaça identificamos sete ações de controle:

- INT-C08 – Participar do experimento é parte obrigatória de uma disciplina;
- INT-C16 – Atribuir recompensa ao desempenho do participante;
- INT-C19 – Selecionar participantes voluntários;
- INT-C20 – Selecionar participantes do contexto real do estudo;
- INT-C37 – Informar aos participantes que exercícios semelhantes podem ser realizados nas provas da disciplina;
- INT-C74 – Informar aos participantes que o tratamento é amplamente utilizado na indústria;
- e
- INT-C77 – Selecionar atividade que traz benefício aos participantes.

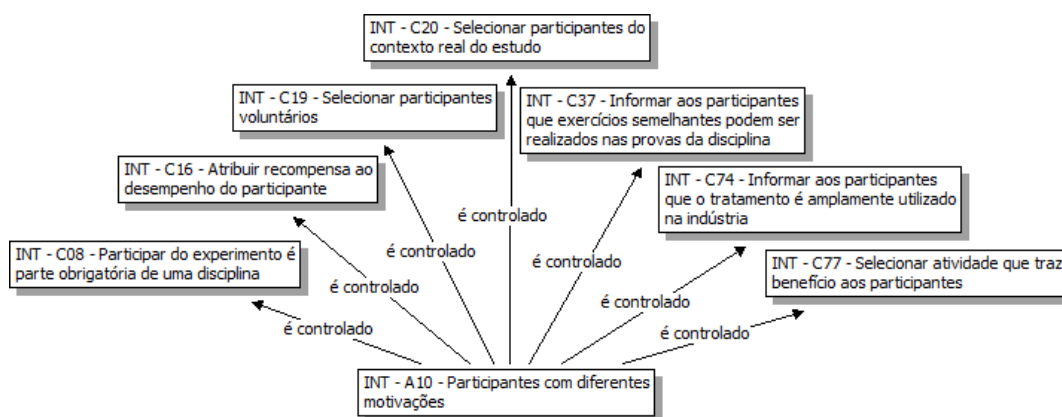


Figura 4.9. Ameaça INT-A10 e suas ações de controle.

A ação de controle INT-C08 trata da obrigatoriedade do participante em realizar o experimento. Dessa forma o participante estará apto a finalizar a disciplina. Outra forma de controle para a ameaça INT-A10 é atribuir recompensa ao desempenho do participante (INT-C16). O

pesquisador responsável pode também selecionar participantes voluntários (INT-C19) ou participantes do contexto real do estudo (INT-C20). Assim, espera-se que os participantes fiquem motivados para realização do experimento. Além disso, as ações INT-C37, INT-C74 e INT-C77 dizem respeito à utilização de procedimentos que podem ser úteis para os participantes no futuro.

A ameaça “*Fatores técnicos podem causar dificuldade aos participantes*” (INT-A11) é apresentada na Figura 4.10. Esta ameaça ocorre quando os participantes não possuem capacidade técnica suficiente para a realização do experimento. Identificamos três ações de controle para esta ameaça:

- INT-C12 – Excluir os dados dos participantes com deficiência técnica;
- INT-C14 – Selecionar pessoas com experiência no tópico em questão; e
- INT-C22 – Permitir que os participantes façam perguntas.

A ação INT-C12 sugere a exclusão dos dados dos participantes que apresentarem deficiência técnica. Para evitar esta ameaça, o pesquisador pode ainda selecionar pessoas com experiência no tópico a ser estudado (INT-C14) e permitir que os participantes façam perguntas para que possíveis dúvidas sejam sanadas (INT-C22).

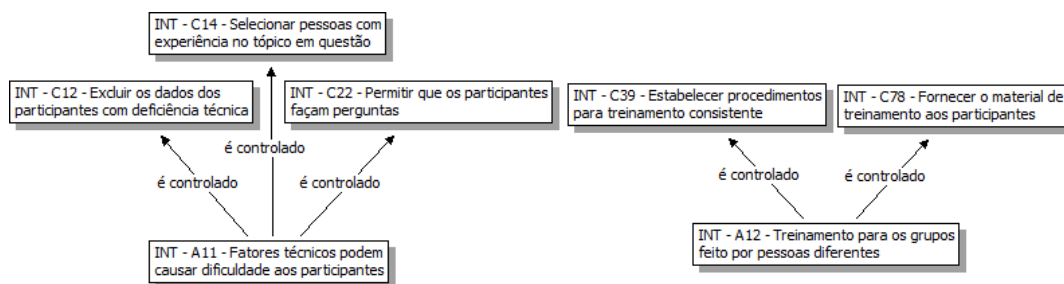


Figura 4.10. Ameaças INT-A11 e INT-A12 e suas ações de controle.

A Figura 4.10 apresenta também a ameaça “*Treinamento para os grupos feito por pessoas diferente*”, INT-A12. Quando os grupos são treinados por pessoas diferentes, o treinamento pode fornecer informações distintas para cada grupo. Para esta ameaça identificamos duas ações de controle:

- INT-C39 – Estabelecer procedimentos para treinamento consistente; e
- INT-C78 – Fornecer material de treinamento aos participantes.

Ao estabelecer procedimentos para condução do treinamento (INT-C39), as informações passadas aos participantes são consistentes. Dessa forma, o viés do treinamento realizado por pessoas diferentes é minimizado. Além disso, o material do treinamento pode ser fornecido aos participantes para que possíveis dúvidas possam ser discutidas posteriormente (INT-C78).

Não identificamos ações de controle para a ameaça “*Diferenças entre os participantes*

relacionadas ao gênero”, INT-A13. Esta ameaça ocorre quando o gênero do participante pode causar viés no experimento.

A ameaça “*Problemas de especificação do material experimental*” (INT-A14) é apresentada na Figura 4.11. Esta ameaça ocorre quando o material experimental apresenta inconsistência, ambiguidade, incorretude, incompletude, etc. Identificamos as seguintes ações de controle para esta ameaça:

- INT-C15 – Buscar o material na literatura;
- INT-C31 – Revisar o material experimental (pessoas diferentes, se possível);
- INT-C50 – Selecionar especialistas para elaboração do material experimental;
- INT-C57 – Realizar estudo piloto para avaliar o material experimental; e
- INT-C59 – Selecionar material utilizado em estudos anteriores.

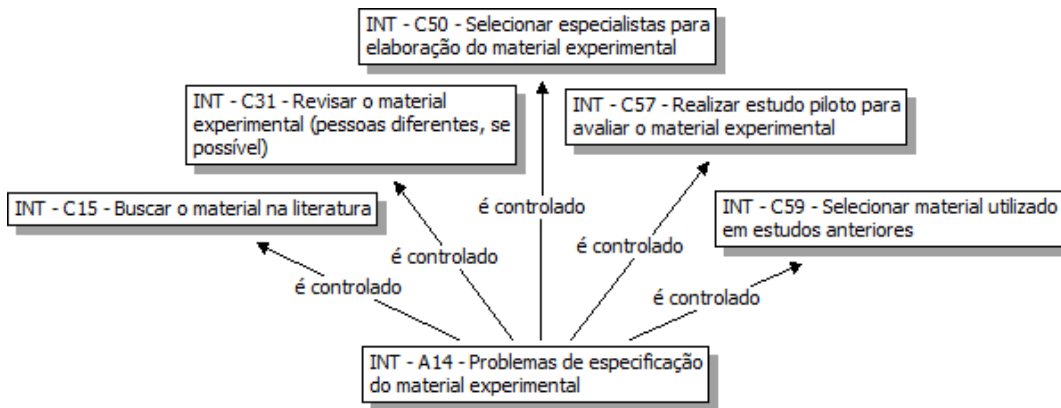


Figura 4.11. Ameaça INT-A14 e suas ações de controle.

A ação de controle INT-C15 e INT-C59 sugere que o material experimental a ser utilizado seja procurado na literatura, ou seja aproveitado de estudos anteriores. Este material já passou por avaliação e validação. O material deve também ser revisado, por pessoas diferentes, se possível (INT-C31). Além disso, é recomendável que o material seja elaborado por especialistas (INT-C50). Por fim, a ação de controle INT-C57 sugere que o material seja avaliado através de estudo piloto.

Não identificamos ações de controle para a ameaça “*Fatores que podem causar problemas no ambiente experimental*”, INT-A15. Esta ameaça ocorre quando os participantes são alocados a ambientes experimentais diferentes. Desse modo, fatores que ocorrerem nesses ambientes podem causar viés ao resultado do experimento.

A Figura 4.12 apresenta a ameaça “*Participante pode apresentar fadiga durante o experimento*”, INT-A16. Caso o experimento seja longo, o participante pode ficar cansado e entediado, prejudicando a execução do experimento. Para esta ameaça identificamos nove ações de controle:

- INT-C19 – Selecionar participantes voluntários;
- INT-C27 – Estabelecer tempo de duração não muito longo;
- INT-C30 – Estabelecer interação com os participantes;
- INT-C49 – Elaborar tarefas com pequeno tempo de duração;
- INT-C53 – Utilizar material experimental diferente durante as execuções;
- INT-C67 – Permitir intervalos entre as execuções;
- INT-C72 – Realizar execuções em dias consecutivos;
- INT-C73 – Possibilitar que o participante escolha o dia do experimento; e
- INT-C79 – Realizar estudo piloto para verificar se o tempo estabelecido causa fadiga.

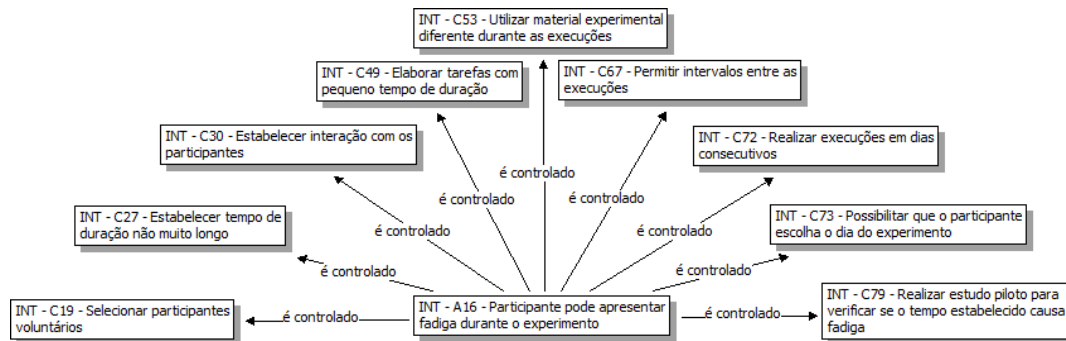


Figura 4.12. Ameaça INT-A16 e suas ações de controle.

A ação de controle INT-C19 sugere que os participantes selecionados sejam voluntários. Isso porque os participantes estarão realmente interessados no experimento. As ações INT-C27 e INT-C49 tem como objetivo limitar o tempo de duração do experimento (ou das sessões do experimento). É sugerido também que haja interação do pesquisador com os participantes (INT-C30), para evitar o tédio. Caso haja mais de uma execução, é sugerido que o material experimental seja diferente (INT-C53). O responsável pelo experimento pode permitir intervalos durante o experimento (INT-C67) ou realizar as execuções em dias consecutivos (INT-C72). Além disso, o pesquisador pode permitir que o participante escolha o dia de execução do experimento (INT-C73), fornecendo comodidade. Por fim, a ação de controle INT-C79 sugere a realização de estudo piloto para verificar se o tempo estabelecido para execução do experimento não causa fadiga ao participante.

Não identificamos ações de controle para a ameaça “*Fatos podem alterar o desempenho do participante durante experimentos com intervalos longos*”, INT-A17. Esta ameaça ocorre quando o experimento possui intervalos. Durante os intervalos podem ocorrer fatos que influenciam no desempenho do participante.

A ameaça “*Comunicação entre os participantes durante os intervalos*” (INT-A18) é apresentada na Figura 4.13. Diferentemente da ameaça INT-A08, a INT-A18 ocorre quando há troca

de informação durante os intervalos do experimento e não durante o experimento. Para esta ameaça identificamos duas ações de controle:

- INT-C54 – Permitir acesso ao material apenas durante o experimento; e
- INT-C68 – Solicitar que os participantes não conversem sobre o experimento durante os intervalos.

A ação de controle INT-C54 sugere que o material seja disponibilizado ao participante apenas durante as execuções. Dessa forma, o pesquisador evita que informações sobre o material sejam compartilhadas. Por fim, a ação INT-C68 sugere que o pesquisador solicite que os participantes não conversem sobre o experimento durante os intervalos. Assim, informações sobre aplicação de tratamento e uso do material experimental não serão compartilhadas.

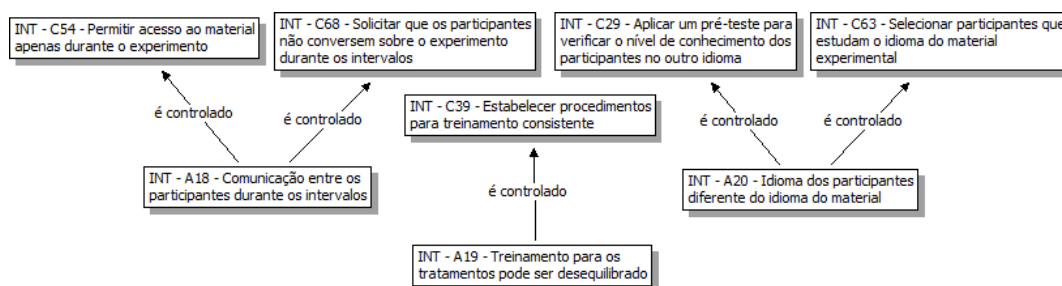


Figura 4.13. Ameaças INT-A18, INT-A19 e INT-A20 e suas ações de controle.

A ameaça “*Treinamento para os tratamentos pode ser desequilibrado*” (INT-A19) ocorre quando existem diferenças no treinamento, mesmo que o treinamento seja dado por apenas uma pessoa. Diferenças de material, tempo e forma de explicação no treinamento podem afetar o resultado do experimento. Identificamos apenas uma ação de controle para esta ameaça:

- INT-C39 – Estabelecer procedimentos para treinamento consistente.

Assim como na ameaça INT-A12, ao estabelecer procedimentos para condução do treinamento (INT-C39), as informações passadas aos participantes são consistentes. Dessa forma, o viés do treinamento realizado é minimizado.

Caso o idioma dos participantes seja diferente do idioma do material experimental (INT-A20), existe um risco para o experimento. Para esta ameaça identificamos duas ações de controle:

- INT-C29 – Aplicar um pré-teste para verificar o nível de conhecimento dos participantes no outro idioma; e
- INT-C63 – Selecionar participantes que estudam o idioma do material experimental.

A ação INT-C29 visa obter informações sobre o conhecimento dos participantes. Dessa forma, o pesquisador tem evidências de que o participante pode ou não participar do experimento.

Por outro lado, a ação INT-C63 sugere que o pesquisador selecione os participantes que fazem curso ou já estudaram o idioma do material experimental.

Não identificamos ações de controle para as ameaças “*Seleção dos participantes com base no desempenho em atividade passadas*” (INT-A21) e “*Aplicação de pré-teste pode alterar o desempenho do participante*” (INT-A22). A primeira ameaça ocorre quando o pesquisador seleciona os participantes com base em dados históricos. Dessa forma, o pesquisador pode causar viés ao selecionando pessoas com melhores resultados. A segunda ameaça ocorre quando uma atividade pré-experimento pode influenciar na execução do experimento. Nesse caso, o participante pode obter informações que possam auxiliar o seu desempenho.

A Figura 4.14 apresenta as seguintes ameaças: “*Treinamento realizado por apenas uma pessoa*” (INT-A23), “*Alteração do pacote experimental original, em caso de replicação*” (INT-A24) e “*Dados coletados incompletamente ou incorretamente*” (INT-A25). A primeira ameaça ocorre quando inconscientemente (ou conscientemente) o pesquisador influencia no treinamento. Assim, o treinamento para um grupo pode ser melhor que para outro. A segunda ameaça ocorre quando o pesquisador conduz uma replicação. Caso o pacote experimental original sofra alterações, o resultado pode não ser causado pela aplicação do tratamento. A última ameaça ocorre quando o participante não finaliza o experimento ou quando os procedimentos para execução do experimento não são seguidos.

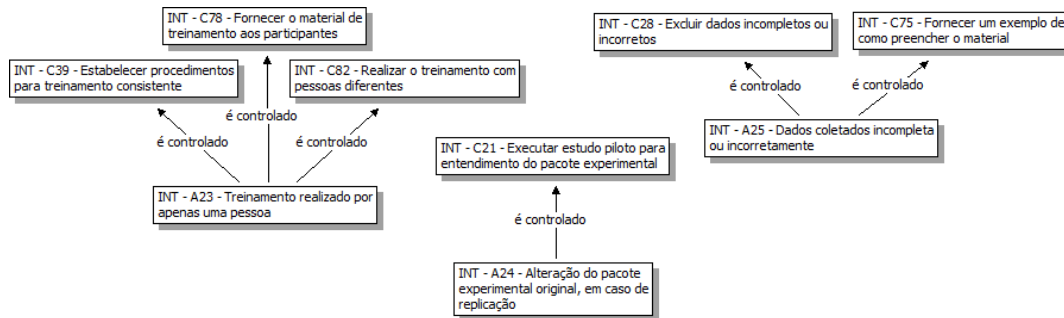


Figura 4.14. Ameaças INT-A23, INT-A24 e INT-A25 e suas ações de controle.

Para a ameaça INT-A23 identificamos três ações de controle:

- INT-C39 – Estabelecer procedimentos para treinamento consistente;
- INT-C78 – Fornecer o material de treinamento aos participantes; e
- INT-C82 – Realizar treinamento com pessoas diferentes.

Assim como para a ameaça INT-A12, as ações de controle INT-C39 e INT-C78 podem ser aplicadas para minimizar a ameaça INT-A23. Por último, é sugerido que o treinamento seja realizado por pessoas diferentes (INT-C82).

Identificamos apenas uma ação de controle para a ameaça INT-A24: “*Executar estudo piloto para entendimento do pacote experimental*” (INT-C21). Dessa forma, o pesquisador pode compreender o material. Com isso, alterações no material podem ser realizadas sem que comprometa o experimento.

Para a ameaça INT-A25 identificamos duas ações de controle:

- INT-C28 – Excluir dados incompletos ou incorretos; e
- INT-C75 – Fornecer um exemplo de como preencher o material.

A ação INT-C28 sugere que os dados incompletos ou incorretos sejam descartados. Por outro lado, a ação INT-C75 sugere que o pesquisador explique e exemplifique a forma de utilização do material para coleta dos dados.

As ameaças “*Atribuição de tempo para execução do experimento*” (INT-A26) e “*Diferenças entre as tarefas*” (INT-A27) são apresentadas na Figura 4.15. A primeira ocorre quando a execução do experimento possui tempo limitado. Dessa forma, o participante pode realizar as tarefas rapidamente dentro do tempo estipulado. Assim, a tarefa pode não ser executada corretamente. A última ocorre quando as tarefas a serem realizadas pelos participantes são diferentes. Assim, o resultado pode não ser causado pela aplicação do tratamento, mas sim pelas diferentes tarefas.

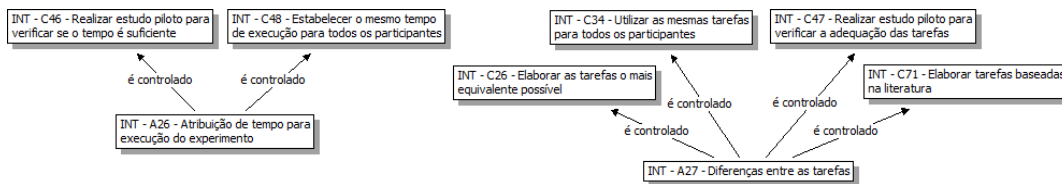


Figura 4.15. Ameaças INT-A26 e INT-A27 e suas ações de controle.

Identificamos duas ações de controle para a ameaça INT-A26:

- INT-C46 – Realizar estudo piloto para verificar se o tempo é suficiente; e
- INT-C48 – Estabelecer o mesmo tempo de execução para todos os participantes.

A primeira ação sugere que seja realizado um estudo piloto para verificar a adequação do tempo estipulado. Dessa forma, o pesquisador pode avaliar se o tempo deve ser alterado ou se as tarefas serão modificadas. A última sugere que todos os participantes tenham o mesmo tempo para executar o experimento. Assim, o efeito sobre o resultado será equivalente para cada participante.

Para a ameaça INT-A27 identificamos quatro ações de controle:

- INT-C26 – Elaborar as tarefas o mais equivalente possível;
- INT-C34 – Utilizar as mesmas tarefas para todos os participantes;

- INT-C47 – Realizar estudo piloto para verificar a adequação das tarefas; e
- INT-C71 – Elaborar tarefas baseadas na literatura.

A ação INT-C26 sugere que as tarefas sejam equivalentes, com o mínimo de diferenças possíveis. Podemos também utilizar as mesmas tarefas para todos os participantes, quando for o caso (INT-C34). Além disso, a ação INT-C47 sugere a realização de estudo piloto para verificar a adequação das diferentes tarefas. Por fim, a ação INT-C71 sugere que o pesquisador elabore tarefas baseadas na literatura.

A Figura 4.16 apresenta as seguintes ameaças:

- INT-A28 – Rivalidade entre os grupos;
- INT-A29 – Os participantes podem procurar respostas em outras fontes;
- INT-A30 – Participantes podem ter realizado estudos semelhantes anteriormente; e
- INT-A31 – Ordem de apresentação do material aos participantes.

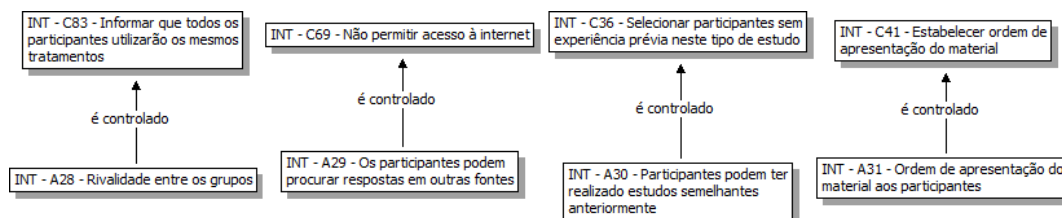


Figura 4.16. Ameaças INT-A28, INT-A29, INT-A30 e INT-A31 e suas ações de controle.

A ameaça INT-A28 ocorre quando um grupo dá o seu melhor ao utilizar um tratamento convencional, para mostrar que é competitivo. Dessa forma, o resultado apresenta viés. Para esta ameaça identificamos a ação de controle “*Informar que todos os participantes utilizarão os mesmos tratamentos*” (INT-C83). Esta ação sugere que o pesquisador informe que os tratamentos aplicados no experimento serão utilizados por todos os participantes. Assim, evita-se a rivalidade entre os grupos.

A ameaça INT-A29 ocorre quando o participante, por exemplo, realiza o experimento em casa. Dessa forma, o participante pode utilizar outras fontes de informação para executar as tarefas. Como ação de controle para esta ameaça temos “*Não permitir acesso à internet*” (INT-C69).

A ameaça INT-A30 ocorre quando os participantes do experimento realizaram estudos semelhantes anteriormente. Dessa forma, a experiência adquirida pode auxiliar o desempenho dos participantes. Identificamos a ação de controle “*Selecionar participantes sem experiência prévia neste tipo de estudo*” (INT-C36). Esta ação sugere que o pesquisador obtenha informações sobre experiência dos participantes em experimentos.

A ameaça INT-A31 ocorre quando a ordem de apresentação do material experimental pode

influenciar o desempenho do participante. Como ação de controle para esta ameaça identificamos “*Estabelecer ordem de apresentação do material*” (INT-C41). Dessa forma, a ameaça pode ser minimizada.

Não identificamos ações de controle para a ameaça “*Parte dos participantes recebe recompensa pela participação e outra parte não*” (INT-A32). Esta ameaça ocorre quando grupos distintos de participantes, por exemplo, participantes que são profissionais e participantes que são estudantes. É dada recompensa para convencer os profissionais a realizarem o experimento.

“*Diferenças entre os participantes relacionadas à personalidade*” é a ameaça INT-A33 (Figura 4.17). Esta ameaça ocorre quando a personalidade influencia o contexto do experimento. Para esta ameaça identificamos duas ações de controle:

- INT-C61 – Agrupar os participantes de acordo com o nível de personalidade; e
- INT-C62 – Caracterizar o nível de personalidade através de questionário.

A ação de controle INT-C61 tem por objetivo agrupar os participantes de acordo com os níveis de personalidade (princípio de blocagem). Por outro lado, a ação INT-C62 visa obter informações do participante para poder formar grupos balanceados (princípio de balanceamento).

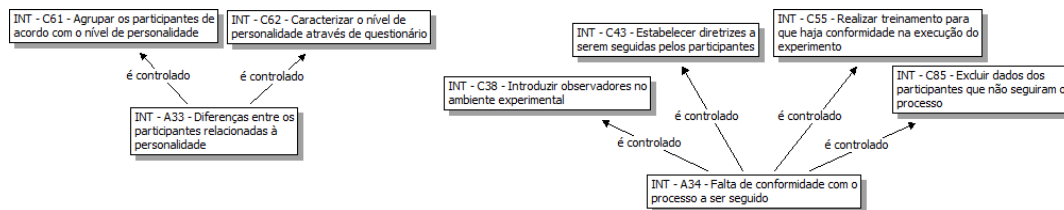


Figura 4.17. Ameaças INT-A33 e INT-A34 e suas ações de controle.

A Figura 4.17 apresenta também a ameaça “*Falta de conformidade com o processo a ser seguido*” (INT-A34). Se o participante não seguir processo corretamente (conforme estabelecido no planejamento do experimento), o resultado do experimento será prejudicado. Como ações de controle para esta ameaça, identificamos:

- INT-C38 – Introduzir observadores no ambiente experimental;
- INT-C43 – Estabelecer diretrizes a serem seguidas pelos participantes;
- INT-C55 – Realizar treinamento para que haja conformidade na execução do experimento; e
- INT-C85 – Excluir dados dos participantes que não seguiram o processo.

A ação INT-C38 sugere a introdução de pesquisadores no ambiente experimental. O pesquisador deve intervir caso perceba alguma inconformidade. Pode-se também estabelecer diretrizes bem definidas (INT-C43) para que os participantes não desviem do processo. Além disso, os participantes podem receber treinamento para que a execução do experimento não seja

prejudicada (INT-C55). Por fim, a ação INT-C85 sugere que o descarte de dados dos participantes que apresentaram inconformidade.

A Figura 4.18 apresenta as ameaças “*Treinamento pode ser insuficiente*” (INT-A35) e “*Adequação das tarefas a serem executadas no experimento*” (INT-A36). A primeira ameaça ocorre quando o treinamento não fornece todas as informações sobre um tratamento, por exemplo. Desse modo, a execução do experimento pode ser comprometida. A última ameaça ocorre quando as tarefas a serem realizadas não são adequadas para o experimento. Por exemplo, podem ser tarefas de outro contexto. Desse modo, o resultado não corresponderá à realidade.

Para a ameaça INT-A35 identificamos a ação de controle “*Permitir que os participantes façam perguntas*” (INT-C22). Desse modo, os participantes podem tirar possíveis dúvidas sobre a realização do experimento. Para a ameaça INT-A36 também identificamos apenas uma ação de controle: “*Realizar estudo piloto para verificar a adequação das tarefas*” (INT-C47). Com a avaliação através do estudo piloto, o pesquisador pode decidir se altera ou mantém as tarefas experimentais.

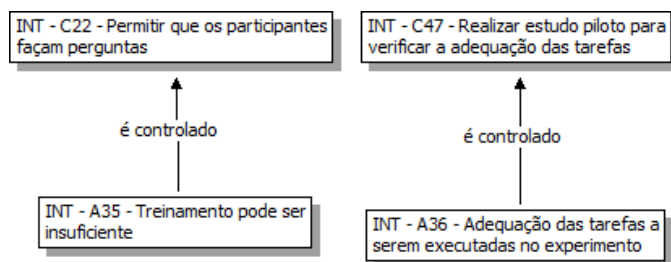


Figura 4.18. Ameaças INT-A35 e INT-A36 e suas ações de controle.

A seguir, serão apresentadas as ameaças à validade externa e as ações de controle identificadas na SLR.

4.1.2. Ameaças à Validade Externa e Ações de Controle

A seguir, apresentaremos as ameaças à validade externa que foram identificadas através da SLR. Identificamos um total de 9 ameaças à validade externa e 23 ações de controle.

A primeira ameaça à validade externa é “*Representatividade dos participantes*”, EXT-A01 (Figura 4.19). Esta ameaça ocorre quando a amostra experimental não representa a população para a qual se quer generalizar o experimento. A amostra pode não representar a população em termos de conhecimento, experiência, habilidade, quantidade, etc. Identificamos seis ações de controle para esta ameaça:

- EXT-C01 – Selecionar participantes referentes ao contexto real do estudo;
- EXT-C02 – Selecionar participantes com conhecimento próximo do contexto real do estudo;

- EXT-C04 – Selecionar participantes próximos de iniciar carreira no contexto real do estudo;
- EXT-C09 – Selecionar participantes que tenham experiência no contexto real do estudo;
- EXT-C15 – Selecionar participantes heterogêneos; e
- EXT-C23 – Realizar treinamento para os participantes.

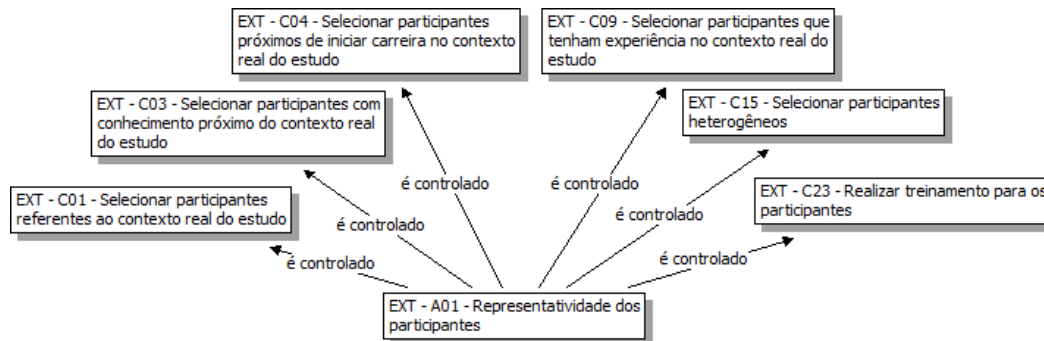


Figura 4.19. Ameaça EXT-A01 e suas ações de controle.

As ações de controle EXT-C01, EXT-C02, EXT-C04 e EXT-C09 sugerem que o pesquisador selecione participantes que tenham determinada experiência no contexto real do estudo: (1) participantes envolvidos no contexto real do estudo; (2) participantes que já se envolveram no contexto real do estudo; e (3) participantes com conhecimento de pessoas que fazem parte do contexto real do estudo. A ação EXT-C15 visa selecionar participantes com nível de conhecimento diferente. Dessa forma, a população pode ser mais bem representada. Por último, a ação EXT-C23 sugere que os participantes selecionados sejam treinados para adquirir experiência semelhante ao de pessoas que estão envolvidas no contexto real do estudo.

A Figura 4.20 apresenta a ameaça “*Representatividade dos artefatos experimentais*” (EXT-A02). Esta ameaça ocorre quando os artefatos não possuem, por exemplo, tamanho e complexidade representativos. Identificamos cinco ações de controle para esta ameaça:

- EXT-C02 – Buscar artefato experimental da literatura;
- EXT-C12 – Selecionar artefatos experimentais do contexto real do estudo;
- EXT-C16 – Selecionar artefatos experimentais mais comuns utilizados no contexto real do estudo, de acordo com pesquisa;
- EXT-C17 – Elaborar artefatos experimentais baseados no contexto real do estudo; e
- EXT-C20 – Enviar os artefatos para análise de especialistas.

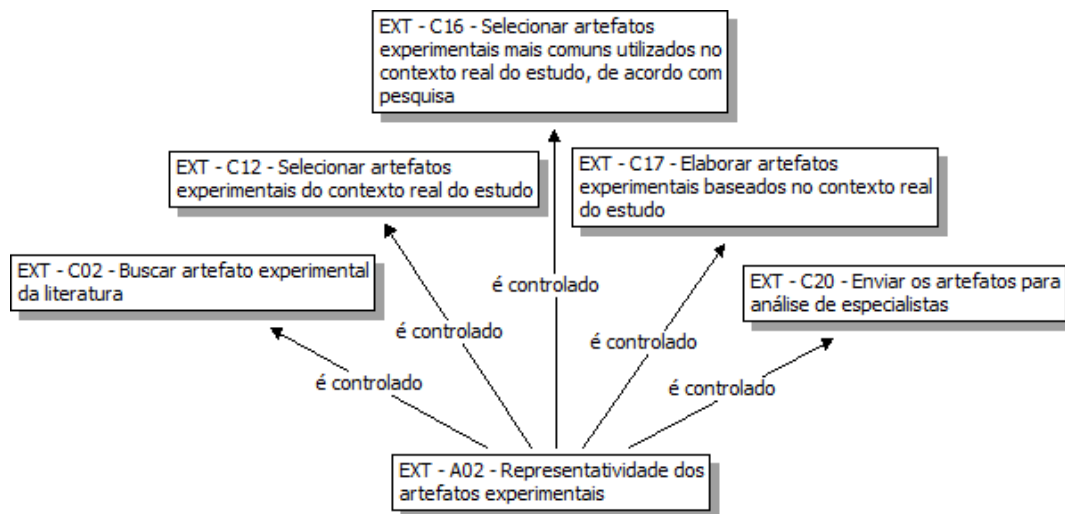


Figura 4.20. Ameaça EXT-A02 e suas ações de controle.

A ação EXT-C02 sugere que o artefato seja selecionado a partir da literatura. Artefatos apresentados na literatura são válidos e representativos em algumas situações. Por outro lado, o pesquisador pode selecionar artefatos diretamente do contexto real do estudo (EXT-C12). Além disso, a ação EXT-C16 sugere uma pesquisa para identificar os artefatos comumente utilizados no contexto real do estudo. Assim, a representatividade é mais elevada. O pesquisador pode também elaborar artefatos baseados nos que são utilizados no contexto real (EXT-C17). Por fim, especialistas poderão verificar a representatividade dos artefatos (EXT-C20).

A ameaça “*Representatividade do domínio dos artefatos experimentais*” (EXT-A03) é apresentada na Figura 4.21. Como ação de controle para esta ameaça, identificamos: “*Utilizar artefatos experimentais com domínios diferentes*” (EXT-C05). Utilizar artefatos de diferentes domínios, principalmente em replicações, faz com que a representatividade do domínio aumente. A Figura 20 apresenta também a ameaça “*Representatividade do processo*” (EXT-A04). Esta ameaça ocorre quando o processo utilizado no contexto do experimento não representa o que ocorre na realidade. Identificamos três ações de controle para esta ameaça:

- EXT-C13 – Selecionar o processo (equivalente) utilizado no contexto real do estudo;
- EXT-C18 – Explicar atividades realizadas anteriormente que compõem o processo; e
- EXT-C19 – Utilizar processo descrito na literatura.

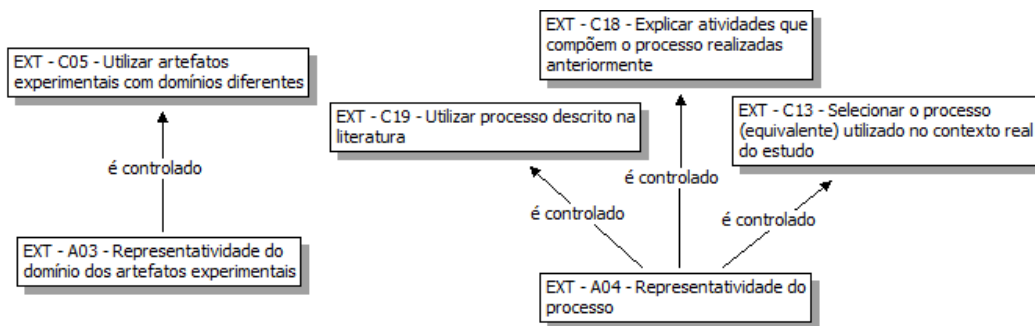


Figura 4.21. Ameaças EXT-A03 e EXT-A04 e suas ações de controle.

A ação EXT-C13 sugere que o processo a ser aplicado no estudo seja utilizado no contexto real, ou equivalente. Caso seja necessário utilizar apenas partes do processo, atividades anteriores devem ser explicadas aos participantes (EXT-C18). Por fim, a ação EXT-C19 sugere que o processo utilizado no estudo seja selecionado a partir da literatura. Processos que foram utilizados em outros estudos são, em sua maioria, representativos do contexto real.

A Figura 4.22 apresenta as ameaças “*Representatividade das tarefas*” (EXT-A05) e “*Representatividade dos tratamentos utilizados*” (EXT-A06). A primeira ocorre quando as tarefas a serem realizadas no experimento não representam as que são desempenhadas no contexto real. A última ocorre quando os tratamentos aplicados não são os mais utilizados no contexto real do estudo. Dessa forma, a representatividade dos tratamentos é comprometida.

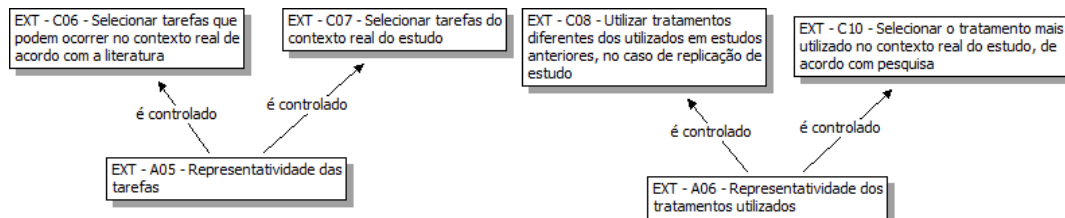


Figura 4.22. Ameaças EXT-A05 e EXT-A06 e suas ações de controle.

Para a ameaça EXT-A05 identificamos duas ações de controle:

- EXT-C06 – Selecionar tarefas que podem ocorrer no contexto real de acordo com a literatura; e
- EXT-C07 – Selecionar tarefas do contexto real do estudo.

A ação EXT-C06 sugere que o pesquisador realize uma pesquisa para identificar tarefas que ocorrem no contexto real do estudo. Por outro lado, a ação EXT-C07 sugere que as tarefas a serem realizadas no estudo devem ser do contexto real.

Identificamos também duas ações de controle para a ameaça EXT-A06:

- EXT-C08 – Utilizar tratamentos diferentes dos utilizados em estudos anteriores, no caso de replicação de estudo; e

- EXT-C10 – Selecionar o tratamento mais utilizado no contexto real do estudo, de acordo com pesquisa.

Caso o estudo seja uma replicação, o pesquisador deve utilizar tratamentos diferentes dos utilizados nos estudos anteriores (EXT-C08). Assim, a representatividade dos tratamentos aumenta. Do mesmo modo, o pesquisador pode realizar uma pesquisa para identificar o tratamento mais utilizado no contexto real (EXT-C10).

A Figura 4.23 apresenta as seguintes ameaças: “*Representatividade dos defeitos inseridos*” (EXT-A07), “*Representatividade dos requisitos utilizados*” (EXT-A08) e “*Representatividade do ambiente experimental*” (EXT-A09). A primeira ocorre quando é necessário inserir defeitos em artefatos (contexto de inspeção, por exemplo). Os defeitos inseridos nos artefatos podem não ser representativos. A segunda ocorre quando é utilizado documento de requisitos como artefato experimental. Os requisitos utilizados podem não ser representativos. A última ameaça ocorre quando o ambiente experimental não corresponde ao que ocorre no contexto real do estudo.

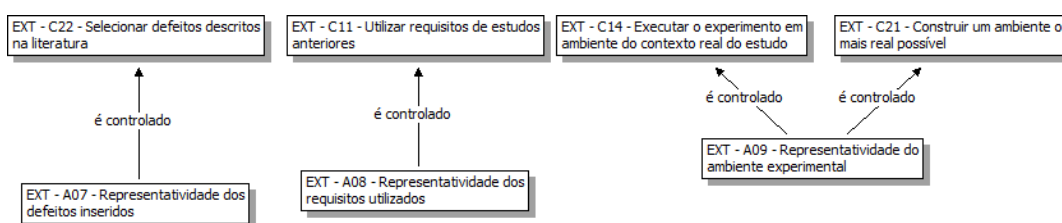


Figura 4.23. Ameaças EXT-A07, EXT-A08 e EXT-A09 e suas ações de controle.

Para a ameaça EXT-A07 identificamos a ação de controle “*Selecionar defeitos descritos na literatura*” (EXT-C22). Acredita-se que os defeitos reportados na literatura sejam representativos do que ocorrem no contexto real. Identificamos também apenas uma ação de controle para a ameaça EXT-A08: “*Utilizar requisitos de estudos anteriores*” (EXT-C11). Assim como os defeitos, acredita-se que os requisitos utilizados em estudos anteriores são representativos do contexto real.

Identificamos duas ações de controle para a ameaça EXT-A09:

- EXT-C14 – Executar o experimento em ambiente do contexto real do estudo; e
- EXT-C21 – Construir um ambiente o mais real possível.

A ação EXT-C14 sugere que o experimento seja realizado no ambiente real do estudo. Dessa forma, o ambiente é representativo. Por outro lado, o pesquisador pode construir um ambiente baseado no ambiente real (EXT-C21). Ambientes simulados podem ser representativos de ambientes reais. A seguir, serão apresentadas as ameaças à validade de constructo e as ações de controle identificadas na SLR.

4.1.3. Ameaças à Validade de Constructo e Ações de Controle

A seguir, apresentaremos as ameaças à validade de constructo que foram identificadas através da SLR. Identificamos um total de 10 ameaças à validade de constructo e 20 ações de controle.

As duas primeiras ameaças à validade de constructo são apresentadas na Figura 4.24. A primeira, “*Os tratamentos podem não representar a causa do resultado*” (COT-A01), ocorre quando o tratamento aplicado no experimento não representa a causa do resultado. Identificamos apenas uma ação de controle para esta ameaça: “*Utilizar tratamentos descritos na literatura*”. Tratamentos descritos na literatura apresentam maior confiabilidade.

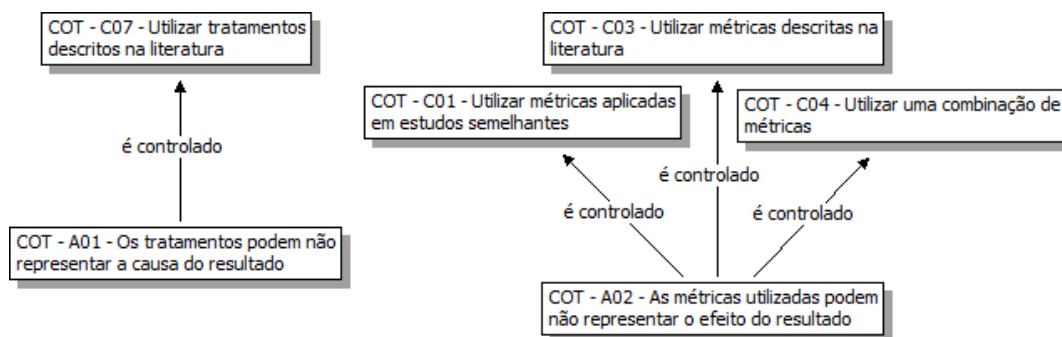


Figura 4.24. Ameaças COT-A01 e COT-A02 e suas ações de controle.

A ameaça “*As métricas utilizadas podem não representar o efeito do resultado*” (COT-A02) ocorre quando a métrica aplicada no experimento não representa o efeito do resultado. Identificamos três ações de controle para esta ameaça:

- COT-C01 – Utilizar métricas aplicadas em estudos semelhantes;
- COT-C03 – Utilizar métricas descritas na literatura; e
- COT-C04 – Utilizar uma combinação de métricas.

As ações COT-C01 e COT-C03 sugerem que as métricas utilizadas sejam selecionadas a partir de estudos anteriores ou a partir da literatura. Estes tipos de métricas são bem definidas e, geralmente, são validadas através de experimentos. Por fim, a combinação de métricas pode ser utilizada para calcular o efeito do resultado (COT-C04).

A ameaça “*Uso de um tratamento favorece o tratamento a ser usado posteriormente*” (COT-A03) ocorre quando o participante utiliza mais de um tratamento no experimento (Figura 4.25). Assim, o tratamento utilizado no início influencia o desempenho do participante ao utilizar os tratamentos posteriores. Para esta ameaça identificamos quatro ações de controle:

- COT-C02 – Estabelecer intervalos entre as aplicações dos tratamentos;
- COT-C15 – Aplicar apenas um tratamento por participante;

- COT-C17 – Estabelecer ordem de utilização dos tratamentos; e
- COT-C18 – Utilizar tratamentos que não influenciam um no outro.

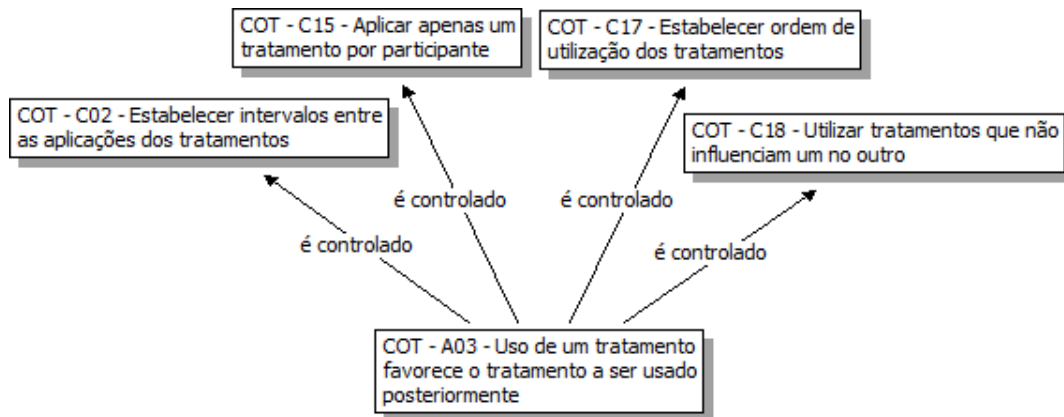


Figura 4.25. Ameaça COT-A03 e suas ações de controle.

A ação de controle COT-C02 sugere que exista intervalo entre as aplicações do tratamento. Desse modo, acredita-se que o tratamento utilizado no início não interfira no que será utilizado posteriormente. Para controlar a ameaça COT-A03 podemos também atribuir apenas um tratamento para cada participante (COT-C15). A ação COT-C17 sugere que seja estabelecida uma ordem de utilização dos tratamentos, de modo que os tratamentos que menos influenciam os demais sejam aplicados primeiro. Por último, podemos utilizar tratamentos que não causam influência, dessa forma a ameaça COT-A03 é controlada.

A Figura 4.26 apresenta a ameaça “*Participante age de forma diferente ao ser observado*” (COT-A04). O participante tem o desempenho afetado quando é observado por outras pessoas. Identificamos quatro ações de controle para esta ameaça:

- COT-C08 – Observar todos os participantes, dessa forma o efeito é equivalente;
- COT-C13 – Realizar atividade piloto para adaptar os participantes ao ambiente;
- COT-C14 – Elaborar tarefa desafiadora, mantendo os participantes concentrados; e
- COT-C20 – Não avaliar o desempenho do participante, mas sim a participação.

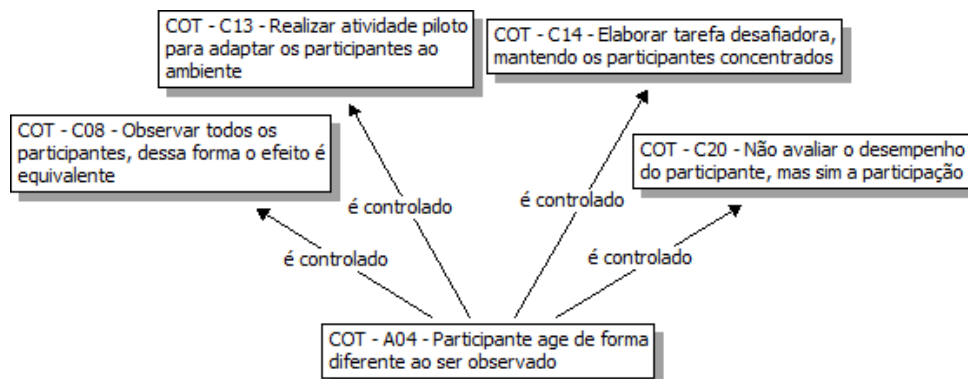


Figura 4.26. Ameaça COT-A04 e suas ações de controle.

A ação de controle COT-C08 sugere que todos os participantes sejam observados, assim se existir, o efeito ocorrerá sobre toda a amostra. Por outro lado, pode-se realizar atividade piloto para adaptar os participantes ao ambiente experimental (COT-C13). Dessa forma, os participantes não se sentirão (ou se sentirão menos) intimidados durante o experimento. Podemos também elaborar tarefas experimentais que mantenham os participantes concentrados (COT-C14). Com isso os participantes podem acabar esquecendo de que estão sendo observados. Por fim, a ação COT-C20 sugere que seja avaliada a participação do participante e não o seu desempenho. Dessa forma, a apreensão por estar sendo avaliado é minimizada.

A ameaça “*Pesquisador pode interferir no resultado do estudo*” (COT-A05) é apresentada na Figura 4.27. Esta ameaça ocorre quando consciente ou inconscientemente o pesquisador afeta o resultado do estudo. Para esta ameaça identificamos duas ações de controle:

- COT-C05 – Incluir pesquisadores independentes no experimento; e
- COT-C11 – Não permitir interação do pesquisador interessado no resultado com os participantes.

A ação COT-C05 sugere a inserção de pesquisadores independentes no experimento. Como estes não estão interessados no resultado da aplicação dos tratamentos, a ameaça é minimizada. Por outro lado, a ação COT-C11 sugere que os pesquisadores interessados não interajam com os participantes. Assim, nenhuma informação que possa afetar o desempenho do participante será transmitida.

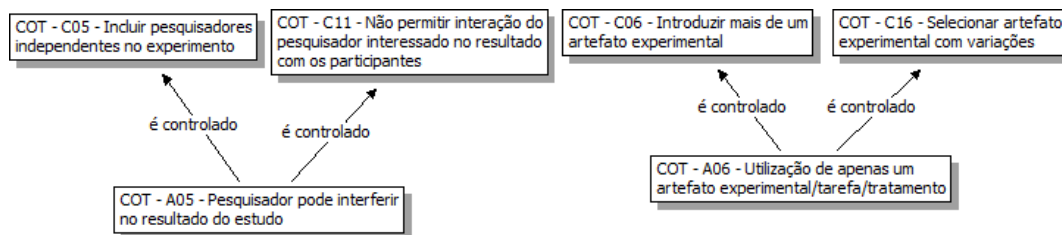


Figura 4.27. Ameaças COT-A05 e COT-A06 e suas ações de controle.

A Figura 4.27 apresenta também a ameaça “*Utilização de apenas um artefato experimental/tarefa/tratamento*” (COT-A06). Esta ameaça ocorre porque ao utilizar apenas um tratamento, por exemplo, a construção da causa fica sub-representada. Identificamos duas ações de controle para esta ameaça:

- COT-C06 – Introduzir mais de um artefato experimental; e
- COT-C16 – Selecionar artefato experimental com variações.

As ações COT-C06 e COT-C16 estão relacionadas aos artefatos experimentais. Estas sugerem que o pesquisador introduza mais de um artefato ou que crie variações de um artefato.

Dessa forma, a construção da causa é maximizada.

A ameaça “*Utilização de apenas uma métrica para o resultado*” (COT-A07) é apresentada na Figura 4.28. Esta ameaça ocorre quando apenas uma métrica é utilizada para checar o resultado. Se a métrica possuir viés, este não será detectado. Com a utilização de mais de uma métrica, o resultado pode ser verificado através de cross-checking. Para esta ameaça identificamos a ação de controle “*Utilizar uma combinação de métricas*” (COT-C04). Dessa forma, poderemos aplicar o cross-checking.

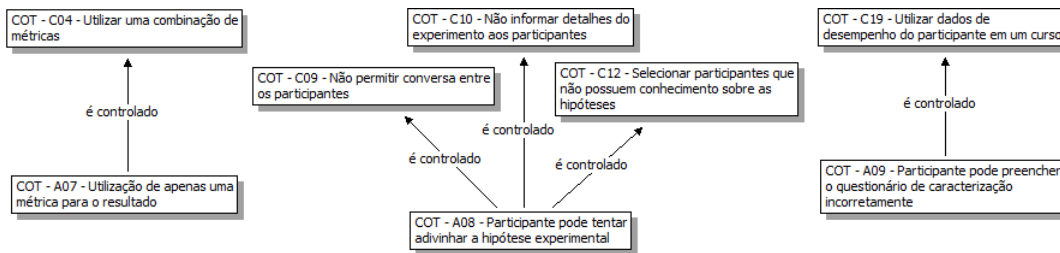


Figura 4.28. Ameaças COT-A07, COT-A08 e COT-A09 e suas ações de controle.

A ameaça “*Participante pode tentar adivinhar a hipótese experimental*” COT-A08 também é apresentada na Figura 4.28. Esta ameaça ocorre quando o participante altera seu comportamento com base no que ele acredita ser correto. Identificamos três ações de controle para esta ameaça:

- COT-C09 – Não permitir conversa entre os participantes;
- COT-C10 – Não informar detalhes do experimento aos participantes; e
- COT-C12 – Selecionar participantes que não possuem conhecimento sobre as hipóteses.

A ação de controle COT-C09 sugere que não seja permitida conversa entre os participantes. Desse modo, as opiniões dos participantes não serão disseminadas. O pesquisador deve também não informar detalhes do experimento aos participantes (COT-C10). Por fim, os participantes selecionados não podem ter conhecimento sobre as hipóteses do experimento (COT-C12).

Por último, a Figura 4.28 apresenta também a ameaça “*Participante pode preencher o questionário de caracterização incorretamente*” (COT-A09). Esta ameaça ocorre quando o participante consciente ou inconscientemente preenche o questionário de caracterização de forma inconsistente com a realidade. Para esta ameaça identificamos apenas uma ação de controle: “*Utilizar dados de desempenho do participante em um curso*” (COT-C19). Com base em dados históricos, o pesquisador tem uma métrica melhor sobre o que se pretende medir. A seguir, serão apresentadas as ameaças à validade de conclusão e as ações de controle identificadas na SLR.

4.1.4. Ameaças à Validade de Conclusão e Ações de Controle

A seguir, apresentaremos as ameaças à validade de conclusão que foram identificadas

através da SLR. Identificamos um total de 8 ameaças à validade de conclusão e 18 ações de controle.

A Figura 4.29 apresenta a ameaça “*Confiabilidade das medidas*” (COC-A01). Esta ameaça ocorre quando as medidas utilizadas no experimento não são confiáveis, ou ainda quando a forma de análise dos dados é duvidosa. Para esta ameaça identificamos cinco ações de controle:

- COC-C01 – Utilizar medidas apresentadas em estudos anteriores;
- COC-C03 – Definir um método de medição dos dados;
- COC-C04 – Definir pessoas diferentes para avaliar os dados coletados;
- COC-C06 – Executar estudo para avaliar o método de medição; e
- COC-C07 – Realizar reuniões com o objetivo de resolver divergências.

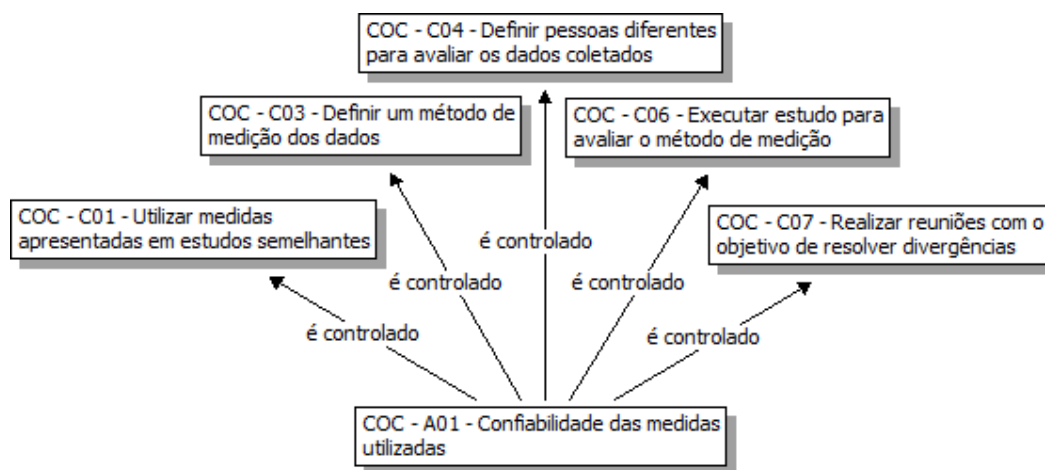


Figura 4.29. Ameaça COC-A01 e suas ações de controle.

A ação de controle COC-C01 sugere que o pesquisador selecione medidas que foram apresentadas em estudos semelhantes. Medidas utilizadas em estudos anteriores possuem confiabilidade porque passam por validação. O pesquisador pode também definir um método para medição dos dados. Além disso, pode ser realizado estudo para avaliar e validar o método definido (COC-C06). Planejar a forma como os dados devem ser medidos evita avaliação subjetiva. Utilizar mais de uma pessoa para analisar os dados coletados também minimiza a subjetividade (COC-C04). Em caso de divergências na análise dos dados, deverão ser realizadas reuniões para que as discordâncias sejam sanadas (COC-C07).

As ameaças “*Violação dos requisitos do teste estatístico*” (COC-A02) e “*Seleção de amostra heterogênea*” (COC-A03) são apresentadas na Figura 4.30. A primeira ocorre quando o pesquisador aplica um teste estatístico inadequado. Isso acontece porque existem diversos fatores que influenciam na seleção do teste estatístico, por exemplo, tamanho da amostra, número de grupos, dentre outros. A segunda ocorre quando a amostra selecionada é heterogênea. Desse modo,

podem existir diferenças, por exemplo, na experiência dos participantes que podem influenciar o resultado do experimento.

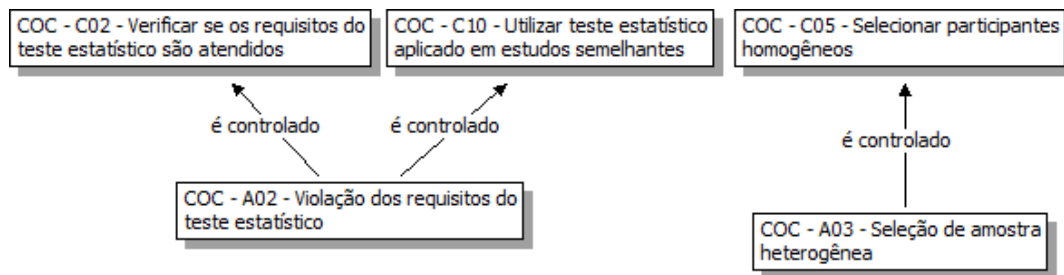


Figura 4.30. Ameaças COC-A02 e COC-A03 e suas ações de controle.

Para a ameaça COC-A02 identificamos duas ações de controle:

- COC-C02 – Verificar se os requisitos do teste estatístico são atendidos; e
- COC-C10 – Utilizar teste estatístico aplicado em estudos semelhantes.

A ação COC-C02 sugere que o pesquisador verifique todos os requisitos para uso dos testes estatísticos. Dessa forma, a escolha de um teste inadequado é evitada. O pesquisador pode também utilizar o teste estatístico aplicado em estudos semelhantes que já foram publicados (COC-C10).

Por outro lado, para a ameaça COC-A03 identificamos apenas uma ação de controle: “*Selecionar participantes homogêneos*” (COC-C05). Assim, as diferenças que possam influenciar no resultado do estudo são menores.

A ameaça “*Seleção de dados para satisfazer a hipótese*” (COC-A04) ocorre quando o pesquisador escolhe os melhores resultados dentre os coletados para que as hipóteses do experimento sejam satisfeitas. Não identificamos ações de controle para esta ameaça.

A Figura 4.31 apresenta três ameaças à validade de conclusão:

- COC-A05 – Amostra de tamanho pequeno;
- COC-A06 – Fatos aleatórios que podem causar problemas no ambiente experimental; e
- COC-A08 – Representatividade do intervalo de confiança.

A ameaça COC-A05 ocorre quando a amostra é considerada pequena. Dessa forma, o poder estatístico é baixo e as conclusões são questionadas. A ameaça COC-A06 ocorre quando problemas fora do comum ocorrem no ambiente experimental e podem causar alteração no desempenho do participante. Por fim, o intervalo de confiança selecionado (COC-A08) para execução do teste estatístico pode não ser adequado. Caso isso ocorra, o poder estatístico também é baixo.

Para cada ameaça identificamos apenas uma ação de controle:

- COC-C09 – Elaborar um design que aumente o número de observações (para COC-A05);

- COC-C08 – Atribuir os participantes ao mesmo ambiente experimental (para COC-A06); e
- COC-C17 – Utilizar intervalo de confiança adequado de acordo com a literatura (para COC-A08).

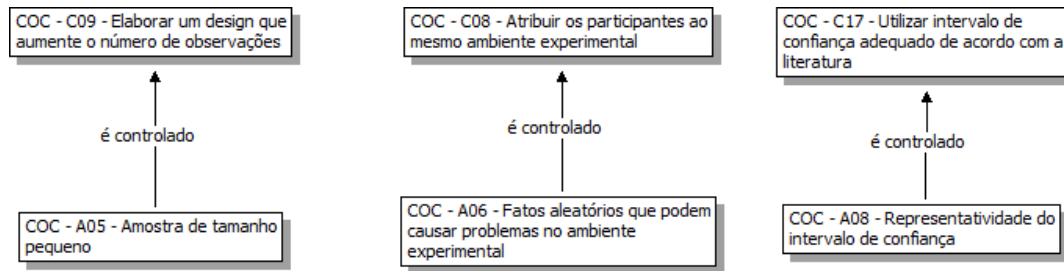


Figura 4.31. Ameaças COC-A05, COC-A06 e COC-A08 e suas ações de controle.

Para tentar minimizar a ameaça COC-A05 o pesquisador pode elaborar um design que aumente o número de datapoints (COC-A09). A ação de controle COC-C08 sugere que os participantes sejam alocados ao mesmo ambiente experimental. Dessa forma, os fatos (COC-A06) que ocorrerem serão equivalentes para todos os participantes. Por fim, a ação COC-C17 informa que deve ser utilizado intervalo de confiança descrito na literatura. Dessa forma, a ameaça COC-A08 pode ser minimizada.

Finalmente, a Figura 4.32 apresenta a ameaça “*Aplicação do tratamento de forma incorreta*” (COC-A07). Esta ameaça ocorre quando o participante não aplica o tratamento da forma que deveria. Isso acontece por falta de conhecimento, ou por não seguir os procedimentos informados pelos pesquisadores, ou, até mesmo, de forma consciente. Identificamos sete ações de controle para esta ameaça:

- COC-C11 – Definir diretrizes em documento para aplicação correta do tratamento;
- COC-C12 – Lembrar os participantes que devem utilizar o tratamento corretamente;
- COC-C13 – Introduzir observadores para verificar a corretude da aplicação do tratamento;
- COC-C14 – Utilizar participantes sem experiência em um tópico;
- COC-C15 – Utilizar câmeras para verificar a aplicação do tratamento;
- COC-C16 – Realizar treinamento para que o tratamento seja aplicado corretamente; e
- COC-C18 – Utilizar ferramenta que automatiza o tratamento.

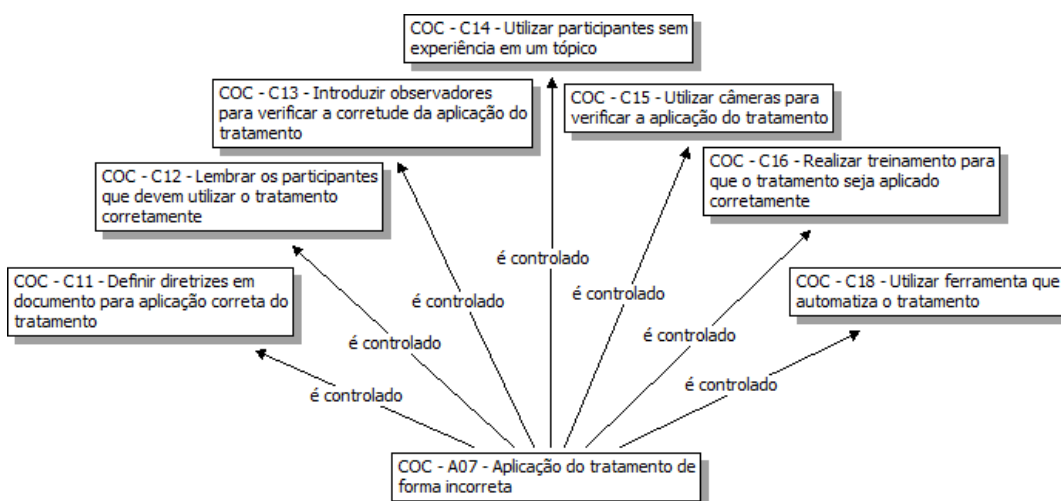


Figura 4.32. Ameaça COC-A07 e suas ações de controle.

A ação COC-C11 sugere que seja elaborado um documento que forneça instruções para o participante aplicar o tratamento corretamente. É uma espécie de guia para que não ocorram inconformidades. O pesquisador pode também lembrar constantemente os participantes que o tratamento deve ser aplicado da forma solicitada (COC-C12). Outra forma de minimizar a ameaça COC-A07 é introduzir observadores para verificar a aplicação do tratamento (COC-C13). Caso ocorram erros, os observadores podem intervir. O pesquisador pode evitar a ameaça se selecionar participantes sem experiência no tópico que está sendo estudado (COC-C14). Pessoas que possuem experiência tem forte tendência em realizar as tarefas da forma que estão acostumados. Com isso, o tratamento pode não ser aplicado corretamente. Outra maneira é utilizar câmeras para verificar se o participante aplica o tratamento de forma correta (COC-C15). O pesquisador pode ainda treinar o participante para que não haja problemas na aplicação do tratamento (COC-C16). Por fim, se existir, pode ser utilizada ferramenta que automatize o uso do tratamento (COC-C18). Dessa forma, o risco de erro ao aplicar o tratamento é mitigado.

5. Assistente de Identificação de Causas e Tratamentos de Ameaças à Validade

Para auxiliar o processo de identificação de causas e tratamentos de ameaças à validade, será desenvolvido um assistente que apresentará uma lista de causas e seus respectivos tratamentos para serem selecionados pelo usuário. O assistente desenvolvido deve atender aos seguintes requisitos:

- Auxiliar o pesquisador na identificação de causas e tratamentos de ameaças à validade, através de uma lista construída com base nas evidências da literatura;

- Facilitar o controle do pesquisador em estudos experimentais, a cerca de identificação e tratamento de ameaças à validade. Isso acontece porque existem relacionamentos entre causas e tratamentos, de forma que ao tratar uma causa pode-se maximizar outra.

As causas serão divididas em “antes do experimento” (pré-experimento) e “durante o experimento” (em experimento), para um melhor entendimento por parte do usuário e, também, para evitar poluição visual. A Figura 5.1 apresenta as duas primeiras causas da aba **em experimento**. Todas as causas serão listadas com seus respectivos tratamentos e o usuário deverá selecionar o que lhe for conveniente.

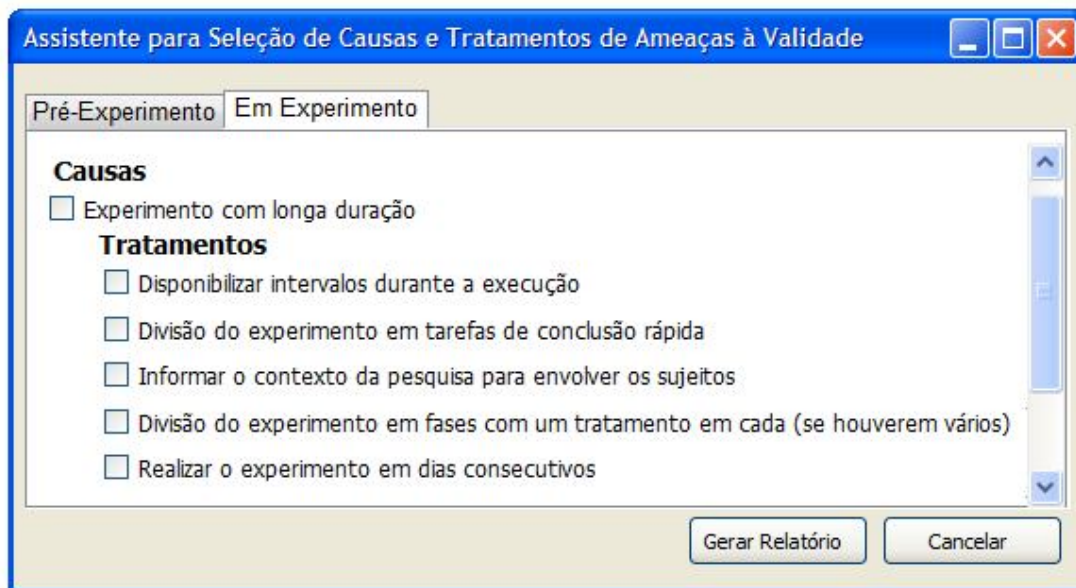


Figura 5.1. Tela apresentando parte da lista de causas de ameaças em experimento e tratamentos.

À medida que as causas e tratamentos forem selecionados, serão exibidos alertas, indicando que determinado tratamento selecionado pode causar outra ameaça ou que uma causa pode potencializar a existência de uma nova ameaça. Conforme a Figura 5.2, a mensagem de alerta surge no momento em que o pesquisador seleciona o tratamento T29. Caso prossiga, a causa C21 será selecionada e um dos seus tratamentos deve ser escolhido.

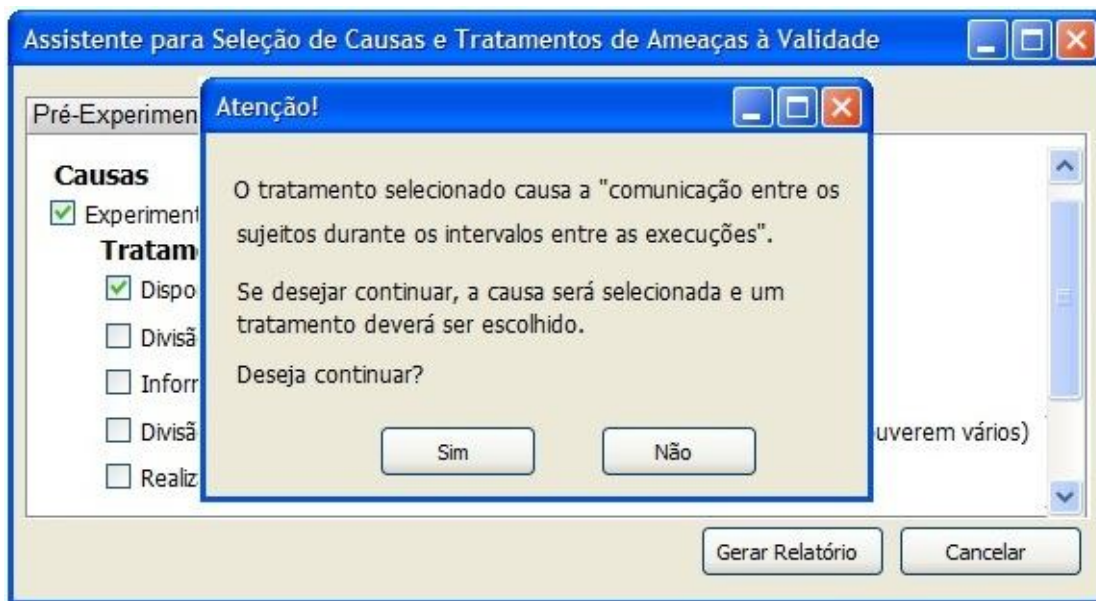


Figura 5.2. Tela apresentando a mensagem de alerta ao pesquisador.

Ao final do uso do assistente, será apresentado um relatório contendo as causas presentes no experimento e os tratamentos que deverão ser aplicados, conforme ilustrado na Figura 5.3.

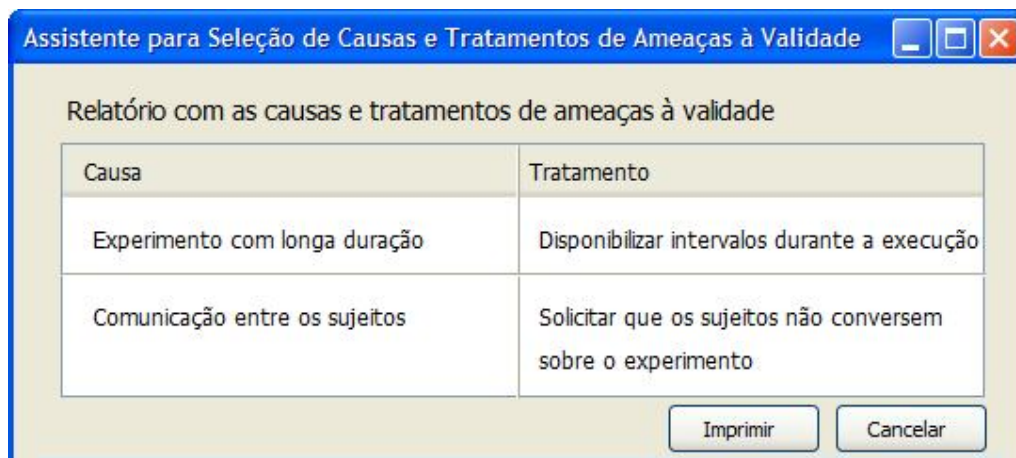


Figura 5.3. Tela apresentando o relatório final.

O diagrama de atividades do assistente é apresentado na Figura 5.4:

- Selecionar tipo de validade: nesta atividade o usuário deve escolher qual tipo de validade pretende tratar (interna, externa, constructo ou conclusão);
- Selecionar causas de ameaças: após ter escolhido o tipo da validade que pretende tratar o usuário deve identificar as causas que estão presentes no estudo experimental;

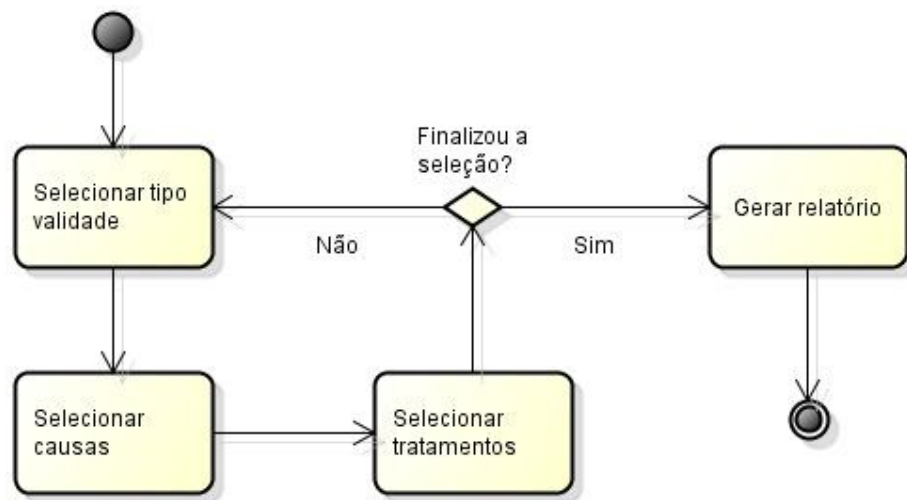


Figura 5.4. Diagrama de atividades do assistente.

- Selecionar tratamentos: o usuário poderá escolher o tratamento desejado para mitigação da causa da ameaça à validade;
- Gerar relatório: caso o usuário tenha finalizado a seleção das causas e dos tratamentos de ameaças à validade, o relatório final é gerado. Se o usuário optar por selecionar outro tipo de validade, retorna para a primeira atividade.

Referências Bibliográficas

- Babbie, E., Survey research methods, Wadsworth, ISBN 0-524-12672-3 (1990).
- Barros, M. O. e Dias Neto, A. C., Threats to validity in search-based software engineering empirical studies, Relatório Técnico RT-006/2011, (2011) 1-12.
- Biolchini, J. C. D. A., Mian, P. G., Natali, A. C. C., Conte, T. e Travassos, G. H., Scientific research ontology to support systematic review in software engineering. *Advanced Engineering Informatics*, v. 21, n. 2, (2007) pp. 133-151.
- Conte, T. U., Técnica de inspeção de usabilidade baseada em perspectivas de projeto web. Tese de doutorado, COPPE/UFRJ, (2009).
- Creswell, J. W., Research design: qualitative and quantitative approaches, Sage Publications (1994).
- Creswell, J. W., Research design: qualitative, quantitative and mixed methods approaches. Sage Publications (2002).
- Denzin, N. K. e Lincoln, Y. S., Handbook of qualitative research, Sage Publications, Londres, 1994.
- Feldt, R. e Magazinius, A., Validity threats in empirical software engineering research – an initial survey, *Proceedings of the 22^o International Conference on Software Engineering and Knowledge Engineering (SEKE)* (2010) 374-379.
- Kampenes, V. B., Dyba, T., Hannay, J. E. e Sjøberg, D. I. K., A systematic review of quasi-experiments in software engineering, *Information and Software Technology* 51 (2009) 71-82.
- Kitchenham, B. e Charters, S., Guidelines for performing systematic literature reviews in software engineering, *EBSE Technical Report EBSE-2007-01*, v. 2.3 (2007).
- Mafra, S., Barcelos, R. e Travassos, G. H., Aplicando uma metodologia baseada em evidência na definição de novas tecnologias de software. *Anais do XX Simpósio Brasileiro de Engenharia de Software (SBES 2006)*, (2006) pp. 239 – 254.
- McGrath, J. E., Methodology matters: doing research in the behavioral and social sciences, em In *Human-Computer Interaction: Toward the Year 2000*. Morgan Kaufmann Publishers (1995) 152-169.
- Shadish, W. R., Cook, T. D. e Campbell, D. T., Experimental and quasi-experimental designs for generalized causal inference, Houghton Mifflin, (2002).

- Shull, F., Carver, J. e Travassos, G. H., An empirical methodology for introducing software processes. *ACM SIGSOFT Software Engineering Notes*, v. 26, n. 5, (2001) 288-296.
- Sjøberg, D. I. K., Hannay, J. E., Hansen, O., Kampenes, V. B., Karahasanovic, A., Liborg, N. K. e Rekdal, A. C., A survey of controlled experiments in software engineering, *IEEE Transactions on Software Engineering* 31 (9) (2005) 733-753.
- Travassos, G. H., Gurov, D. e Amaral, E. A. G., Introdução à Engenharia de Software Experimental, Relatório Técnico RT-ES-590/02, (2002) 1-52.
- Travassos, G. H. e Barros, M. O., Contributions of in virtuo and in silico experiments for the future of empirical studies in software engineering, *Proc. of 2003 ACMIEEE International Symposium on Empirical Software Engineering (ISESE 2003)*, Italia, (2003).
- Travassos, G. H., Santos, P. S. M., Mian, P. G., Dias Neto, A. C. e Biolchini, J., A Environment to Support Large Scale Experimentation in Software Engineering, *13^o IEEE International Conference on Engineering of Complex Computer Systems (ICECCS)*, (2008) 193-202.
- Wazlawick, R. S., Metodologia de pesquisa para Ciência da Computação. Elsevier (2008).
- Wöhlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B. e Wessl, A., Experimentation in software engineering: an introduction. Kluwer Academic Publishers (2000).

Referências da Revisão da Literatura

- [EP1] Arisholm, E., Sjøberg, D. I. K. e Jørgensen, M., Assessing the changeability of two objectoriented design alternatives? A controlled experiment, *Empirical Software Engineering* 6 (3) (2001) 231-237.
- [EP2] Basili, V. R., Green, S., Laitenberger, O., Lanubile, F., Shull, F., Sorumgard, S. e Zelkowitz, M. V., The empirical investigation of perspective-based reading, *Empirical Software Engineering* 1 (2) (1996) 133-164.
- [EP3] Bianchi, A., Lanubile, F. e Visaggio, G., A controlled experiment to assess the effectiveness of inspection meetings, *Proceedings of the Seventh International Software Metrics Symposium (METRICS'01)*, IEEE Computer Society (2001) 42-50.
- [EP4] Biffl, S. e Halling, M., Investigating the influence of inspector capability factors with four inspection techniques on inspection performance, *Proceedings of the 8^o International Software Metrics Symposium (METRICS'02)*, IEEE Computer Society (2002) 107-117.
- [EP5] Briand, L. C., Bunse, C. e Daly, J. W., A controlled experiment for evaluating quality guidelines on the maintainability of object-oriented designs, *IEEE Transactions on Software Engineering* 27 (6) (2001) 513-530.
- [EP6] Briand, L. C., Bunse, C., Daly, J. W. e Differding, C., Technical communication: an experimental comparison of the maintainability of object-oriented and structured design documents, *Empirical Software Engineering* 2 (3) (1997) 291-312.
- [EP7] Calero, C., Piattini, M. e Genero, M., Empirical validation of referential integrity metrics, *Information and Software Technology* 43 (15) (2001) 949-957.
- [EP8] Dunsmore, A., Roper, M. e Wood, M., Systematic object-oriented inspection an empirical study, *ICSE. Proceedings of the 23^o International Conference on Software Engineering*, (2001) 135-144.
- [EP9] Dunsmore, A., Roper, M. e Wood, M., Further investigations into the development and evaluation of reading techniques for object-oriented code inspection, *ICSE. Proceedings of the 24^o International Conference on Software Engineering*, (2002) 47-57.
- [EP10] Freimut, B., Laitenberger, O. e Biffl, S., Investigating the impact of reading techniques on the accuracy of different defect content estimation techniques, *Proceedings of the 7^o International Software Metrics Symposium (METRICS'01)*, IEEE Computer Society (2001) 51-62.

- [EP11] Fusaro, P., Lanubile, F. e Visaggio, G., A replicated experiment to assess Requirements inspection techniques, *Empirical Software Engineering* 2 (1) (1997) 39-57.
- [EP12] Harrison, R., Counsell, S. e Nithi, R., Experimental assessment of the effect of inheritance on the maintainability of object-oriented systems, *The Journal of Systems and Software* 52 (2-3) (2000) 173-179.
- [EP13] Höst, M. e Wohlin, C., An experimental study of individual subjective effort estimation and combinations of the estimates, *Proceedings of the 20th International Conference on Software Engineering (ICSE)*, IEEE Computer Society (1998) 332-339.
- [EP14] Höst, M. e Johansson, C., Evaluation of code review methods through interviews and experimentation, *The Journal of Systems and Software* 52 (2-3) (2000) 113-120.
- [EP15] Höst, M., Regnell, B. e Wohlin, C., Using students as subjects - a comparative study of students and professionals in lead-time impact assessment, *Empirical Software Engineering* 5 (3) (2000) 201-214.
- [EP16] Johnson, P. M. e Tjahjono, D., Does Every Inspection Really Need a Meeting?, *Empirical Software Engineering* 3 (1) (1998) 9-35.
- [EP17] Kiper, J. D., Auernheimer, B. e Ames, C. K., Visual depiction of decision statements: what is best for programmers and non-programmers?, *Empirical Software Engineering* 2 (4) (1997) 361-379.
- [EP18] Laitenberger, O. e DeBaud, J. M., Perspective-based reading of code documents at Robert Bosch GmbH, *Information and Software Technology* 39 (11) (1997) 781-791.
- [EP19] Laitenberger, O., El Emam, K. e Harbich, T. G., An internally replicated quasi-experimental comparison of checklist and perspective based reading of code documents, *IEEE Transactions on Software Engineering* 27 (5) (2001) 387-421.
- [EP20] Laitenberger, O., Atkinson, C., Schlich, M. e El Emam, K., An experimental comparison of reading techniques for defect detection in UML design documents, *The Journal of Systems and Software* 53 (2) (2000) 183-204.
- [EP21] Land, L. P. W., Sauer, C. e Jeffery, R., The use of procedural roles in code inspections: an experimental study, *Empirical Software Engineering* 5 (1) (2000) 11-34.
- [EP22] Lott, C. M., Technical communication: a controlled experiment to evaluate on-line process guidance, *Empirical Software Engineering* 2 (3) (1997) 269-289.

- [EP23] MacDonald, F. e Miller, J., A comparison of tool-based and paper-based software inspection, *Empirical Software Engineering* 3 (3) (1998) 233-253.
- [EP24] McCarthy, P., Porter, A. A., Siy, H. e Votta, L. G., An experiment to assess costbenefits of inspection meetings and their alternatives: a pilot study, *3^o IEEE International Software Metrics Symposium (METRICS)*, (1996) 100-111.
- [EP25] Miller, J., Wood, M. e Roper, M., Further experiences with scenarios and checklists, *Empirical Software Engineering* 3 (1) (1998) 37-64.
- [EP26] Ohlsson, M. C., Wohlin, C. e Regnell, B., A project effort estimation study, *Information and Software Technology* 40 (11-12) (1998) 831-839.
- [EP27] Otero, M. C. e Dolado, J. J., An initial experimental assessment of the dynamic modeling in UML, *Empirical Software Engineering* 7 (1) (2002) 27-47.
- [EP28] Pfahl, D., Koval, N. e Ruhe, G., An experiment for evaluating the effectiveness of using a system dynamics simulation model in software project management education, *7^o IEEE International Software Metrics Symposium (METRICS)*, IEEE Computer Society (2001) 97-109.
- [EP29] Porter, A. A. e Votta, L. G., An experiment to assess different defect detection methods for software requirements inspections, *Proceedings of the 16^o International Conference on Software Engineering (ICSE)*, (1994) 103-112.
- [EP30] Porter, A. A. e Johnson, P. M., Assessing software review meetings: results of a comparative analysis of two experimental studies, *IEEE Transactions on Software Engineering* 23 (3) (1997) 129-145.
- [EP31] Porter, A. A. e Votta, L. G., Comparing detection methods for software requirements inspections: a replication using professional subjects, *Empirical Software Engineering* 3 (4) (1998) 355-379.
- [EP32] Porter, A. A., Votta, L. G. e Basili, V. R., Comparing detection methods for software requirements inspections: a replicated experiment, *IEEE Transactions on Software Engineering* 21 (6) (1995) 563-575.
- [EP33] Prechelt, L. e Tichy, W. F., A controlled experiment to assess the benefits of procedure argument type checking, *IEEE Transactions on Software Engineering* 24 (4) (1998) 302-312.
- [EP34] Prechelt, L. e Unger, B., An experiment measuring the effects of personal software process (PSP) training, *IEEE Transactions on Software Engineering* 27 (5) (2000) 465-472.

- [EP35] Prechelt, L., Unger-Lamprecht, B., Philippsen, M. e Tichy, W. F., Two controlled experiments assessing the usefulness of design pattern documentation in program maintenance, *IEEE Transactions on Software Engineering* 28 (6) (2002) 595-606.
- [EP36] Prechelt, L., Unger, B., Tichy, W. F., Brossler, P. e Votta, L. G., A controlled experiment in maintenance: comparing design patterns to simpler solutions, *IEEE Transactions on Software Engineering* 27 (12) (2001) 1134-1144.
- [EP37] Ramanujan, S., Scamell, R. W. e Shah, J. R., An experimental investigation of the impact of individual, program, and organizational characteristics on software maintenance effort, *The Journal of Systems and Software* 54 (2) (2000) 137-157.
- [EP38] Regnell, B., Runeson, P. e Thelin, T., Are the perspectives really different? Further experimentation on Scenario-Based reading of requirements, *Empirical Software Engineering* 5 (4) (2000) 331-356.
- [EP39] Roper, M., Wood, M. e Miller, J., An empirical evaluation of defect detection technique, *Information and Software Technology* 39 (11) (1997) 763-775.
- [EP40] Rothermel, K. J., Cook, C. R., Burnett, M. M., Sconfeld, J., Green, T. R. G. e Rothermel, G., WYSIWYT testing in the spreadsheet paradigm: an empirical evaluation, *Proceedings of the 22^o International Conference on Software Engineering (ICSE)* (2000) 230-239.
- [EP41] Sabaliauskaite, G., Matsukawa, F., Kusumoto, S. e Inoue, K., Experimental comparison of checklist-based reading and perspective-based reading for UML design document inspection reading, *International Symposium on Empirical Software Engineering (ISESE)*, IEEE Computer Society (2002) 148-160.
- [EP42] Sandahl, K., Blomkvist, O., Karlsson, J., Krysanter, C., Lindvall, M. e Ohlsson, N., An extended replication of an experiment for assessing methods for software requirements inspections, *Empirical Software Engineering* 3 (4) (1998) 327-354.
- [EP43] Thelin, T., Runeson, P. e Regnell, B., Usage-based reading - an experiment to guide reviewers with use cases, *Information and Software Technology* 43 (15) (2001) 925-938.
- [EP44] Thelin, T., Runeson, P., Wohlin, C., Olsson, T. e Anderson, C., How much information is needed for usage-based reading, *International Symposium on Empirical Software Engineering (ISESE)*, IEEE Computer Society (2002) 127-138.
- [EP45] Zendler, A., Pfeiffer, T., Eicks, M. e Lehner, F., Experimental comparison of coarse-grained concepts in UML, OML, and TOS, *The Journal of Systems and Software* 57 (1) (2001) 21-30.

[EP46] Zhang, Z., Basili, V. R. e Shneiderman, B., Perspective-based usability inspection: an empirical validation of efficacy, *Empirical Software Engineering* 4 (1) (1999) 43-69.