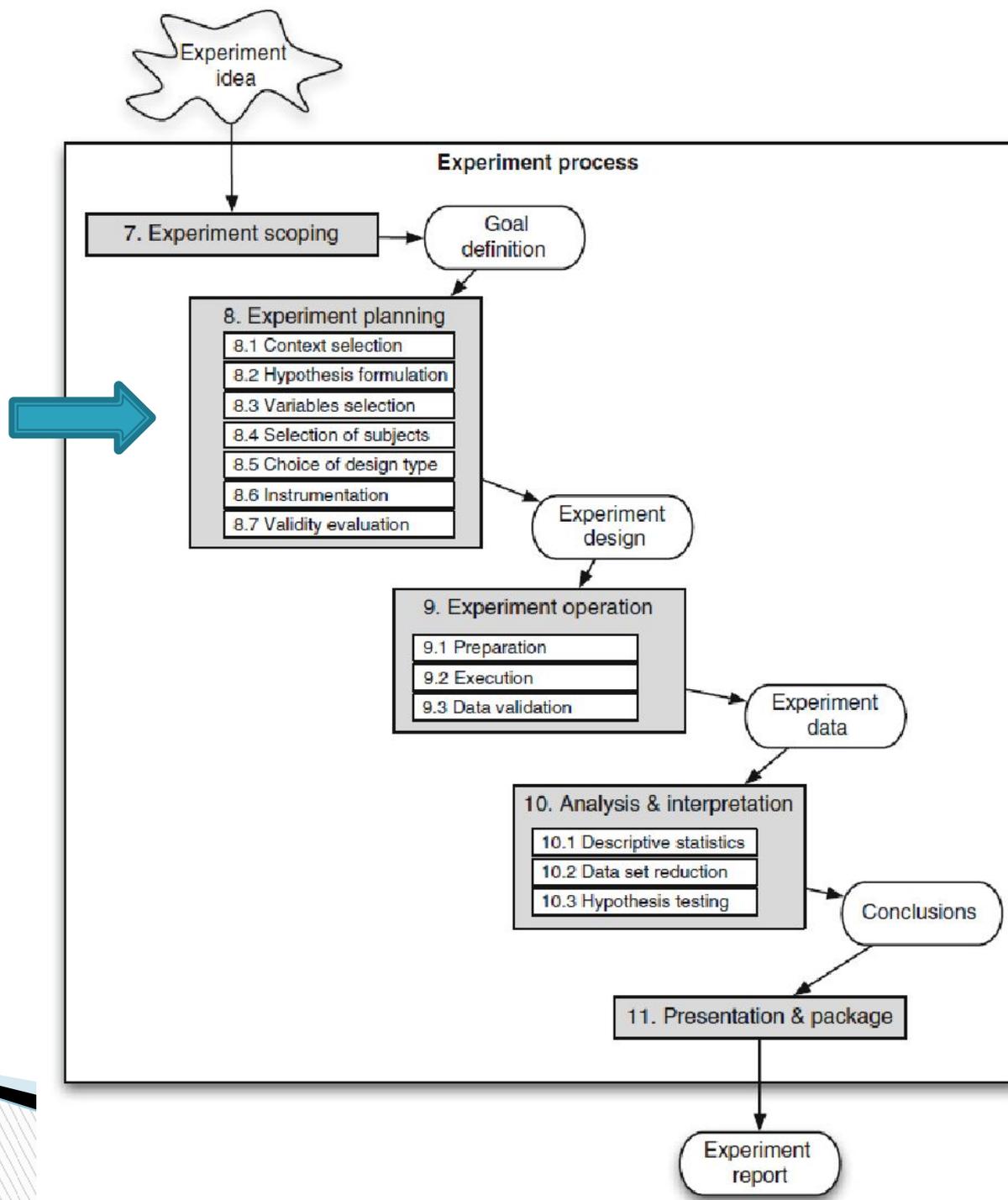


Planejamento e Avaliação

Capítulo 8 - Wohlin

Adaptado a partir do material elaborado pelo Prof. Paulo Masiero





Organização

- Introdução
- 1. Seleção do contexto
- 2. Formulação das hipóteses
- 3. Seleção das variáveis
- 4. Seleção dos sujeitos
- 5. Design do experimento
- 6. Instrumentação
- 7. Avaliação da Validade do experimento

Introdução

- O objetivo desta fase é gerar um documento de planejamento, que servirá de base para a execução e análise.
- Os verbos devem estar (quase) todos no futuro do presente
- O documento pode ser alterado, a partir de uma execução-piloto.
- O que se escreve em um artigo ou na tese ou dissertação é diferente do documento de planejamento
 - Quais as diferenças?

Introdução (cont.)

- ❑ Na tese o experimento todo deve ser descrito em um capítulo.
- ❑ É apresentado apenas um resumo do planejamento, que deve ser referenciado no texto. O foco é nas decisões mais importantes. Aqui os verbos podem ir para o passado.
- ❑ O documento de planejamento (ou parte dele) pode ir para o apêndice ou colocado em um arquivo acessível por um link.
- ❑ Idem para os formulários de coletas de dados e resultados coletados, se forem numerosos.

Introdução (cont.)

- ❑ Em um artigo, a escrita deve focar primordialmente na análise dos resultados do experimento.
- ❑ Recomenda-se que no máximo 20% a 30% do artigo seja dedicado a descrever o escopo, planejamento e execução.
- ❑ O documento de planejamento completo pode ir para o apêndice ou ser disponibilizado em um arquivo acessível por um link (Idem para os formulários de coletas de dados e resultados coletados)
- ❑ Na análise, aprofundar na discussão!

1. Seleção do contexto

- Objetivo geral: obter resultados que sejam os mais gerais possíveis em um experimento.
- Para isso, selecionar projetos reais com profissionais (ou com estudantes?)
 - Riscos de atrasos, os projetos podem não ter os problemas que se quer estudar; outros riscos devidos a complexidade, tamanho, custos.
- Alternativa: executar projetos em paralelo com os projetos reais que podem ser mais controlados.

1. Seleção do contexto (cont.)

- Resumindo, considerar:
 - Projetos on-line vs offline
 - Equipes de profissionais vs estudantes
 - Problemas reais vs “de brinquedo”
 - Generalidade vs especificidade.

- Experimento de Basili (métodos de inspeção por leitura):
 - Como caracterizar o método antigo?
 - Aprender um novo método pode influenciar como o antigo é aplicado.

2. Formulação das hipóteses

- Uma hipótese é uma teoria ou suposição que pode explicar um determinado comportamento de interesse da pesquisa.
- A base para a análise estatística é o teste de hipóteses.
- Um estudo experimental tem como objetivo coletar dados, em um ambiente controlado, para confirmar ou negar a hipótese.

Testes de Hipóteses

- Hipóteses avaliadas por testes estatísticos definidos pelos pesquisadores da estatística inferencial
- Normalmente são definidas duas hipóteses
 - **Hipótese nula (H_0):** indica que não há padrões ou tendências subjacentes no experimento e que as diferenças observadas no estudo são coincidências, ou seja, é a hipótese que o analista deseja rejeitar com a maior significância possível
 - **Hipótese alternativa (H_1):** é a hipótese inversa à hipótese nula, que será aceita caso a hipótese nula seja rejeitada
- Os testes estatísticos verificam se é possível rejeitar a hipótese nula, de acordo com um conjunto de dados observados e suas propriedades estatísticas

Exemplos

- Hipótese nula – Um novo método de inspeção encontra em média o mesmo número de falhas que o antigo.
 - $H_0: \mu_{\text{old}} = \mu_{\text{new}}$
- Hipótese alternativa: Um novo método de inspeção encontra em média mais falhas que o antigo.
 - $H_A: \mu_{\text{old}} < \mu_{\text{new}}$
- Dois erros são possíveis: rejeitar uma hipótese verdadeira ou aceitar uma hipótese falsa.

Exemplo: Definição de Hipóteses e Seleção de Variáveis

5. PLANNING (detailed)

Hypothesis Formulation:

Example 1

➤ **O1': Do PBR teams detect more defects than Checklist teams?**

H0: There is no difference in the defect detection rates of teams applying PBR compared to teams applying the Checklist technique. That is, every successive dilution of a PBR team with a non-PBR reviewer has only random effects on team scores.

Ha: The defect detection rates of teams applying PBR are higher compared to teams using the Checklist technique. That is, every successive dilution of a PBR team with a non-PBR reviewer decreases the effectiveness of the team

RT=reading technique
DOC = requirements document
ATM = sistema ATM
PG = sistema Parking Garage

5. PLANNING (detailed)

Hypothesis Formulation:

Example 2

- **O2': Do individual PBR or Checklist reviewers find more defects?**

Group effect (RT X DOC interaction)

H0: There is no difference between Group 1 and Group 2 with respect to individual effectiveness/efficiency.

Ha: There is a difference between Group 1 and Group 2 with respect to individual effectiveness/efficiency

Main effect RT

H0: There is no difference between subjects using PBR and subjects using Checklist with respect to individual effectiveness/efficiency.

Ha: There is a difference between subjects using PBR and subjects using Checklist with respect to individual effectiveness/efficiency.

Main effect DOC

H0: There is no difference between subjects reading ATM and subjects reading PG with respect to individual effectiveness/efficiency.

Ha: There is a difference between subjects reading ATM and subjects reading PG with respect to individual effectiveness/efficiency.

3. Seleção das Variáveis

- Escolha das variáveis independentes
 - São aquelas que podemos controlar e mudar no experimento
 - Devem ter algum efeito sobre as variáveis dependentes
 - Exige conhecimento do domínio
- Escolha das variáveis dependentes
 - O efeito do tratamento é medido nas variáveis dependentes.
 - A escolha das variáveis dependentes implica que as escalas de medida e os intervalos das variáveis são determinados.
 - As hipóteses devem ser refinadas quando as variáveis dependentes forem escolhidas

Hipóteses e Variáveis

- Hipóteses levam à definição de variáveis
- Variáveis independentes (ou fatores, quando controladas)
 - Referem-se à entrada do processo de experimentação, podendo ser controladas durante este processo
 - Representam a causa que afeta o resultado do processo de experimentação. Quando é possível seu controle, os valores são chamados de "tratamentos"
- Variáveis dependentes
 - Referem-se à saída do processo de experimentação, sendo afetadas durante o processo de experimentação
 - Representam o efeito da combinação dos valores das variáveis independentes (incluindo os fatores). Seus possíveis valores são chamados de "resultados"

Fator: Wohlin

An experiment studies the effect of changing one or more independent variables. Those variables are called *factors*. The other independent variables are controlled at a fixed level during the experiment, or else we cannot say if the factor or another variable causes the effect. A *treatment* is one particular value of a factor.

5. PLANNING (detailed)

Variables Selection:

➤ **Independent Variables**

- **Reading techniques:** we have two alternatives: the PBR technique and an usual technique like Checklist.
- **Perspectives:** Within PBR, a subject uses a technique based on one of the review perspectives. For this experiment we used the three perspectives previously described: Designer, Tester and User.
- **Requirements documents** (Problem Domain)
- **Subjects Experience**

➤ **Dependent Variables**

- **Effectiveness** in defect detection
- **Efficiency** to apply the techniques

4. SELEÇÃO DOS SUJEITOS

- É também chamada de “amostra da população”.
- A seleção é altamente relacionada com a generalização dos resultados. Deve ser representativa da população.
- A amostra pode ser **probabilística** ou **não-probabilística** (indica se a probabilidade de escolher um sujeito é conhecida ou não).

Amostragens probabilísticas

- ❑ **Amostragem aleatória simples:** os sujeitos são escolhidos aleatoriamente de uma lista da população.
- ❑ **Amostragem aleatória estratificada:** a população é dividida em grupos ou estratos, com uma distribuição conhecida entre os grupos. Amostragem aleatória é aplicada a cada grupo.
- ❑ **Amostragem sistemática:** o 1º Sujeito é escolhido aleatoriamente da lista da população e em seguida é escolhido cada n -ésimo elemento da lista.... (implica em ordenação da população)

Amostragens não-probabilísticas

- ▣ **Amostragem por conveniência**: as pessoas mais próximas e mais convenientes são selecionadas.
- ▣ **Amostragem por quotas**: o objetivo é obter elementos variados da população.
Amostragem por conveniência é usada para selecionar cada elemento.

Sugestões

- O tamanho da amostra impacta os resultados quando se generaliza.
- Princípios gerais:
 - Se há grande variabilidade na população, uma amostra maior é necessária.
 - A forma como se analisará os dados coletados influencia a escolha do tamanho da amostra. Então é necessário definir, antecipadamente, como os dados serão analisados.

5. DESIGN DO EXPERIMENTO

- Consiste de uma série de testes do tratamento.
- Descreve cuidadosamente como os testes são organizados e executados.
- As análises estatísticas a serem usadas dependem do design escolhido e das escalas de medidas.
- É preciso também se basear nas hipóteses para definir as estatísticas a serem usadas.

Princípios gerais - Aleatorização

- Aplica-se à escolha dos objetos, sujeitos e em que ordem os testes serão executados.
- É usado para “levar para a média” o efeito de um fator que pode estar presente em caso contrário.
- Exemplos: que programa escolher para inspeção? que analistas da empresa escolher para participar?

Princípios gerais – Grupamento (blocking)

- Usualmente há algum fator (variável independente) que provavelmente tem um efeito sobre o resultado, mas não estamos interessados nesse efeito.
- Em um grupo, o efeito indesejado do tratamento é o mesmo e podemos usar o efeito do tratamento dentro do grupo.
- Exemplo: Os sujeitos escolhidos têm diferentes experiências: alguns usaram OO antes e outros não. Então dividimos os sujeitos em dois grupos.

Princípios gerais - Balanceamento

- Há um balanceamento quando atribuímos o tratamento a um número igual de sujeitos.
- É desejável porque simplifica e torna mais forte a análise estatística dos dados, mas não é necessária.
- Exemplo: há o mesmo número de sujeitos em cada bloco.

Exemplos típicos de design

- Um fator com dois tratamentos
- Um fator com mais de dois tratamentos
- Dois fatores com dois tratamentos
- Mais do que dois fatores, cada um com dois tratamentos.

Um fator com dois tratamentos

- Em experimentos deste tipo, quer-se comparar os dois tratamentos um com o outro.
- O mais comum é comparar as médias das variáveis dependentes para cada tratamento.
- Notação – seja:
 - μ_i - a média da variável dependente para o tratamento i
 - Y_{ij} - a j -ésima medida da variável dependente para o tratamento i

Um fator com dois tratamentos

- Exemplo

- Um novo método de projeto produz software com mais qualidade do que o método usado anteriormente.
- Fator - o método de projeto
- Tratamentos - o método novo e o velho.
- Variável dependente : número de falhas encontradas durante o desenvolvimento.

Table 8.1 Example of assigning subjects to the treatments for a randomized design

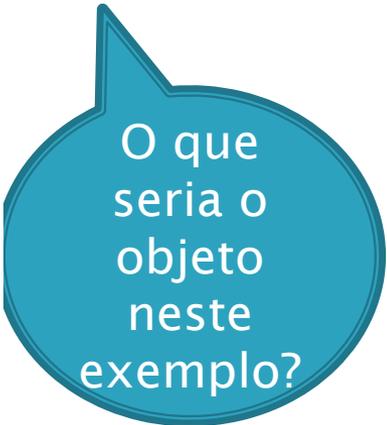
Subjects	Treatment 1	Treatment 2
1	X	
2		X
3		X
4	X	
5		X
6	X	

Example of hypothesis:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2, \mu_1 < \mu_2 \text{ or } \mu_1 > \mu_2$$

Examples of analysis: t-test, Mann-Whitney, see Sect. 10.3.



O que seria o objeto neste exemplo?

Exemplo : projeto completamente aleatório

- ▣ Usar o mesmo objeto experimental para os dois tratamentos.
- ▣ Atribuir os sujeitos aleatoriamente a cada tratamento
- ▣ Cada sujeito usa apenas um tratamento em um objeto.
- ▣ Se o número de sujeitos por tratamento é o mesmo, o projeto é balanceado

Projeto com comparação pareada (paired)

- Este projeto melhora a precisão dos resultados.
- Cada sujeito usa ambos os tratamentos no mesmo objeto.
- Para minimizar o efeito da ordem em que os tratamentos são aplicados, a ordem é atribuída aleatoriamente aos sujeitos.
- Não pode ser aplicado em qualquer caso de comparação, pois o sujeito pode obter muita informação do primeiro experimento e usar no segundo tratamento.
- Se o número de sujeitos iniciando com o primeiro tratamento é o mesmo que o do segundo □ balanceado.

Table 8.2 Example of assigning the treatments for a paired design

Subjects	Treatment 1	Treatment 2
1	2	1
2	1	2
3	2	1
4	2	1
5	1	2
6	1	2

Example of hypothesis:

$d_j = y_{1j} - y_{2j}$ and μ_d is the mean of the difference.

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d \neq 0, \mu_d < 0 \text{ or } \mu_d > 0$$

Examples of analysis: Paired t-test, Sign test, Wilcoxon, see Sect. 10.3.

Y_{ij} - a j -ésima
medida da variável
dependente para o
tratamento i

Um fator, com mais de dois tratamentos

- O objetivo é comparar um tratamento com os outros
- A comparação é usualmente baseada nas médias
- Exemplo: Pesquisar a qualidade do software quando se usa diferentes linguagens de programação.
 - Fator: linguagens de programação
 - Tratamentos: C, C++ e Java

Um fator, com mais de dois tratamentos

Projeto completamente aleatorizado

- O experimento deve ser feito em uma ordem aleatória, de tal forma que os tratamentos são usados em um ambiente tão uniforme quanto possível.
- O projeto do experimento usa um objeto para todos os tratamentos
- Os sujeitos são atribuídos aleatoriamente aos tratamentos.

Table 8.3 Example of assigning the treatments to the subjects

Subjects	Treatment 1	Treatment 2	Treatment 3
1		X	
2			X
3	X		
4	X		
5		X	
6			X

Example of hypothesis, where a is the number of subjects:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_a$$

$$H_1 : \mu_i \neq \mu_j \text{ for at least one pair } (i, j)$$

Examples of analysis: ANOVA (ANalysis Of VAriance) and Kruskal-Wallis, see Sect. 10.3.

Um fator, com mais de dois tratamentos

Projeto com blocos completos aleatórios

- É um dos tipos de projetos de experimentos mais usado.
- Útil quando a variabilidade entre os sujeitos é grande.
- Cada sujeito usa todos os tratamentos, isto é o projeto é “bloqueado” nos sujeitos.
- Usa um objeto para todos os tratamentos e a ordem de uso é aleatória.

Table 8.4 Example of assigning the treatments to the subjects

Subjects	Treatment 1	Treatment 2	Treatment 3
1	1	3	2
2	3	1	2
3	2	3	1
4	2	1	3
5	3	2	1
6	1	2	3

Example of hypothesis:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_a$$

$$H_1 : \mu_i \neq \mu_j \text{ for at least one pair } (i, j)$$

Examples of analysis: ANOVA (ANalysis Of VAriance) and Kruskal-Wallis, see Sect. 10.3.

Dois Fatores

- O experimento fica mais complexo
- A hipótese única com um fator se divide em três hipóteses:

τ_i	The effect of treatment i on factor A.
β_j	The effect of treatment j on factor B.
$(\tau\beta)_{ij}$	The effect of the interaction between τ_i and β_j .

Dois fatores com dois tratamentos

- ▣ Atribui-se aleatoriamente os sujeitos a cada combinação de tratamento.
- ▣ Exemplo: Um experimento para analisar a facilidade de entendimento de um documento de projeto usando projeto estruturado e projeto OO e dois documentos de requisitos, um “bom” e um “ruim”.

Dois fatores com dois tratamentos

Table 8.5 Example of a 2*2 factorial design

		Factor A	
		Treatment A1	Treatment A2
Factor B	Treatment B1	Subject 4, 6	Subject 1, 7
	Treatment B2	Subject 2, 3	Subject 5, 8

Example of hypothesis:

$$H_0 : \tau_1 = \tau_2 = 0$$

H_1 : at least one $\tau_i \neq 0$

$$H_0 : \beta_1 = \beta_2 = 0$$

H_1 : at least one $\beta_j \neq 0$

$$H_0 : (\tau\beta)_{ij} = 0 \text{ for all } i, j$$

H_1 : at least one $(\tau\beta)_{ij} \neq 0$

Example of analysis: ANOVA (ANalysis Of VAriance), see Sect. 10.3.

T_i: efeito do tratamento i no fator A

B_j: efeito do tratamento j no fator B

(TB)_{ij}: efeito da interação entre T_i e B_j

Tabela do Marco Mello

Qual teste estatístico devo usar?

<http://marcoarmello.wordpress.com>

por Marco Mello
(adaptado de
Jutta Schmid)

Qual é a distribuição dos meus dados?

variável independente qualitativa ou quantitativa

(V3, 19/09/2015)

Há uma relação entre as minhas variáveis?

variável independente nominal/ordinal,
variável dependente quantitativa³

categorias independentes categorias ligadas

distrib. normal distrib. outra distrib. normal distrib. outra

teste t Mann-Whitney teste t pareado Wilcoxon

1 variável independente
2 categorias

1 variável independente
>2 categorias

ANOVA unifatorial Kruskal-Wallis ANOVA de medidas repetidas Friedman

ANOVA multifatorial GLM ou GAM²

>1 variável independente

todas as variáveis quantitativas³

sem causalidade com causalidade

distrib. normal distrib. outra distrib. normal distrib. outra

correlação de Pearson correlação de Spearman regressão linear/não-linear simples regressão não-paramétrica

1 variável independente

>1 variável independente

correlação parcial/múltipla correlação de Kendall regressão múltipla/stepwise SEM e análise de caminhos⁴

GLM ou GAM²

1 variável independente

Shapiro-Wilk¹

Kolmogorov-Smirnov¹

>1 variável independente

qui-quadrado
teste G
teste exato de Fisher
teste binomial

1. Costumam ser usados para testar a normalidade dos dados.

2. Pode-se usar também modelos lineares mistos quando há variáveis independentes quantitativas e qualitativas no teste. GAM pode ser usado quando se espera relações não-lineares entre as variáveis.

3. Se a variável independente for quantitativa, mas a variável dependente for nominal e binária (e.g., sim ou não), você pode usar uma regressão logística ou GLM com função logit.

4. Estas análises são especialmente recomendadas quando se quer testar relações diretas e indiretas entre diversas variáveis.

Exemplo

Projeto de Experimento e

Seleção de Participantes

5. PLANNING (detailed)

Experiment Design:

- follows the principles of blocking and balancing
- two factors and two treatments for each factor
- 2*2 factorial design

- **Factor A:** Technique
- **Treatments:** PBR and Checklist

- **Factor B:** Requirements Document
- **Treatments:** ATM and PG documents

	Group 1			Group 2			
	<i>Designer</i> 3 Subjects	<i>Tester</i> 3 Subjects	<i>User</i> 3 Subjects	<i>Designer</i> 3 Subjects	<i>Tester</i> 3 Subjects	<i>User</i> 3 Subjects	
Checklist	Training in Checklist						First Day
	ATM document inspection			PG document inspection			
PBR Technique	Training in PBR						Second Day
	PG document inspection			ATM document inspection			

http://www.cs.umd.edu/projects/SoftEng/ESEG/manual/pbr_package/manual.html

ID	Perspective			Thread	
	Design(1)	Test(2)	Use(3)	1	2
01	X			X	
02		X		X	
03			X	X	
04	X				X
05		X			X
06			X		X
07	X			X	
08		X		X	
09			X	X	
10	X				X
11		X			X
12			X		X
13	X			X	
14		X		X	
15			X	X	
16	X				X
17		X			X
18			X		X

5. PLANNING (detailed)

Subjects Selection:

- 18 undergraduate students with slightly more than one year of classroom experience on average, from the Software Engineering course at University of São Paulo at São Carlos, randomly divided into two groups of nine

6. INSTRUMENTAÇÃO

- Um dos passos mais trabalhosos de um experimento é o desenvolvimento de toda a instrumentação necessária à sua realização
- Os instrumentos são escolhidos no planejamento. Antes da execução eles devem ser criados para o experimento específico.
- São de três tipos: objetos, diretrizes e instrumentos de medição.

6. INSTRUMENTAÇÃO

(Cont.)

- Mecanismos de coleta de dados
 - *Documento de consentimento*
 - *Formulários de caracterização de participantes*
 - *Formulários de coleta de dados*
 - *Formulários de feedback*
- Validar os instrumentos antes de usar pela primeira vez.
- A coleta pode ser feita por entrevista, automaticamente ou preenchida pelos sujeitos.

5. PLANNING (detailed)

Instrumentation:

➤ Artifacts:

- Consent Form 
- Analyst Survey 
- List of Participants 
- Training Material 
- Objects: ATM and PG Requirement Documents 
- Defect Form 
- Defect List 
- Feedback Questionnaire 

Formulário de Participação

Project title and purpose

The experiment is commonly called "The Reading Experiment" and is intended to compare the effectiveness, in terms of defect detection, of various techniques for reading requirements specifications. For this particular experiment, a technique called "Perspective-based Reading" will be compared to the standard reading technique currently applied in the university and industry.

Statement of age

I state that I am over 18 years of age and wish to participate in an experiment conducted by Dr. José Carlos Maldonado at the Universidade de São Paulo and Dr. Sandra Camargo Pinto Ferraz Fabbri at the Universidade Federal de São Carlos, under the scope of the **NSF-CNPq Readers Project**.

Procedures

The experiment involves two sessions, each of about 6 hours. The first session is held on ___/___/___ and the second is held on ___/___/___. Each session will include various presentations, tutorials, training, and sample documents that are to be read and reviewed.

Confidentiality

All information collected in the experiment is confidential, and my name will not be identified at any time.

Survey do Analista

Reviewer Background Questionnaire – E1

Reviewer ID: _____ Date: ___/___/___

This form asks you a few questions about your background and experience.

General Questions

1. How many years/months of experience have you had in each function?

Time	Years	Months
Manager		
Developer		
Tester		
Analyst		
Other _____		

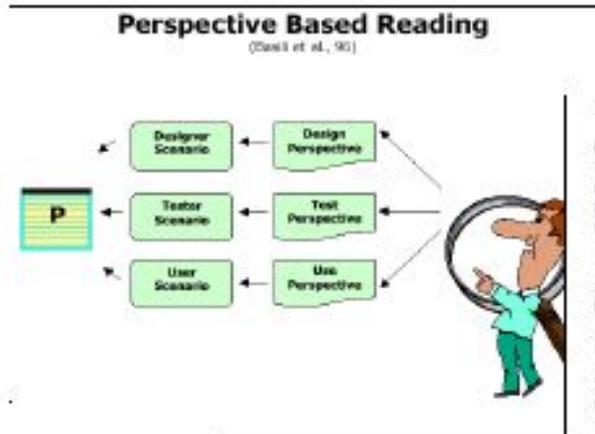
2. How comfortable are you with reading or reviewing requirements documents?
[Please mark with an x in the scale]

Value	0	1	2	3
Comfort level	Not at all	Low	Moderate	High

Lista de Participantes

ID #	Name (Print)	Signature
ID 1		
ID 2		
ID 3		
ID 4		
ID 5		
ID 6		
ID 7		
ID 8		

Material de Treinamento



Test-based Reading

For each Requirement/functional specification, generate a test or set of tests that allow you to ensure that an implementation of the system satisfies the requirement/functional specification. Use your standard test approach and technique, and incorporate test criteria in the test suite. In doing so, ask yourself the following questions for each test.

1. Do you have all the information necessary to identify the item being tested and the test criteria? Can you generate a reasonable test case for each item based upon the criteria?

Page 11: Functional requirement 10: What it means to "update" the different files never is specified.

2. Can you be sure that the tests generated will yield the correct values in the correct units?

Page 9: Functional requirement 3: "information" is not specified. There are

requirements that the implementer might interpret differently.

Fault Taxonomy

Omission

- **Missing Functionality:** Information describing internal operational behavior of the system has been omitted from the SRS.
- **Missing Performance:** Information describing performance specifications has either been described in a way that is unacceptable for testing.
- **Missing Interface:** Information describing how the system will interface and communicate with other systems.

DEFECT REPORT - FORM E5

Reviewer ID: _____ Date: ___/___/___

Page 1 of 1

Document name: ABC Video System

Defects				
# Defect	Page	Class	Description	
1	11	O	FR10: What it means to "update" the different files is never specified.	
2	9	O	FR3: "Information" is not specified. There are several interpretations for this	

Formulário de Defeitos

DEFECT REPORT – FORM E5

Reviewer ID: _____ Date: ___/___/___

Page 1 of 1

Document name: **ABC Video System**

Defects			
# Defect	Page	Class	Description
1	11	O	FR10: What it means to "update" the different files is never specified.
2	10	O	FR6: It is not specified that the account number must be entered before the number of a tape can be entered.
3	9	O	FR3: What is a "transaction record"?

Lista de defeitos

Defects for the Generic Documents

1. Defect list for the generic documents

The defects lists are written according to the format defined by the tables below. Please note that page number refers to the page numbering of the separately printed documents, i.e. not as they appear in this manual. If the defect is pertinent to one requirement, the number of the requirements as it appears in the generic document should also be indicated here.

The following classification is used for analysis:

Ambiguous information
E – Extraneous information
I – Inconsistent Information
IF – Incorrect Fact
MD – Miscellaneous Defect
MI – Missing Information

Noting the type of effort required to

O – Error of Omission

It is necessary to add information to

C – Error of Commission

It is necessary to edit/delete inform

Defects for the Generic Documents

1.1 ABC Video System

Def. #	Page	Req. #	O/C	Class	Description
1	8	FR1	O	MI	Only clerk functions are listed. How manager functions are accessed is not clear.
2	9	FR2	O	MI	Definition omitted: "current status".
3	9	FR3	O	MI	Information required for the transaction needed is not crucial!

Questionário de feedback

Feedback Questionnaire – Form E7

Reviewer ID: _____ Date: ___/___/___

This form asks you a few questions about the experiment itself after it is conducted.

General Questions

1. Please assess the explanations before and throughout the training session.
[Please mark with an x in the scale]

	0	1	2
Training, Explanation	Not Enough	Enough	Too Much

2. Did you need more time than allocated for reviewing the documents?

Yes	<input type="checkbox"/>
No	<input type="checkbox"/>

7. Validade do experimento

- Todos os experimentos têm riscos à validade de seus resultados: Quão válidos são eles?
- O projetista do experimento tem a obrigação de identificar e discutir esses riscos, ainda no planejamento
- As principais categorias de riscos são:
 - Validade da conclusão
 - Validade interna
 - Validade externa
 - Validade da construção

Validade da conclusão

- Trata-se da habilidade de chegar a uma conclusão correta a respeito dos relacionamentos entre o tratamento e o resultado do experimento.
- É preciso considerar:
 - O teste estatístico a utilizar
 - Baixo poder da estatística ou suposições inadequadas (ex. distribuição dos resultados)
 - Confiabilidade das medidas
 - Formulário mal projetado, questões mal formuladas...

Validade da conclusão (Cont.)

- É preciso considerar:
 - Pescaria e Taxa de erros
 - Evitar “pescar” um resultado: a análise não seria independente e o pesquisador influenciaria o resultado por estar procurando um resultado específico.
 - A taxa de erro está relacionada com a significância do teste. Ex. três experimentos relacionados com um nível de significância de 0.05 implica na significância total de $1 - (1 - 0.05)^3 = 0,14$
 - Irrelevâncias aleatórias no ambiente de operação
 - Heterogeneidade aleatória dos sujeitos

Validade interna

- É a verdade aproximada sobre relações de causa e efeito (relações causais).
- Validade interna só é relevante em estudos que tentam estabelecer uma relação causal.
- A questão-chave na validade interna é se as mudanças observadas podem ser atribuídas ao tratamento e **não** a outras causas possíveis (às vezes chamadas de explicações alternativas ou fatores de confusão)
- Há ameaças a experimentos com grupos únicos e com múltiplos grupos

Validade interna – Grupo único

- História, quando as aplicações ocorrem em diferentes datas (por exemplo, numa delas ocorre uma catástrofe nacional).
- Mortalidade : sujeitos que desistem
- Teste: se o teste é repetido mais de uma vez, a partir da segunda pode haver diferença.
- Maturação: os sujeitos reagem diferentemente com o passar do tempo

Validade de Construção

- Considera os relacionamentos entre a teoria e a observação, isto é, se o tratamento reflete bem a causa e o resultado reflete bem o efeito.
- Os problemas podem surgir por falhas do experimentador ou dos participantes:
 - Os participantes podem basear seu comportamento em suposições sobre a hipótese.
 - O ser humano geralmente tenta parecer melhor do que é quando está sendo avaliado.
 - Os pesquisadores podem projetar o experimento pensando nos resultados que esperam (viés).

Validade externa

- ▣ Trata da validade da generalização dos resultados do experimento para a prática industrial.
- ▣ Deve-se observar:
 - A interação entre a seleção dos sujeitos e o tratamento
 - A interação entre o ambiente de execução e o tratamento
 - A interação entre a história e o tratamento
 - Ex. Um questionário sobre sistemas críticos respondido alguns dias depois de uma grande falha de software.

Exemplos de Análise de Riscos à Validade

5. PLANNING (detailed)

Results Validity:

➤ Internal Validity

- Language: The class lecture notes, assignment instructions, techniques and artifacts are written in English, so the lack of proficiency in English can affect the results of the study.
- Learning: PBR is a more procedurally defined technique than a checklist. If subjects are trained in the use of the more procedural PBR techniques prior to using the less-procedural checklist, there is a danger that they will perform the tasks on the checklist in a more ordered fashion.
- Conformance to the Original study: There are some changes made to the experimental procedures by the replicators before running the study. There are two main issues that must be considered for this threat:
 - The replicators made some adjustments to the training time but are keeping it equal for both techniques and they provide different levels of detail and require different levels of detail and background.
 - The techniques will be applied just after training, without giving the subjects time to mature and assimilate the underlying concepts.
- Process conformance of the subjects: We do not have mechanism to observe the subjects while they are working nor will collect any intermediate artifacts. Thus, we can not be certain that the subject will follow the technique.

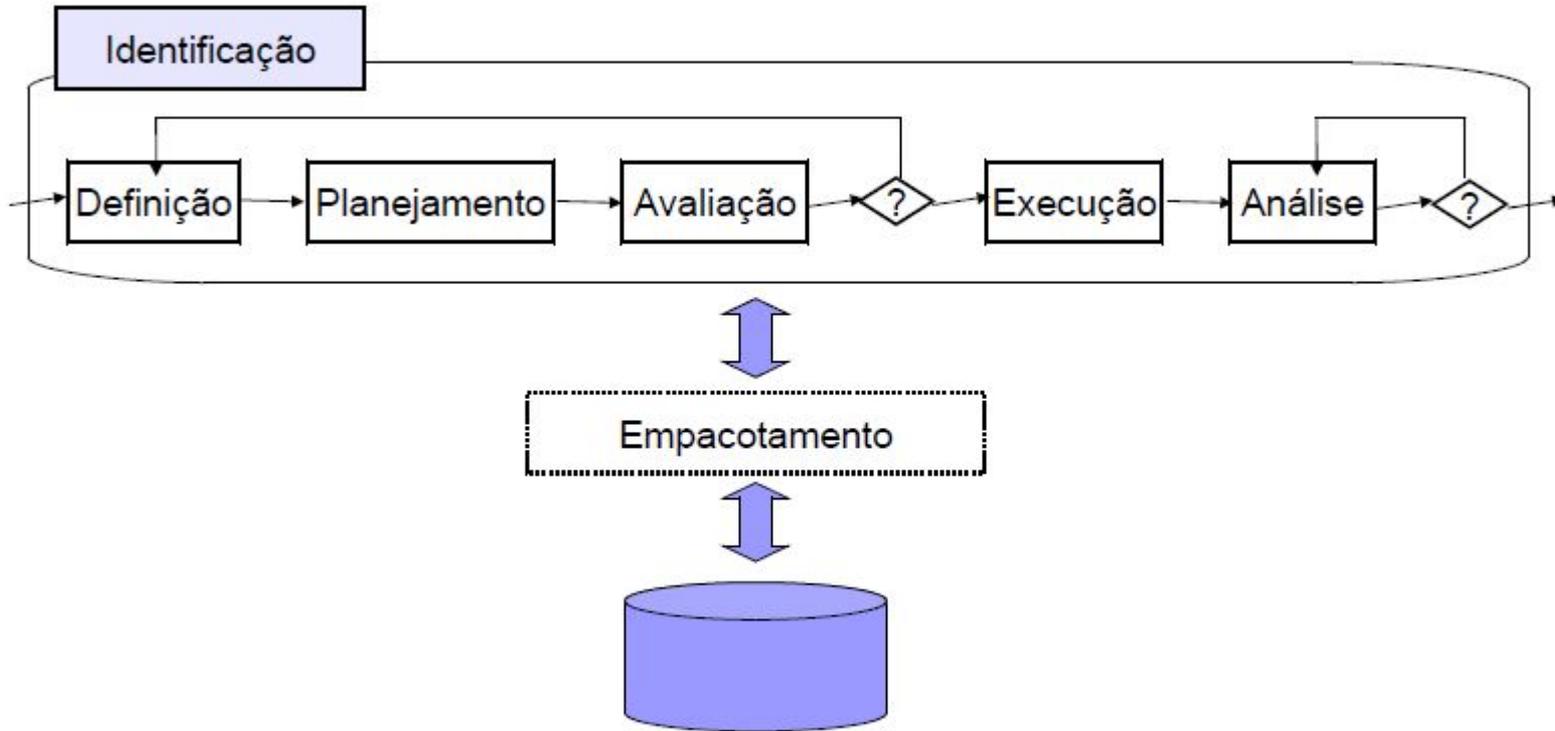
5. PLANNING (detailed)

Results Validity:

➤ **External Validity**

- This study will run in the classroom at a university and the subjects are not as experienced as industrial professionals. The subjects are not experienced in PBR perspectives. Thus, the conclusions of this study may not be directly transferable to industrial inspectors.
- The small number of subjects who will participate in the replication. It is possible that any result of the study be a function of this small sample size

Processo de Experimentação



Avaliação do Planejamento do Experimento

- A avaliação tem o objetivo de verificar se o projeto experimental é consistente e pode realmente ser executado
- Esta fase deve avaliar o projeto experimental e seus artefatos
- Os pontos mais críticos a serem avaliados em um experimento são o design e o processo subjacente:
 - Tempo destinado a cada atividade
 - Treinamento nas técnicas e artefatos a serem utilizados

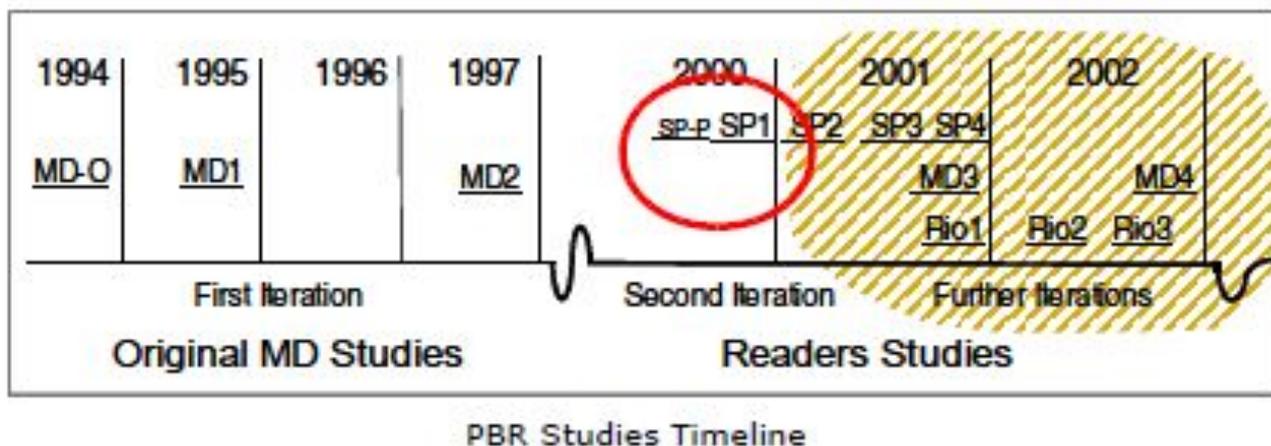
Avaliação do Planejamento

- **Consulte especialistas**
 - Consulte especialistas no domínio para avaliar seus artefatos experimentais
 - Consulte especialistas em ES para avaliar objetivos e projeto experimental
 - Consulte especialistas em ESE e estatísticos para avaliar formato experimental e planos de coleta e análise de dados
- **Se possível, execute um estudo-piloto**
 - Execute uma versão simplificada do experimento para avaliar seu projeto e artefatos.
- **Redesenhe o experimento e melhore os artefatos conforme necessário**

ESTUDOS-PILOTO

- Versões simplificadas do estudo experimental em dimensões tais como:
 - Número de participantes;
 - Volume de tarefas a serem executadas;
 - Tamanho dos artefatos utilizados.
- Resultados do estudo piloto não devem ser usados na análise de dados e teste de hipóteses.
- O estudo piloto deve ser usado para avaliar a razoabilidade do projeto experimental, seus artefatos e processos.

Exemplo: Avaliação e Melhoria do Projeto Experimental



- Um estudo piloto foi realizado para entender as tarefas associadas ao estudo experimental PBR
- Este estudo ajudou a:
 - Determinar os tempos associados às tarefas listadas no projeto
 - Avaliar formulários e questionários
 - Refinar e avaliar o material de treinamento

6. Training

- Definition and procedure:
 - The training will be done in two 2-hour sessions using another artifact, ABC Video Store.
 - The sessions will consist of 30 minutes of theoretical presentation and 90 minutes of practice with the techniques.
 - At the end of the training, the researchers will give the subjects feedback on their performance and the full list of defects for the ABC Video Store document.
 - This feedback will allow the subjects to see the types of defects that they would not uncover and use this information in future applications of the technique.

- Instructor: Emerson Sillas Dória

- Participants: 18 undergraduate students

- Artifacts:
 - ABC Video Store Requirements Document
 - Defect Form
 - Defect List

7. EXECUTION PROCEDURE

➤ Definition:

- The subjects will apply the Checklist and PBR techniques in sessions of 1 hour and 45 minutes each.
- First Day:
 - after receiving training in the Checklist method, the subjects from Group 1 will review the ATM document and subjects from Group 2 will review the PG document.
 - Each subject will be assigned to one of three subgroups for PBR (one for each perspective).
- Second Day:
 - after receiving training in the assigned PBR perspective, the subjects reviewed the other requirements document.
- The subjects will perform the inspections in a classroom while the experimenters will be present.
- During the inspection of the documents the subjects should record any defects they find along with a classification for the defect.

➤ Instructor: Emerson Sillas Dória

7. EXECUTION PROCEDURE

➤ First Day Detailed Definition: Example

