

Revista Eletrônica de Sistemas de Informação

ISSN 1677-3071

Sumário

Editorial

[Editorial](#)

Alexandre Reis Graeml

Ensino e pesquisa

[COMPUTAÇÃO UBÍQUA: ESTADO DA ARTE E OPORTUNIDADES DE PESQUISA PARA A ÁREA DE NEGÓCIOS](#)

Fabio Miguel Junges, Amarolinda Zanela Klein, Jorge L. V. Barbosa

Foco nas pessoas

[SITES DE REDES SOCIAIS CORPORATIVAS: ENTRE O PESSOAL E O PROFISSIONAL](#)

Jorge da Silva Correia-Neto, Jairo Simião Dornelas, Guilherme Vilar

Foco nas organizações

[AVALIAÇÃO DA CAPACIDADE DOS PROCESSOS DE GOVERNANÇA CORPORATIVA DE TI BASEADA NO COBIT 5](#)

Diana Leite Nunes dos Santos, João Souza Neto

Foco na tecnologia

[VISUALIZAÇÃO DE INFORMAÇÃO DE OPINIÕES ONLINE SOBRE RESTAURANTES: USO DE TÉCNICAS ORIENTADAS À VISUALIZAÇÃO DE GRAFOS](#)

Elizabeth Simão Carvalho, Marcírio Silveira Chaves, Nélia Sacramento

[INVESTIGAÇÃO EXPERIMENTAL E PRÁTICAS ÁGEIS: AMEAÇAS À VALIDADE DE EXPERIMENTOS ENVOLVENDO A PRÁTICA ÁGIL PROGRAMAÇÃO EM PAR](#)

Vagner Carlos Marcolino Lima, Adolfo Gustavo Serra Seca Neto, Maria Claudia Figueiredo Pereira Emer



Este trabalho está licenciado sob uma [Licença Creative Commons Attribution 3.0](#).

Esta revista é (e sempre foi) eletrônica para ajudar a proteger o meio ambiente, mas, caso deseje imprimir esse artigo, saiba que ele foi editorado com uma fonte mais ecológica, a *Eco Sans*, que gasta menos tinta.

This journal is (and has always been) electronic in order to be more environmentally friendly. Now, it is desktop edited in a single column to be easier to read on the screen. However, if you wish to print this paper, be aware that it uses Eco Sans, a printing font that reduces the amount of required ink.

INVESTIGAÇÃO EXPERIMENTAL E PRÁTICAS ÁGEIS: AMEAÇAS À VALIDADE DE EXPERIMENTOS ENVOLVENDO A PRÁTICA ÁGIL PROGRAMAÇÃO EM PAR

EXPERIMENTAL INVESTIGATION AND AGILE PRACTICES: THREATS TO THE VALIDITY OF EXPERIMENTS INVOLVING THE PAIR PROGRAMMING AGILE PRACTICE

(artigo submetido em janeiro de 2013)

Vagner Carlos Marcolino Lima

Mestre pelo Programa de Pós-Graduação em
Computação Aplicada (PPGCA) da Universi-
dade Tecnológica Federal do Paraná (UTFPR)
vagnercml@hotmail.com

Adolfo Gustavo Serra Seca Neto

Professor do Programa de Pós-Graduação em
Computação Aplicada (PPGCA) da Universi-
dade Tecnológica Federal do Paraná (UTFPR)
adolfo@dainf.ct.utfpr.edu.br

Maria Claudia Figueiredo Pereira Emer

Professor do Programa de Pós-Graduação em Computação Aplicada (PPGCA)
da Universidade Tecnológica Federal do Paraná (UTFPR)
mclaudia@dainf.ct.utfpr.edu.br

ABSTRACT

Even though the process of investigation by experimentation, in the context of software engineering, is not something new, it still lacks new quality studies. In studies that adopt the experiment as a method of research, a complete and detailed evaluation of the threats to the internal, external, construct and conclusion validity is an extremely important quality factor. Based on a detailed analysis of two empirical studies, this work presents a list of 21 possible threats to the validity of an experiment involving the pair programming agile practice, giving an exemplification of 17 possible threats.

Key-words: experimental research; pair programming; threats to experiments; agile practice; experimental software engineering.

RESUMO

A investigação por experimentação no contexto da engenharia de software não é algo recente, mas ainda carece de novos estudos de qualidade. Em estudos que adotam o experimento como método de pesquisa, uma avaliação completa e detalhada das ameaças à validade interna, externa, de construção e de conclusão é um fator de qualidade importantíssimo. A partir de uma análise detalhada de dois estudos empíricos, este trabalho apresenta uma lista de 21 possíveis ameaças à validade de um experimento envolvendo a prática ágil programação em par, proporcionando a exemplificação de 17 possíveis ameaças.

Palavras-chave: pesquisa experimental; programação em par; ameaças à experimentos; prática ágil; engenharia de software experimental.

1 INTRODUÇÃO

O interesse da indústria de software em métodos ágeis para o desenvolvimento de sistemas, na última década, vem crescendo ano após ano (DYBÅ e DINGSØYR, 2008). Os estudos sugerem que a academia deve aumentar o número e a qualidade de estudos envolvendo tais métodos (DYBÅ e DINGSØYR, 2009). Esses métodos, diferentemente das abordagens tradicionais de desenvolvimento de software, valorizam mais as pessoas e interações que processos e ferramentas (BECK *et al.*, 2001).

Buscando contribuir com este cenário, realizamos este estudo que busca investigar por meio de experimentos o desempenho da prática ágil programação em par perante diferentes tarefas de modelagem e construção de software. O desempenho será medido a partir da satisfação, motivação e produtividade dos participantes e também da qualidade do software por eles produzido. E, na época, buscávamos a eventual existência de uma forma diferenciada de avaliação de desempenho das duplas *versus* esforço individual.

A programação em par é uma prática de engenharia de software, conforme Beck e Andres (2004), primária do método ágil *extreme programming*, ou XP. De acordo com VersionOne (2011) e Melo *et al.* (2012), a XP é um dos métodos ágeis mais adotados pela indústria de software.

A investigação por experimentação no contexto da engenharia de software não é algo recente, mas ainda carece de novos estudos de qualidade. O experimento, o estudo de caso e o survey são as três principais formas primárias de se conduzir uma investigação experimental. No contexto deste trabalho, destaca-se o primeiro. O experimento caracteriza-se pela manipulação de algumas variáveis e a observação de outras em um ambiente controlado. Em geral, essa forma traz um maior controle sobre a execução, medição, investigação e facilidade de repetição. Porém, o custo e o risco são altos (MAFRA e TRAVASSOS, 2006; TRAVASSOS, 2011; TRAVASSOS, GUROV e AMARAL, 2002).

As etapas de uma investigação experimental, independentemente da forma adotada, são definição, planejamento, avaliação, execução, análise e interpretação e, por fim, realizada em paralelo, empacotamento. Entretanto, definição e planejamento são referências para as etapas seguintes. É na etapa de planejamento de um experimento em que o pesquisador se preocupa em como o estudo experimental será conduzido. Por exemplo, qual o contexto e o desenho do experimento (MAFRA e TRAVASSOS, 2006).

É também especificamente na etapa de planejamento da investigação experimental que o pesquisador deve analisar e definir as ameaças à validade do seu experimento segundo quatro tipos de validade: interna, externa, de construção e de conclusão. A eventual existência de uma determinada ameaça que não foi previamente definida e devidamente tratada pode comprometer os resultados do experimento e, com isso, invalidar o estudo como um todo (MAFRA e TRAVASSOS, 2006).

Há obras que descrevem as ameaças mais comuns para os tipos de validade citados anteriormente (TRAVASSOS *et al.*, 2002; WAINER, 2007). É por meio dessas descrições que o pesquisador definirá as ameaças específicas à validade de seu experimento. Contudo, logo que chegamos à etapa de planejamento do estudo experimental descrito no início desta seção, principalmente no passo de análise e definição das ameaças, levantaram-se os seguintes questionamentos: as ameaças à validade de um experimento envolvendo a prática ágil programação em par limitam-se às ameaças apontadas como as mais comuns? E, se as descrições das ameaças mais comuns são genéricas e em geral sem exemplos práticos, como avaliar essa possível limitação?

Perante esse contexto, este artigo apresenta o resultado de uma identificação preliminar de exemplos práticos de ameaças à validade de um experimento envolvendo a prática ágil programação em par. Espera-se, com o resultado deste trabalho, auxiliar outros pesquisadores orientando seus esforços para avaliações mais abrangentes e detalhadas das ameaças à validade de seus experimentos. E, também, busca-se contribuir com a qualidade do estudo experimental que está em andamento.

Nas seções a seguir, apresentamos um resumo sobre programação em par, conceitos pertinentes à validade de um experimento, a análise e identificação de exemplos práticos de ameaças à validade de experimentos envolvendo programação em par (PP), uma discussão dessas análises e, finalmente, as conclusões.

2 PROGRAMAÇÃO EM PAR

A programação em par é uma prática em que dois desenvolvedores trabalham lado a lado em um computador colaborando continuamente em um mesmo projeto, algoritmo, código ou teste (WILLIAMS e KESSLER, 2000). Nela, os dois desenvolvedores podem ser vistos como um único organismo pensante, o qual é responsável por todos os aspectos do artefato que está sendo elaborado. Um desenvolvedor, o piloto, controla o computador e escreve o código. O outro, chamado de navegador, continua e ativamente observa o trabalho do piloto, procurando por defeitos, pensando em alternativas, buscando recursos e considerando implicações estratégicas. E, durante o processo, eles deliberadamente e periodicamente trocam de papéis (WILLIAMS *et al.*, 2000).

Alguns dos benefícios esperados com o uso de PP no desenvolvimento de software são: (a) melhoria da qualidade do projeto, o trabalho em equipe, a comunicação e a disseminação do conhecimento e, combinada com outras práticas, tais como desenvolvimento orientado a teste (test driven development - TDD), é uma prática chave para melhoria da qualidade do código (SFETSOS e STAMELOS, 2010); (b) aumento da produtividade, em geral, no médio e longo prazo, segundo Poppendieck e Poppendieck (2011), pois a qualidade e a robustez do código são melhoradas quando “dois conjuntos de olhos” continuamente inspecionam o código.

Embora os benefícios citados anteriormente sirvam de grande incentivo para o uso da PP, tanto na indústria de software quanto em outros contextos, há alguns desafios a serem enfrentados. Alguns destes podem inviabilizar a adoção da prática na indústria de software, são eles: o relacionamento humano, que é um dos grandes desafios da PP (TELES, 2004; BECK e ANDRES, 2004) ou a “guerra” de ferramentas de desenvolvimento e padrões de codificação e pares do tipo “professor-estudante” (SHORE e WARDEN, 2008).

3 VALIDADE DE UM EXPERIMENTO

Em um experimento o pesquisador possui controle sobre a quantidade e o tipo de grupos de participantes, qual grupo fará o quê, em que momento haverá a intervenção e, em alguns casos, quem participará de cada grupo (WAINER, 2007). Os tipos de variáveis possíveis são: (a) variáveis independentes (ou fatores), as quais apresentam a causa que afeta o processo de experimentação e (b) variáveis dependentes, as quais se referem à saída de tal processo. O valor de (a) se chama ‘tratamento’ e o valor de (b) se chama ‘resultado’ (MAFRA e TRAVASSOS, 2006; TRAVASSOS *et al.*, 2002).

E mais, há vários desenhos experimentais – em linhas gerais, como os grupos de participantes serão arranjados (ver seção 3.5). Ressalta-se que a presença ou não de determinadas ameaças à validade do experimento dependerá do desenho experimental escolhido pelo pesquisador (WAINER, 2007).

A validade de um experimento está relacionada ao nível de confiança que se pode ter no processo de investigação experimental como um todo. Ou seja, o quão confiáveis são os elementos envolvidos nesse processo – desde a base teórica adotada até os resultados obtidos, inclusive a forma como são apresentados, considerando validade de construção, interna e externa (TRAVASSOS *et al.*, 2002; WAINER, 2007).

3.1 VALIDADE DE CONSTRUÇÃO

A validade de construção, para Travassos *et al.* (2002, p. 13), “considera os relacionamentos entre a teoria e a observação, ou seja, se o tratamento reflete a causa bem e o resultado reflete o efeito bem”. As ameaças mais comuns a este tipo de validade estão relacionadas com: projeto do experimento: em geral, má definição da base teórica ou da definição do processo de experimentação; fatores humanos (ou sociais): por exemplo, os participantes podem basear seus comportamentos nas hipóteses de pesquisa ou eles podem estar envolvidos em outros experimentos.

3.2 VALIDADE INTERNA

A validade interna, para Travassos *et al.* (2002, p. 13), “define se o relacionamento observado entre o tratamento e o resultado é causal, e não é resultado da influência de outro fator – não controlado ou medido”. Para Wainer (2007), as ameaças mais comuns a este tipo de validade são:

instrumentação: a diferença nos resultados é consequência de uma medição incorreta ou instrumento inadequado; testagem: o desenho proporciona que os participantes aprendam com seus próprios erros; está relacionada à sequência da coleta dos resultados *versus* o momento da intervenção, por exemplo; maturação: os sujeitos podem tornar-se mais capazes ou ficar desmotivados com o passar do tempo; história: possibilidade de que o resultado do processo seja consequência de um evento externo ao experimento.

Continuando a lista de ameaças, Wainer (2007) menciona ainda: seleção: os participantes não foram selecionados de maneira aleatória ou os grupos não foram divididos de maneira igualitária tanto no aspecto quantitativo quanto qualitativo; mortalidade seletiva: a evasão de participantes com características específicas e relevantes durante o experimento; contaminação: por exemplo, participantes que fazem parte de um grupo ensinam, ou aprendem, com participantes do outro grupo; comportamento competitivo: os membros do grupo de controle se sentem preteridos frente aos do outro grupo, e podem se mostrar motivados a competir com o grupo experimental; comportamento compensatório: alguma pessoa, ou entidade, externa ou não ao experimento, sente que o grupo de controle foi preterido e, com isso, cria medidas compensatórias para ele.

E mais, regressão à média: fenômeno que pode trazer ameaça ao experimento. Em geral, conforme Bandeira (2012), essa ameaça está presente quando se trabalha apenas com um grupo (somente o grupo experimental) e seus integrantes são selecionados de maneira deliberada com base nos piores ou nos melhores escores do pré-teste, ou seja, não há um grupo de controle e seleção aleatória de participantes; efeito da expectativa do sujeito: os participantes esperam ou buscam um determinado resultado – positivo ou negativo – para o experimento, ou ainda, pode haver um efeito positivo pelo simples fato de os participantes saberem que estão sendo observados – conhecido também por efeito Hawthorne.

Por fim, Wainer (2007) ainda destaca o efeito da expectativa do experimentador (o pesquisador interage intensamente com os participantes, as crenças dele causam um efeito no sujeito ou nos testes realizados por ele, ou ainda, o simples fato do pesquisador querer que o experimento “dê certo”) e a influência de parte da intervenção (aparece em diferentes domínios com diferentes nomes, mas se refere ao fato de que o efeito observado não é devido à intervenção como um todo, mas apenas a parte dela).

3.3 VALIDADE DE CONCLUSÃO

A validade de conclusão está, segundo Travassos *et al.* (2002, p. 12), “relacionada à habilidade de chegar a uma conclusão correta a respeito dos relacionamentos entre o tratamento e o resultado do experimento”, e isso envolve a correta análise e interpretação estatística do resultado (por

exemplo, teste estatístico *versus* tamanho da amostra), confiabilidade das medidas e confiabilidade da implementação dos tratamentos.

3.4 VALIDADE EXTERNA (VE)

A validade externa, conforme Travassos *et al.* (2002, p. 13), “define as condições que limitam a habilidade de generalizar os resultados de um experimento para a prática industrial”, as ameaças mais comuns a este tipo de validade são: participantes: selecionar participantes que não possuem relação ou refletem o comportamento da população, ou, trabalhar com uma amostra não representativa quantitativamente e qualitativamente da população; tempo: por exemplo, impor ou suprimir restrições de tempo; configuração do experimento: realizar o experimento em um ambiente muito diferente do ideal, ou adotar matérias ou instrumentos distantes da realidade.

3.5 AMEAÇAS À VALIDADE *VERSUS* DESENHO EXPERIMENTAL

Dentre os desenhos experimentais, existe um desenho que é considerado verdadeiramente experimental, pois a seleção dos participantes entre o grupo experimental e de controle é realizada de maneira aleatória. Dessa forma, algumas ameaças são neutralizadas ou minimizadas (Quadros 1, 2, 3 e 4), algo que não acontece em desenhos como os citados no final do parágrafo seguinte (WAINER, 2007). Destaca-se que os estudos selecionados e avaliados na próxima seção adotaram esse desenho.

O desenho comentado anteriormente é o seguinte: dois grupos, pós-teste e seleção aleatória (veja sua representação a seguir). Entretanto, quando a seleção não é aleatória, mas há grupo de controle, o desenho é chamado de quase experimento, e quando não há grupo de controle o desenho é chamado de pré-experimental – ou não-experimental em outras obras (WAINER, 2007). O Grupo 1 é o que sofre a intervenção, ele é o grupo experimental. O Grupo 2 é o que não sofre a intervenção, ou seja, o grupo de controle. Os sujeitos são selecionados e divididos (fase **A**) entre os grupos de forma aleatória e a observação (fase **O**) é realizada somente após o teste, ou intervenção (fase **X**).

<u>Grupo 1:</u>	A	X	O
<u>Grupo 2:</u>	A		O

Segundo Wainer (2007), a análise estatística apropriada seria algo como a comparação entre o resultado do pós-teste do Grupo 1 *versus* Grupo 2.

Ameaça à validade interna (VI)	Situação da ameaça <i>versus</i> experimento controlado	
	permanece	neutraliza/minimiza
Instrumentação	X	
Testagem		X
Maturação		X
História		X
Seleção		X
Mortalidade seletiva		X
Contaminação	X	
Comportamento competitivo	X	
Comportamento compensatório		X
Regressão à média		X
Efeito da expectativa do sujeito	X	
Efeito da expectativa do experimentador	X	
Influência de parte da intervenção	X	

Quadro 1. Situação das ameaças à validade interna *versus* experimento controlado

Fonte: elaborado pelos autores

Ameaça à validade externa (VE)	Situação da ameaça <i>versus</i> experimento controlado	
	permanece	neutraliza/minimiza
Participantes	X	
Tempo	X	
Configuração do experimento	X	

Quadro 2. Situação das ameaças à validade externa *versus* experimento controlado

Fonte: elaborado pelos autores

Ameaça à validade de construção (VCS)	Situação da ameaça <i>versus</i> experimento controlado	
	permanece	neutraliza/minimiza
Projeto do experimento	X	
Fatores humanos (ou sociais)	X	

Quadro 3. Situação das ameaças à validade de construção *versus* experimento controlado

Fonte: elaborado pelos autores

Ameaça à validade de conclusão (VCC)	Situação da ameaça <i>versus</i> experimento controlado	
	permanece	neutraliza/minimiza
Análise e interpretação estatística	X	
Confiabilidade das medidas	X	
Confiabilidade da implementação dos tratamentos	X	

Quadro 4. Situação das ameaças à validade de conclusão *versus* experimento controlado

Fonte: elaborado pelos autores

4 EXPERIMENTOS ENVOLVENDO PP

Nesta seção descrevemos a seleção dos estudos para análise, o método de identificação e classificação dos exemplos de ameaças à validade e, por fim, apresentamos os exemplos práticos analisados.

4.1 SELEÇÃO DOS ESTUDOS PARA ANÁLISE

A prática programação em par vem sendo investigada em inúmeros estudos como, por exemplo (separados conforme o principal método de pesquisa): (a) experimento: Balijepally *et al.* (2009), Choi *et al.* (2008), Hannay *et al.* (2010), Walle e Hannay (2009), Salleh *et al.* (2009), Salleh *et al.* (2011b), Sfetsos *et al.* (2008) e Williams *et al.* (2000); (b) revisão sistemática da literatura: Sfetsos e Stamelos (2010), Salleh *et al.* (2011a); (c) estudo de caso: Melo *et al.* (2011), Williams e Kessler (2000), Fila e Loui (2010) e Han *et al.* (2010).

Dentre os estudos citados no parágrafo anterior, foram selecionados para este trabalho os estudos que atenderam os seguintes critérios ou diretrizes: (a) o método de pesquisa adotado pelo estudo foi a investigação experimental, especificamente, o experimento; (b) o estudo teve seções e subseções específicas para as ameaças à validade; ou, caso não tenha apresentado este detalhamento, o estudo apresenta formas diferenciadas (ou mais rigorosas) de avaliar o desempenho das duplas *versus* esforço individual.

Além disso, procuramos selecionar no mínimo um estudo feito com profissionais e outro com estudantes universitários. Os estudos que atenderam os critérios descritos anteriormente e que foram selecionados para a análise são os seguintes:

1. Sfetsos *et al.* (2008) [Estudo A]: nesse estudo os pesquisadores realizaram um experimento controlado com 70 estudantes do curso de engenharia de software de um instituto tecnológico na Grécia. Nele foi investigado o impacto do temperamento e da personalidade do estudante sobre a comunicação, desempenho do par e viabilidade de colaboração. Por fim, nesse estudo há uma seção e várias subseções que descrevem a avaliação da validade da investigação experimental;

2. Balijepally *et al.* (2009) [Estudo B]: nesse estudo os pesquisadores realizaram um experimento controlado com 120 desenvolvedores –

graduados ou graduandos matriculados no curso de Sistemas de Informação de uma grande universidade pública nos Estados Unidos. Nele foram examinados os efeitos de diferentes meios de programação (pares *versus* esforço individual) e complexidades de tarefas sobre a qualidade de software, satisfação e a confiança no desempenho. Enfim, nesse estudo há uma modesta seção que descreve a avaliação da validade da investigação experimental. Entretanto, foi o único estudo que apresentou uma forma diferenciada de avaliar o desempenho das duplas *versus* esforço individual.

4.2 MÉTODO DE IDENTIFICAÇÃO E CLASSIFICAÇÃO DOS EXEMPLOS DE AMEAÇAS À VALIDADE

Para identificar e classificar os exemplos de ameaças à validade de um experimento envolvendo PP, um dos autores do presente trabalho leu os estudos selecionados e gerou uma tabela intermediária onde as ameaças relatadas nos estudos foram analisadas, resumidas e classificadas originalmente segundo os quatro tipos de validade definidos anteriormente.

Na elaboração da tabela intermediária foi considerado o seguinte: para o Estudo A foi mantido o nome que os pesquisadores definiram para cada ameaça. Para o Estudo B, ver “Correlação por Definição (CD)” descrita a seguir. Ressalta-se que cada ameaça foi identificada e classificada mesmo se o autor do estudo justificou a ausência dela no experimento ou se ele definiu um mecanismo de proteção para ela.

Na sequência, a tabela intermediária junto com restante do trabalho foi enviada aos outros dois autores. Eles revisaram o artefato e, em seguida, cada um enviou suas considerações para o relator. Com as considerações em mãos buscando um maior nível de detalhe e compreensão o relator elaborou uma nova tabela [re]classificando novamente as ameaças da seguinte forma:

1. Correlação total (ou CT): caso o tipo de validade, a identificação (nome) e a definição da(s) ameaça(s) relatada(s) pelo estudo em análise coincidiram com a literatura sobre ameaças à validade utilizada neste trabalho, ou somente literatura de referência;
2. Correlação parcial (ou CP): se somente o tipo de validade e a definição da(s) ameaça(s) relatada(s) pelo estudo em análise coincidiram com a literatura de referência.
3. Tipo de validade não correspondente - reclassificadas (RC): caso o tipo de validade da(s) ameaça(s) relatada(s) pelo estudo em análise não coincidiu com a literatura. Nessa situação a ameaça foi [re]classificada a partir da sua descrição segundo a literatura de referência;
4. Correlação por definição (CD): não havia descrição explícita da ameaça no estudo em questão. Logo, foi necessário definir as ameaças a partir do conteúdo das seções “Avaliação da validade” ou “Limitações” do estudo em questão, conforme as possibilidades descritas na literatura de referência.

Continuando o processo, essa nova tabela foi enviada novamente aos outros dois autores. Eles revisaram o artefato e, em seguida, cada um enviou suas considerações para o relator. A partir dessas considerações o relator atualizou a tabela e, na sequência, a dividiu em uma tabela para cada tipo de validade¹. Por fim, para facilitar a busca pela descrição ou exemplo da ameaça, foram gerados quadros de consolidação (Quadros 5, 6, 7 e 8).

4.3 IDENTIFICAÇÃO E CLASSIFICAÇÃO DOS EXEMPLOS DE AMEAÇAS À VALIDADE

Nos quadros apresentados a seguir pode-se identificar em qual estudo há a descrição ou exemplificação de uma determinada ameaça e sua correlação com a literatura de referência.

Ameaça à validade interna (VI)	Estudo onde a ameaça está descrita/exemplificada <i>versus</i> nível de correlação com a literatura			
	Correlação total (CT)	Correlação parcial (CP)	Correlação por definição (CD)	Tipo de validade não correspondente (RC)
Instrumentação	Estudo A		Estudo B	
<i>Testagem</i>				
Maturação	Estudo A			
História	Estudo A			Estudo A (2*): originalmente uma ameaça a VE e outra a VCC.
Seleção	Estudo A		Estudo B	
Mortalidade seletiva		Estudo A	Estudo B	
Contaminação		Estudo A		
Comportamento competitivo		Estudo A		
Comportamento compensatório		Estudo A		
<i>Regressão à média</i>				
<i>Efeito da expectativa do sujeito</i>				
Efeito da expectativa do experimentador				Estudo A: originalmente uma ameaça a VCS.
<i>Influência de parte da intervenção</i>				

Quadro 5. Resultado da análise dos estudos experimentais para as ameaças à validade interna (* Quantidade de diferentes descrições ou exemplos por ameaça)

Fonte: elaborado pelos autores

¹ Estão disponíveis em: <https://docs.google.com/open?id=0B-lo3hrzjKFXWWRqZ204NIB1LVk> ou http://vagnercml.files.wordpress.com/2012/08/wbma2012_lima_secaneto_emer_tabelas_intermediarias_v1-0.pdf.

Ameaça à validade externa (VE)	Estudo onde a ameaça está descrita/exemplificada <i>versus</i> nível de correlação com a literatura			
	Correlação total (CT)	Correlação parcial (CP)	Correlação por definição (CD)	Tipo de validade não correspondente (RC)
Participantes		Estudo A	Estudo B	
Tempo			Estudo B	Estudo A: originalmente uma ameaça a VI.
Configuração do experimento		Estudo A	Estudo B	

Quadro 6. Resultado da análise dos estudos experimentais para ameaças à validade externa (* Quantidade de diferentes descrições ou exemplos por ameaça).

Fonte: elaborado pelos autores

Ameaça à validade de construção (VCS)	Estudo onde a ameaça está descrita/exemplificada <i>versus</i> nível de correlação com a literatura			
	Correlação total (CT)	Correlação parcial (CP)	Correlação por definição (CD)	Tipo de validade não correspondente (RC)
Projeto do experimento		Estudo A(7*)		
Fatores humanos (ou sociais)		Estudo A(2*)	Estudo B	

Quadro 7. Resultado da análise dos estudos experimentais para ameaças à validade de construção

(* Quantidade de diferentes descrições ou exemplos por ameaça).

Fonte: elaborado pelos autores

Ameaça à validade de conclusão (VCC)	Estudo onde a ameaça está descrita/exemplificada <i>versus</i> nível de correlação com a literatura			
	Correlação total (CT)	Correlação parcial (CP)	Correlação por definição (CD)	Tipo de validade não correspondente (RC)
Análise e interpretação estatística	Estudo A(4*)			
Confiabilidade das medidas	Estudo A			
Confiabilidade da implementação dos tratamentos	Estudo A			

Quadro 8. Resultado da análise dos estudos experimentais para as ameaças à validade de conclusão (* Quantidade de diferentes descrições ou exemplos por ameaça.)

Fonte: elaborado pelos autores

5 DISCUSSÃO

Ao analisarmos os resultados obtidos e presentes nos quadros da seção anterior, vemos que foi possível identificar exemplos práticos para mais de 80% das ameaças mais comuns (17 das 21) relatadas por Wainer (2007) e Travassos *et al.* (2002).

Uma ameaça à validade interna é, segundo Wainer (2007), “uma outra possível explicação para o efeito observado que não a ação ou a manipulação”. Então, caso exista essa outra possível explicação, há uma ameaça e a validade do estudo pode ser questionada. Sendo assim, independentemente do tipo de validade, ou o pesquisador cita a ameaça e justifica o porquê de ela não ocorrer no seu estudo, ou ele cita a ameaça, assumindo sua presença, e descreve algum mecanismo de proteção que minimize ou neutralize seu efeito. Caso contrário, a validade do estudo poderá ser questionada.

Por exemplo, como ficariam as ameaças testagem, regressão à média, efeito da expectativa do sujeito e influência de parte da intervenção não citadas em ambos os estudos? A primeira e a segunda ameaças o estudo não precisaria necessariamente citar, pois, conforme Wainer (2007), elas são neutralizadas pelo desenho experimental escolhido. Porém, os autores do estudo poderiam ser questionados sobre as outras duas.

O estudo A apresentou uma quantidade alta de ameaças correlacionadas parcialmente em projeto do experimento e análise e interpretação estatística, fato que nos chamou a atenção e nos levou a analisar brevemente alguns dos referenciais teóricos utilizados pelo estudo. Constatou-se, então, que tais referenciais merecem uma investigação mais minuciosa, pois se demonstraram um pouco mais abrangentes do que o referencial teórico adotado neste trabalho.

O estudo B adotou uma forma diferenciada de avaliação de desempenho das duplas *versus* esforço individual. Em resumo, nela trabalha-se com o desempenho dos pares reais *versus* pares nominais. Um par nominal é resultado da combinação artificial (ou aleatória) de dois participantes que desenvolveram o software individualmente. Após essa combinação, os membros de cada par nominal são classificados em melhor e segundo melhor integrante da dupla, considerando, por exemplo, a variável qualidade de software. Então, avaliação do desempenho é realizada comparando o desempenho dos pares reais com os melhores membros das duplas artificiais, ou com os segundos melhores membros das duplas artificiais, ou ambos.

Algumas possíveis causas de ameaças à validade de experimentos envolvendo PP:

1. Dificuldade em lembrar a sintaxe da linguagem de programação (pode-se disponibilizar uma documentação *online* da respectiva linguagem) [Estudo B, Validade interna, Instrumentação];
2. Diferença de desempenho e de produtividade entre desenvolvedores [Estudo A, Validade interna, Seleção];

3. Os participantes não estarem familiarizados com a prática [Estudo B, validade externa, participantes];
4. Não houver um período de aquecimento para as duplas [Estudo B, validade externa, tempo];
5. Desmotivação por parte dos participantes em consequência do uso de uma documentação *online* da linguagem de programação [Estudo B, validade de construção, fatores humanos (ou sociais)];
6. Não implementar a prática de maneira correta [Estudo A, validade de conclusão, confiabilidade da implementação dos tratamentos];
7. Vários aspectos relacionados à análise estatística dos dados [Estudo A, validade de conclusão, análise e interpretação estatística].

Por fim, o pesquisador deve ficar atento a situações em que um mecanismo de proteção de uma determinada ameaça pode causar uma nova ameaça ou intensificar outra, por exemplo, situações 1 e 5 apresentadas anteriormente. E mais, o eventual questionamento por parte de alguns pesquisadores sobre a não validade do experimento quando são usados estudantes para representar profissionais de software, de acordo com Höst *et al.* (2000), em certas condições, os estudantes podem ser usados sem necessariamente causar ameaça a validade do experimento.

6 CONCLUSÃO

A realização de um experimento de qualidade envolvendo a prática ágil programação em par é algo custoso e arriscado. Pois, dentre vários elementos a serem definidos e planejados, vimos por meio deste trabalho que a quantidade de ameaças à validade de um experimento desse tipo é elevada (pelo menos 21 ameaças foram identificadas). Entretanto, considerando o nível de detalhe do Estudo A, a quantidade não se limitaria a esse número. Uma vez que apenas as ameaças relacionadas a Projeto do Experimento nesse estudo são sete e não apenas uma. E, dependendo a situação, caso o pesquisador desconsidere alguma dessas ameaças, o resultado do experimento se torna sem importância ou inválido, havendo a necessidade de realizar um novo experimento com algumas alterações ou até de recomeçar todo o estudo experimental.

Por outro lado, independentemente da prática ágil investigada, para diminuir o risco associado às ameaças à validade de um experimento, o pesquisador deve: (a) fazer a escolha correta do desenho experimental e (b) realizar análise e descrição das ameaças à validade a partir de um “catálogo de ameaças” mais abrangente e detalhado, inclusive especificando para cada ameaça o seu mecanismo de proteção e eventuais relacionamentos de causa desse mecanismo com outras ameaças.

Há uma forma diferenciada (ou mais rigorosa) de avaliação de desempenho das duplas *versus* esforço individual. Ela foi utilizada no estudo B.

Como trabalho futuro sugere-se complementar essa lista de ameaças a validade de experimentos envolvendo PP, buscando em outros estudos as ameaças não identificadas neste trabalho. Outro possível trabalho futuro é, usando o método apresentado aqui, realizar o mesmo estudo, mas abordando outra prática ágil que possa vir a ser objeto de investigações experimentais, por exemplo, desenvolvimento orientado a testes.

REFERÊNCIAS

BALIJEPAALLY, V.; MAHAPATRA, R.; NERUR, S.; PRICE, K. H. Are two heads better than one for software development? The productivity paradox of pair programming. *MIS Quarterly*, v. 33, n.1, p. 91-118, março, 2009.

BANDEIRA, M. Validade interna e externa de uma pesquisa: vieses. São João del-Rei: UFSJ. 2012. Disponível em: <http://www.ufsj.edu.br/portal2-repositorio/File/lapsam/Texto%204-VALIDADE.pdf>. Acesso em: 28/04/2012.

BECK, K. e ANDRES, C. *extreme programming explained: embrace chance*. Boston: Addison-Wesley, 2004.

BECK, K.; BEEDLE, M.; BENNEKUM, A. v.; COCKBURN, A.; CUNNINGHAM, W.; FOWLER, M.; GRENNING, J.; HIGHSMITH, J.; HUNT, A.; JEFFRIES, R.; KERN, J.; MARICK, B.; MARTIN, R. C.; MELLOR, S.; SCHWABER, K.; SUTHERLAND, J.; THOMAS, D. Manifesto for agile software development. 2001. Disponível em: <http://agilemanifesto.org>. Acesso em: 02/06/2012.

CHOI, K. S.; DEEK F. P.; IM I. Exploring the underlying aspects of pair programming: The impact of personality. *Information and Software Technology*, v. 50, n. 11, p. 1114-1126, 2008.

DYBÅ, T.; DINGSØYR, T. Empirical studies of agile software development: a systematic review. *Information and Software Technology*, v. 50, n. 9, p. 833-859, 2008.

DYBÅ, T.; DINGSØYR, T. What do we know about agile software development? *IEEE Software*, v. 26, n. 5, p. 6-9, 2009.

FILA, N. D.; LOUI, M. C. Work-in-progress - Who's driving? Structured pairs in an introductory electronics laboratory. *IEEE*, p. F3C-1 – F3C-2, 2010.

HÖST, M.; REGNELL, B.; WOHLIN, C. Using students as subjects - a comparative study of students and professionals in lead-time impact assessment. *Empirical Software Engineering*, v. 5, n. 3, p. 201-214, 2000.

HAN, K.-W.; LEE, E.; LEE, Y. The impact of a peer-learning agent based on pair programming in a programming course. *IEEE Transactions on Education*, v. 53, n. 2, p. 318-327, 2010.

HANNAY, J. E.; ARISHOLM, E.; ENGVIK, H.; SJØBERG, D. I. K. Effects of personality on pair programming. *IEEE Transactions on Software Engineering*, v. 36, n. 1, p. 61-80, 2010.

MAFRA, S. N.; TRAVASSOS, G. H. Estudos primários e secundários apoiando a busca por evidência em engenharia de software. Rio de Janeiro:

COPPE/UFRJ, Rio de Janeiro, 2006. Disponível em: <http://www.cos.ufrj.br/uploadfiles/1149103120.pdf>. Acesso em: 29/04/2012.

MELO, C.; CRUZES, D. S.; KON, F.; CONRADI, R. Agile team perceptions of productivity factors. In: AGILE 2011 Conference, Salt Lake City. *Proceedings...* Rally Software and VersionOne, 2011.

MELO, C.; SANTOS, V. A.; CORBUCCI, H.; KATAYAMA, E.; GOLDMAN, A.; KON, F. Métodos ágeis no Brasil: estado da prática em times e organizações. São Paulo: Departamento de Ciência da Computação do IME/USP, 2012. Disponível em: <http://www.agilcoop.org.br/files/MetodosAgeisBrasil2011.pdf>. Acesso em: 12/03/2012.

POPPENDIECK, M. e POPPENDIECK, T. *Implementando o desenvolvimento lean de software: do conceito ao dinheiro*. Porto Alegre: Bookman, 2011.

SALLEH, N.; MENDES, E.; GRUNDY, J. Empirical studies of pair programming for CS/SE teaching in higher education: a systematic literature review. *IEEE Transactions on Software Engineering*, v. 37, n. 4, p. 509-525, 2011[a].

SALLEH, N.; MENDES, E.; GRUNDY, J. The effects of openness to experience on pair programming in a higher education contexto. In: Conference on Software Engineering Education and Training, 24. ed., *Proceedings...* IEEE, 2011[b].

SALLEH, N.; MENDES, E.; GRUNDY, J.; BURCH, G. An empirical study of the effects of personality in pair programming using the five-factor model. In: International Symposium on Empirical Software Engineering and Measurement, 3. ed., Lake Buena Vista. *Proceedings...* IEEE, 2009.

SFETSOS, P.; STAMELOS, I.; ANGELIS, L.; DELIGIANNIS, I. An experimental investigation of personality types impact on pair effectiveness in pair programming. *Empirical Software Engineering*, v. 14, p. 187-226, 2008.

SFETSOS, P.; STAMELOS, I. Empirical studies on quality in agile practices: a systematic literature review. In: International Conference on the Quality of Information and Communications Technology, 7. ed., *Proceedings...* IEEE, 2010.

SHORE, J. e WARDEN, S. *A arte do desenvolvimento ágil*. Rio de Janeiro: Alta Books, 2008.

TELES, V. M. *Extreme programming: aprenda como encantar seus usuários desenvolvendo software com agilidade e alta qualidade*. São Paulo: Novatec, 2004.

TRAVASSOS, G. H. Experimentação em engenharia de software: fundamentos e conceitos. In: Simpósio Brasileiro de Qualidade de Software, 10 ed., Curitiba. *Anais...* 2011.

TRAVASSOS, G. H.; GUROV, D.; AMARAL, E. A. G. Introdução à engenharia de software experimental. Rio de Janeiro: COPPE/UFRJ, 2002. Disponível em:

<http://www.ufpa.br/cdesouza/teaching/topes/4-ES-Experimental.pdf>. Acesso em: 07/03/2012.

VERSIONONE. State of agile development. Atlanta: VersionOne, 2011. Disponível em: http://www.versionone.com/pdf/2011_State_of_Agile_Development_Survey_Results.pdf. Acesso em: 20/03/2012.

WAINER, J. Métodos de pesquisa quantitativa e qualitativa para a Ciência da Computação. Porto Alegre: Sociedade Brasileira de Computação, 2007.

WALLE, T.; HANNAY, J. E. Personality and the nature of collaboration in pair programming. In: International Symposium on Empirical Software Engineering and Measurement, 3 ed., *Proceedings...* IEEE, 2009.

WILLIAMS, L. A.; KESSLER, R. R. All I really need to know about pair programming I learned in Kindergarten. *Communications of the ACM*, v. 43, n. 5, p. 108-114, 2000.

WILLIAMS, L.; KESSLER, R. R.; CUNNINGHAM, W.; JEFFRIES, R. Strengthening the case for pair programming. *IEEE Software*, v. 17, n. 4, p. 19-25, 2000.