Fraunhofer Institut
Experimentelles
Software Engineering

Experience Package from the ESSI Process Improvement Experiment HYPER

Authors:
Bernd Freimut
Brigitte Klein
Oliver Laitenberger
Günther Ruhe

Partially supported by
ESSI Process Improvement Experiment (PIE)
project n. 27839 HYPER

IESE-Report No. 015.00/E
Version 1.0
February 2000

A publication by Fraunhofer IESE

Fraunhofer IESE is an institute of the Fraunhofer Gesellschaft.
The institute transfers innovative software development techniques, methods and tools into industrial practice, assists companies in building software competencies customized to their needs, and helps them to establish a competetive market position.

Fraunhofer IESE is directed by
Prof. Dr. Dieter Rombach
Sauerwiesen 6
D-67661 Kaiserslautern

Abstract

The development of high quality software satisfying cost, schedule, and resource requirements is an essential prerequisite for improved competitiveness of life insurance companies. One major difficulty to master this challenge is the inevitability of defects in software products. Since defects are known to be significantly more expensive if detected in later development phases or testing, companies in this marketplace must use cost-effective technologies to detect defects early on in the development process. A particular promising one is software inspection.

This report describes the results of the ESPRIT/ESSI Process Improvement Experiment "High Quality of Software Products by Early Use of Innovative Reading Techniques (HYPER)". The core of this project has been the transfer of innovative software inspection technologies to the Allianz EURO conversion projects. The innovation in the area of software inspection is based on a systematic reading technique, that is, Perspective-based reading (PBR), that tells inspection participants what to look for and - more important - how to scrutinize a software artifact for defects.

The report packages the experience regarding the application of PBR inspections on requirements and design documents in the ESSI PIE. The experience consists of results from the measurement program and lessons learned both with respect to inspections and their accompanying measurement program.

Keywords: Perspective-based Inspection, Systematic Quality Improvement, GQM-based measurement program

Copyright © Fraunhofer IESE 2000

Table of Contents

Copyright © Fraunhofer IESE 2000

Copyright © Fraunhofer IESE 2000

Copyright © Fraunhofer IESE 2000

1 Introduction

Development of high quality software satisfying cost, schedule, and resource requirements is a challenging task for all software producing units in the insurance business. Competition will be more and more decided by those organizations which have the best quality software products, and which are able to react flexible on new market requirements.

Since the quality of delivered products is strongly determined by the quality of the employed software development processes, the demand for high quality products calls for continuous and systematic quality improvement of the software development processes. Therefore, Allianz Life, one of Germanys largest insurance companies, has been performing a large-scale software process improvement program since 1993.

The ingredients of the software process improvement program are the following three activities:

1.) Identifying and understanding the weaknesses of the current development processes by means of a measurement program.
2.) Selecting and implementing appropriate techniques that are to overcome the detected weaknesses.
3.) Evaluating the impact of the selected techniques by means of a measurement program.

To identify and understand the weaknesses of Allianz development processes, initial measurement programs based on the Goal-Question-Metric paradigm [3][10] were performed. These measurement programs exhibited the following results:

- The importance of testing for quality assurance is overestimated (there is too much effort spent on testing).
- Communication and common understanding among all departments participating in a project is currently a weakness and has to be improved.
- Too many defects are injected in the early stages (requirements analysis, design) and are detected in testing.
- Effectiveness and efficiency of verification (inspections) is currently low and will be improved by application of perspective-based reading techniques.
- Overall effort for testing (currently 30% of development effort) has to be reduced by early use of verification.

Copyright © Fraunhofer IESE 2000

In order to tackle these weaknesses of the development process, Allianz Life decided to investigate Perspective-based Inspections performed in the early phases of the software life cycle.

Consequently, Perspective-based Inspections were implemented in the Euro-Conversion project, a development project of strategic importance for Allianz Life, and their impact on the development process were evaluated by means of a measurement program based on the GQM paradigm. This overall evaluation was performed in the framework of the ESSI Process Improvement Experiment "High Quality of Software Products by Early Use of Innovative Reading Techniques (HYPER)".

This report summarizes the results of this Process Improvement Experiment (PIE). The report's main purpose is to highlight the overall project results and form the background for further dissemination and exploitation of results.

The report is structured as follows: Section 2 summarizes Perspective-based Inspections and their overall impact on the development process as determined in the PIE. Section 3 presents the results from the measurement program. Section 4 presents lessons learned with respect to the application of Perspective-based Inspections and the accompanying measurement program.

Copyright © Fraunhofer IESE 2000

2 Executive Summary: Experience using Perspective-based Inspections at Allianz

2.1 A Brief Description of Perspective-based Inspections

Software Inspection is an industry-proven best practice for software quality assurance. It consists of the activities planning, defect detection, defect collection, and defect correction. As depicted in Figure 1, the planning activity is performed by the organizer who is responsible for setting up an inspection for a particular software artifact. Throughout the defect detection step inspectors individually scrutinize a software artifact for potential defects using the Perspective-based reading technique. Other documentation, such as company-specific guidelines, may support this activity. Inspectors document all potential defects they find on a defect report form. As some of the potential defects documented on the defect report forms might prove not to be real defects, inspectors together with the author and a moderator perform an inspection meeting. The goal of the inspection meeting is to decide upon which of the potential defects are real ones. In addition, new defects might be detected during the inspection meeting. Throughout the meeting, one of its participants documents all real defects on a meeting report form. In the final activity, the author corrects the real defects.

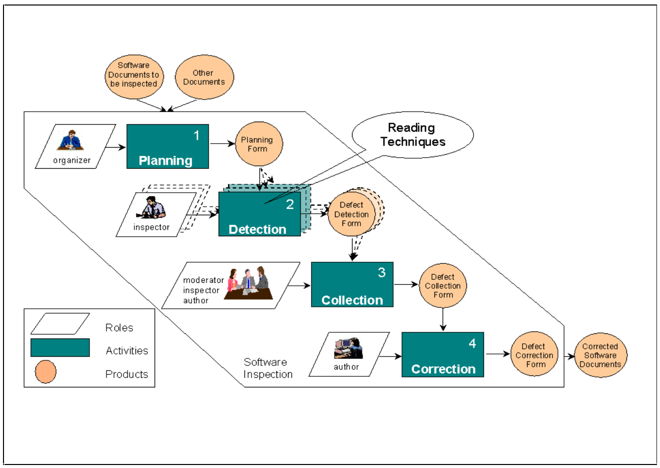


Figure 1 The inspection process

During the defect detection step, the inspectors use the Perspective-based reading technique that tells inspection participants what to look for and more important how to scrutinize a software artifact for defects.

Copyright © Fraunhofer IESE 2000

The basic goal of PBR is to examine the various descriptions of a software artifact from the perspectives of the artifact's various stakeholders for the purpose of identifying defects. Each software artifact is inspected from the perspective of each stakeholder involved in the software lifecycle in such a way as to determine if the descriptions satisfy the stakeholders' particular needs.

An inspector in a perspective-based inspection reads the inspected descriptions from the perspective of a particular stakeholder. In doing so, the inspector follows a perspective-based reading scenario (in short: scenario). A scenario tells an inspector how to go about reading an artifact from one particular perspective and what to look for.

As shown in Figure 2, the scenario consists of an introduction, instructions, and questions framed together in a procedural manner. The introductory part describes the stakeholder's interest in the artifact and explains the quality factors most relevant for this perspective. The instruction part describes what kind of descriptions an inspector is to use, how to read the descriptions, and how to extract the appropriate information from them.

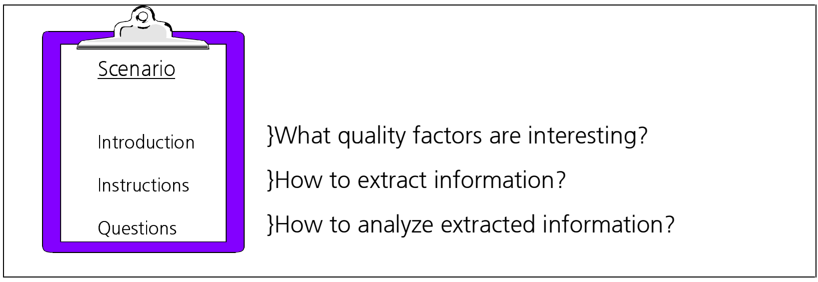


Figure 2 A Perspective-based reading scenario

Copyright © Fraunhofer IESE 2000

2.2 Impacts and Objectives of PBR Inspections

In the framework of Allianz' Euro-Conversion projects, inspections have been performed in two important sub-projects. In the remainder of this report these sub-projects are referred to as Project A and Project B.

Project A had an effort of 38 person months, 28 person months of which were effort from the IT department. It comprised two stages and had the task to adapt software systems to the EURO currency for the Allianz Investment Trust (KAG), a subsidiary of the Allianz Group. The project team consisted of 4 project members each from the IT and investment departments. Project B had an effort of 33 person months, 22 person months of which were effort from the IT department. It also comprised two stages to convert the amounts of insurance policies to the EURO. The project team consisted of 6 project members from IT departments and 11 project members from the insurance departments.

The overall results of the PIE in these two projects can be summarized as follows:

- Inspections in early phases find defects that would have been detected in much later (testing-) phases without inspections. Therefore, defects are detected earlier in the life cycle.
- In the two considered projects the test effort was lower compared to a baseline derived from two former projects similar in size to Project A and B. The effort for unit test and integration test, and introduction (equivalent to acceptance testing) accounted for 23.7% resp. 29% of the overall IT development effort compared to a baseline of between 32% and 47%. Although different factors contribute to this lower testing effort (e.g., a different testing strategy), the developers considered inspections to be a major factor.
- Performing inspections for user-output descriptions (screen definitions, letters to be sent to Allianz' insurance clients) contributes significantly to the clarity and customer-friendliness of the resulting products.
- Including the future (Allianz-internal) users of the system as inspectors in analysis inspections leads to a system appropriate for the future users. This contributes to a higher user satisfaction of Allianz' software systems.
- Design Inspections were useful for detecting performance and reliability problems early-on.

Copyright © Fraunhofer IESE 2000

2.3 Cost-Benefit of PBR Inspections

Effort Savings

In total, both projects spent 96 person days on performing inspections. In contrast to this cost, the project members estimated the effort savings from these inspections in later phases at 192 person days. Thus, the return-on-investment equals 2 and is therefore clearly larger than 1.

Project	Cost	Estimated Savings
Project A	52 pd	89 pd
Project B	44 pd	102 pd
Total	96 pd	192 pd

Table 1 Cost-Benefit of PBR inspections.

In addition to this economical impact of inspections, indirect and qualitative benefits could be observed as well.

User Satisfaction

The focus of Project A's inspections was on user-output descriptions such as letters to be send to customers and screen definitions to be targeted to people working in call-centers. Thus, scenarios for the corresponding target groups (Allianz' investment and legal experts as well as Allianz' call-center personnel) were developed. Additionally, these scenarios were used by inspectors being employees in the call-centers, who also know about the requirements of Allianz' investment clients. As a result of these two measures, many defects regarding the user-friendliness could be detected. This led to the definition of a more appropriate system contributing to Allianz' business objective "Better customer satisfaction of delivered products" that also motivated the introduction of inspections.

Education

The emphasis of Project B was to design and implement crucial and complex requirements for which many different aspects such as financial, actuarial, organizational, and implementation related issues had to be taken into account. The involvement of experts in the respective domains as inspectors contributed to the learning of the developers since the document authors could gain insight into the various domains. As a result, the document authors considered this additional knowledge as valuable for future development activities.

Copyright © Fraunhofer IESE 2000

2.4 Prerequisites for the Implementation of PBR Inspections

In order to perform inspections successfully, the following prerequisites have to be fulfilled:

- The inspection process has been defined.
- The inspection process is embedded in the overall software development process.
- Persons responsible for the performance of the inspection process have been identified.
- Cost and benefit of inspections are made explicit by means of accompanying measurement programs.
- The inspection participants have been trained both with respect to the inspection approach and the accompanying measurement program.

2.5 Auxiliary Means for PBR Inspections and their Measurement

In order to satisfy the prerequisites listed in the previous section, the following training material has been prepared for usage within Allianz' Euro-Conversion projects and the HYPER project:

- A training course on Perspective-based Inspections [7].
- A handbook on Perspective-based Inspections [8].
- A training course on Goal-oriented Measurement using GQM [5].
- A GQM measurement plan for characterizing the Perspective-based inspection approach and its impact on the development process [4].

Copyright © Fraunhofer IESE 2000

3 Measurement Results

3.1 Introduction

In order to evaluate the introduction of Perspective-based (PBR) inspections in the context of the Process Improvement Experiment HYPER, the inspection process and its impact on the overall development process is investigated in a qualitative and quantitative manner. The basis of this investigation is a measurement program following the GQM paradigm [3][10].

In this section, the main results of the measurement program as defined in [4] are presented. The organization is as follows. Section 3.2 presents measurement results characterizing the effectiveness and efficiency of PBR inspections. Section 3.3 characterizes the defects slipping through inspections, Section 3.4 characterizes the development effort of the entire project, Section 3.5 characterizes the cost-benefit relationship of inspections, Section 3.6 investigates the impact of the inspectors' experience on the effectiveness and efficiency.

In each section the relevant measurement results according to the GQM plan are presented. For each question, the object of measurement, its quality aspect and a description of the quality description are given. Next, the measurement results for both projects are presented along with the interpretation of the results as given by the project members.

Copyright © Fraunhofer IESE 2000

3.2 Characterization of PBR Inspections wrt. effectiveness and efficiency

3.2.1 What is the effectiveness of PBR inspections, where effectiveness is defined as the ratio between number of detected defects (of severity critical and very critical) and size of the document?

Description

In order to determine the capability of inspections to detect defects, the measure 'defect density' can be used. This measure is defined as the number of defects detected per page. In order to compare the effectiveness of different inspections or inspection processes, it has to be assumed that all documents have the same actual defect density. Therefore, in comparing different inspections or inspection processes the different document types have to be taken into account, since different document type can vary in their actual defect density.

Object

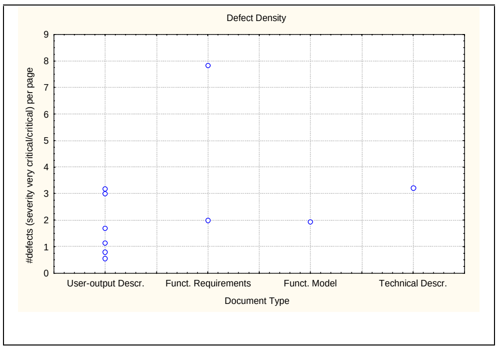
PBR inspections for analysis and high-level design inspections

Quality Aspect

Effectiveness

Situation

The defect density for document types belonging to analysis and high-level design inspections is shown below. On average 1.73 defects per page are detected in user-output descriptions, 4.92 defects per page in functional requirements, 1.93 defects per page in the functional model, and 3.22 defects per page in technical descriptions.



Interpretation

In Project A user-output descriptions were inspected. For this document type, newly developed documents showed – as it can be expected – the highest defect density. In particular, the first two inspections yielded a defect density much higher (3 and 3.2 defects per page) than the following inspections. This is due to the fact that the developers considered the results of the first two inspections when developing the documents that were inspected later on. Thus, a side effect of inspections is that developers become aware of possible defects and can prevent them in the future.

Copyright © Fraunhofer IESE 2000

In Project B, in which the remaining document types were inspected, one inspection yielded a very high defect density of 7.83 defects per page. This inspection can be seen as exceptional, as the inspectors found many defects that the developers were aware of but that had not to be addressed in the first iteration of the project.

Copyright © Fraunhofer IESE 2000

3.2.2 What is the effectiveness of PBR inspections where effectiveness is defined as the percentage of defects (of severity critical and very critical) found?

Description

In order to determine the capability of inspections to detect defects, the percentage of defects detected by inspections can be used as metric. Compared to other measures such as the number of defects or the defect density, this measure can be used to compare inspections across different document types. Ideally, 100% of all present defects should be detected. Depending on the effectiveness of the inspections and the detectability of defects, this percentage can also be lower.

Object

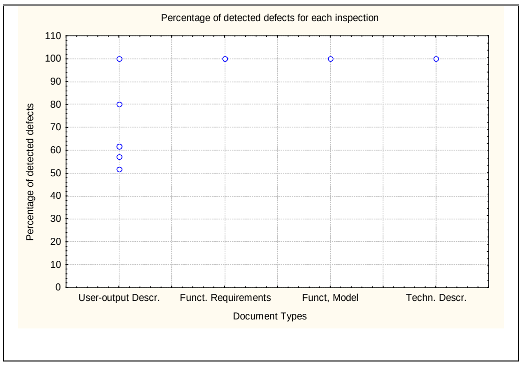
PBR inspections for analysis and high-level design inspections

Quality Aspect:

Effectiveness

Situation

In Project A on average 69% of all defects are detected in user-output descriptions, whereas in Project B no defect was traced back to the inspected documents, thus yielding a value of 100%.



Interpretation

In Project A, in which user-output descriptions were inspected, 69% of all analysis defects in the inspected document were already detected at the end of the analysis phase by inspections. Thus, the developers of this project considered inspections to prevent many defects from slipping through into later development and testing activities.

In Project B the remaining document types were inspected. Here 100% of all analysis defects were detected by inspections. However, this figure has to be interpreted with care. In this project the analysis documents dealt more with strategic aspects of the Euro-Conversion whereas details were refined during design. Defects from later (testing-) phases were traced back to the detailed documentation being the low-level design. Thus, no defect was traced back to the analysis or high-level design phase.

Copyright © Fraunhofer IESE 2000

3.2.3 What is the distribution of findings, questions, improvement proposals, and comments?

Description

Besides defects (*findings*, i.e., issues that have to be corrected[1]), inspections can also uncover issues that have to be investigated further in coordination with other departments or projects (*questions*), proposals to improve the development process (*improvement proposals*), and comments capturing important explanations regarding the inspected document (*comments*).

Object

PBR inspections

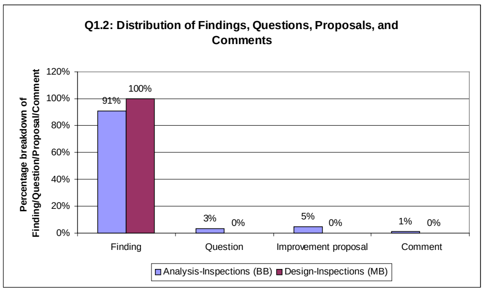
Quality Aspect

Detected issues

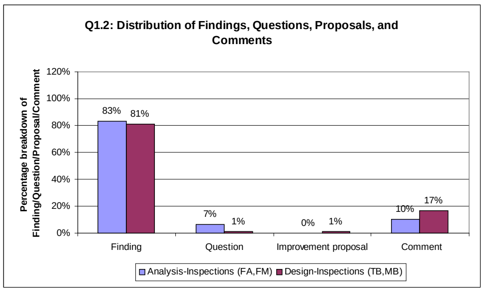
Situation

In both projects mostly defects (*findings*) are recorded as shown below.

Project A



Project B



Interpretation

Inspections can uncover different kinds of issues, such as questions, improvement proposals, comments and defects. The emphasis is, however, to detect defects.

[1] see also Section 4.1.2 for an explanation of the term 'finding'

Copyright © Fraunhofer IESE 2000

3.2.4 What is the distribution of findings, questions, proposals, comments with respect to the inspected product, other products, process, other projects?

Description

During an inspection, a defect/finding can be detected in the inspected document (*inspected document*) or it can be discovered that another document in the project has to be changed or corrected (*other product*). Moreover, it can be decided, that coordination with a parallel project has to be performed (*other project*). Finally, a finding can refer to the process where either the development process or the process of applying the developed system is meant (*process*).

Object

PBR inspections

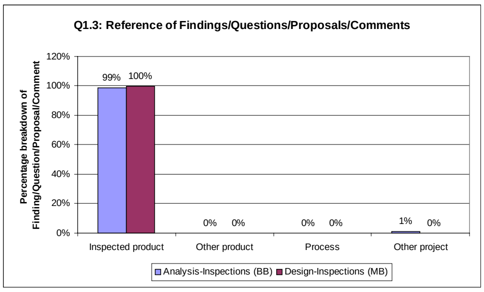
Quality Aspect

Reference of detected findings, questions, proposals, comments

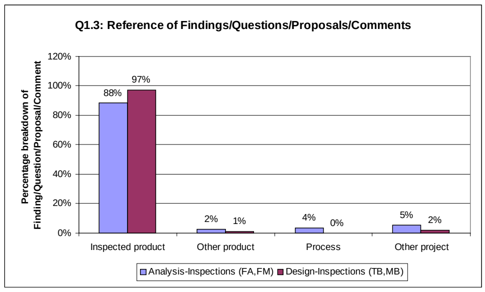
Situation

In both projects the findings refer mostly to the inspected product as shown below.

Project A



Project B



Interpretation

Although the defects/findings detected in inspections, can refer to several sources, a clear emphasis is on the inspected document. This result demonstrates that the focus is on checking and improving the document as recommended in the literature.

Copyright © Fraunhofer IESE 2000

3.2.5 What is the severity of the detected defects?

Description

Inspections can detect defects that differ in their impact on the customer, other systems, or the system under inspection. Depending on the severity of this impact, defects can be classified as *very critical*, *critical*, or *interesting*.

Object

PBR inspections

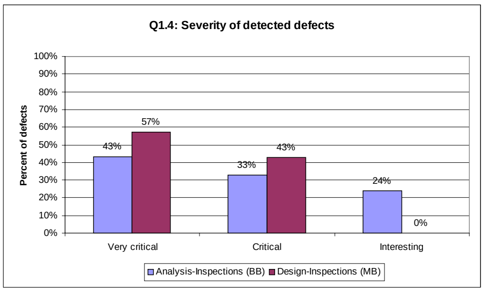
Quality Aspect

Severity of detected defects

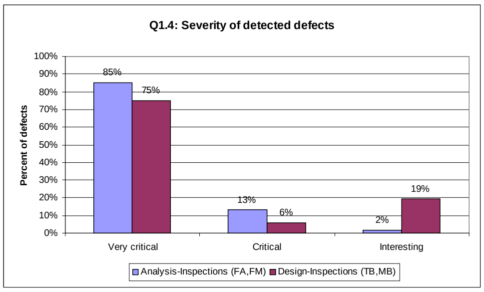
Situation

In both projects mainly defects of severity *very critical* and *critical* were detected. A difference can be observed between Project A and Project B in the sense that in Project B a high proportion of defects are of severity *very critical*.

Project A



Project B



Interpretation

The inspected products in Project A were user-output descriptions, such as screen definitions and letters to be sent to Allianz' insurance clients. Since user-friendliness was the focus of the inspection, defects with an impact on the user or customer were classified *very critical*, whereas other important defects were classified as *critical*. Due to the subjective nature of user-friendliness, also nice-to-have defects that were not critical to the system were detected and classified as *interesting*.

The inspected products of Project B were documents defining (parts of) the functionality for the entire system. Thus, most defects were *very critical* as the defects would have affected system functionality significantly.

Due to the different nature of the defects and classifications as described above, the graphs shown above are difficult to compare. For a reasonable comparison, the classes *very critical* and *critical* have to be collapsed into one single class. The interpretation of this result then shows that the detected defects have a significant impact on the customer or the developed system indicating the quality improvement achieved by inspections.

Copyright © Fraunhofer IESE 2000

3.2.6 What is the distribution of detected defects (of severity *critical* and *very critical*) broken down by defect type?

Description

The defects detected in inspections can be due to the fact that the inspected document is inconsistent in itself (*inconsistent*), that the requirements or design is ambiguous and allows different interpretations (*ambiguous*), that the document is not concrete enough to allow an implementation (*not concrete*), that an important aspect has been omitted (*forgotten*), or that a mistake has been made (*wrong*). Additionally, inspectors can raise new aspects that should be put into the requirements or design (*new*) or can provide different ideas for implementation (*alternative*). Finally, unimportant aspects leading to gold-plating can be identified and removed (*extraneous*).

Object

PBR inspections

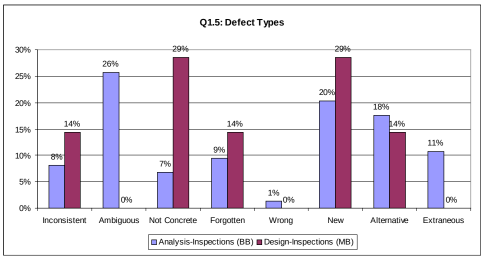
Quality Aspect

Type of detected defects

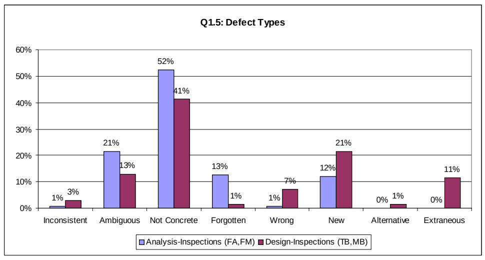
Situation

In Project A the majority of defects detected in analysis are of type *ambiguous*, *new*, and *alternative*. In Project B the majority of defects are of *type not concrete* and *ambiguous*.

Project A[2]



Project B



Interpretation

For Project A the distribution is explained by the type of the inspected document (user-output descriptions such as screen definitions and letters to be sent to Allianz' insurance clients). The focus on user-friendliness in these inspections uncovered many aspects that were not clear enough (type *ambiguous*). Alternative solutions were often provided to increase user-friendliness explaining the high proportion of defects of type *alternative*. The proportion of defects of type *new* reflects the fact that due to the diverse selection of inspection participants many different aspects of the inspected products could be covered.

In Project B the high proportion of defects of type *not concrete* is explained by the nature of the project. In this project, the system was developed in two it-

[2] Only a very small number of design defects were detected. Therefore, an interpretation of their results is not reasonable here.

Copyright © Fraunhofer IESE 2000

erations. During the inspections in the first iteration cycle, many defects arose that were not subject of the first iteration but of the second. Thus, the developers had a benefit after the inspections in the first cycle for the development activities in the second cycle as weaknesses were already identified before the second cycle started. The proportion of *ambiguous* defects could be partly explained by the fact that different people can have different interpretations of a phrasing. Thus, inspections contributed to clearer and more precise requirements.

Copyright © Fraunhofer IESE 2000

3.2.7 What would have been the impact of the detected defects (of severity *critical* and *very critical*) if they had not been detected?

Description

Defects detected in inspections can have different impacts on the customer if they had not been found in inspections but had been detected in the final system. A standard classification scheme to capture the impact of defects is given in [6]. This scheme defines the impact of defects on the customer as *user-friendliness*, *documentation*, *functionality*, *integrity*, *performance*, *standards*, *maintainability*, or *reliability.*

Object

PBR inspections

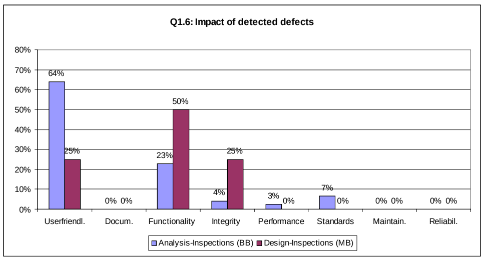
Quality Aspect

Impact of detected defects

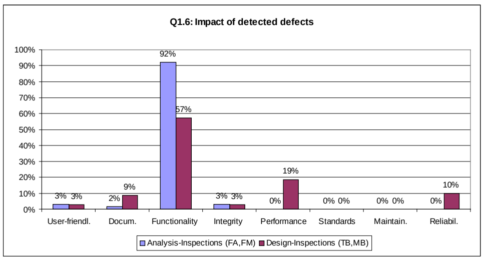
Situation

In Project A, the majority of defects impact the *user-friendliness* of the system, whereas in Project B the majority of defects have an impact on the *functionality* of the system. In Project B, design inspections additionally uncover *performance* and *reliability* defects.

Project A[3]



Project B



Interpretation

These results are explained by the type of inspected products. In Project A user-output descriptions are inspected. The focus of these inspections is on user-friendliness explaining naturally the high proportion of *user-friendliness* defects. The focus of Project B was the definition and implementation of (a part of) the functionality for the entire system. Thus, the inspected documents types (functional descriptions, functional model, module descriptions) explain the high proportion of *functionality* defects. In High-Level Design inspections, additionally defects were detected impacting *performance* and *reliability* issues.

Thus, depending on the type of inspected document, inspections contributed to a more user-friendly system that implemented the correct functionality. This will result in a higher acceptance of the system from the future users.

---

[3] Only a very small number of design defects were detected. Therefore, an interpretation of the results is not reasonable here.

Copyright © Fraunhofer IESE 2000

3.2.8 What is the efficiency (number of defects detected divided by the detection effort) of PBR inspections?

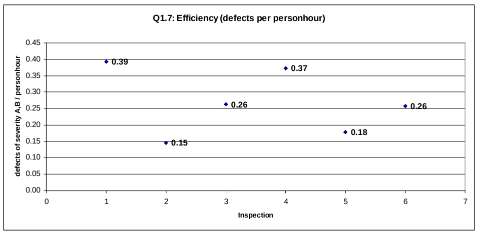
Description Inspection efficiency characterizes the costs associated with finding defects by inspections. A common measure for inspection efficiency is the number of detected defects divided by the detection effort

Object PBR inspections for analysis and high-level design documents

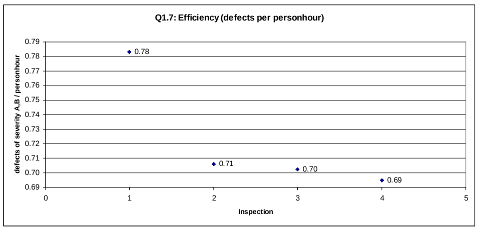
Quality Aspect Efficiency

Situation In Project A on average 0.27 defects (of severity *very critical* or *critical*) per person hour were detected in user-output descriptions whereas in Project B on average 0.72 defects (of severity *very critical* or *critical*) per person hour were detected.

Project A



Project B



Interpretation A direct interpretation of the efficiency values in the two projects was not performed. However, since efficiency combines both the effort required for inspections as well their effectiveness (here: the number of defects), the interpretations and improvement proposals for the quality aspects effort (Section 3.5.4) and effectiveness (Section 3.2.1 and 3.2.2) apply here as well.

Copyright © Fraunhofer IESE 2000

3.3 Characterization of the defect slippage

3.3.1 What is the number of defects broken down by activity where defect was detected and where it was injected?

Description

Inspections in early phases can uncover defects that otherwise would have been detected in much later (testing) phases at higher costs. Thus, defects can be detected much earlier in the life cycle.

Object

Overall Verification and Validation Approach

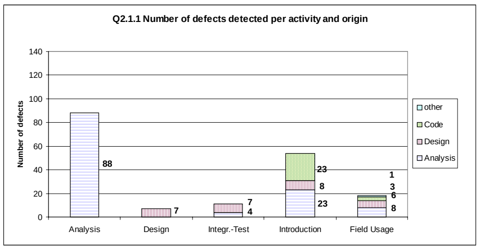
Quality Aspect

Time of defect detection

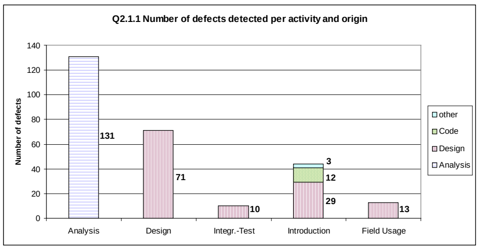
Situation

In Project A 95 defects (i.e., 53%) where detected in early analysis and design inspections, whereas in Project B 202 defects (i.e., 75%) defects where detected in early inspections as shown below. (Note: In integration test, only defects of origin analysis and design detected in the second iteration of the project were recorded and thus considered in this analysis).

Project A



Project B



Interpretation

In this figure the number of defects found in analysis and design inspections contains only defects of severity *very critical* and *critical*, which could have resulted in a test defect. Therefore, due to inspections defects are detected much earlier in the life cycle than with testing alone.

In Project A most of the defects of origin analysis were indeed detected in analysis inspections (72%). Thus, inspections helped to assure the quality of the requirements specification early on, which resulted in fewer defects and thus less rework effort in the testing phase.

In Project B the inspected analysis documents provided the concept for the implementation from the viewpoint of the application domain. The design documents contained the technical aspects of the system. In later testing phases defects were only tracked back to this technical concept. Therefore, no defects of origin analysis are detected in later phases.

Copyright © Fraunhofer IESE 2000

3.3.2 What is the ratio of the number of defects detected during analysis and design inspections and the total number of defects with origin in analysis and design documents?

Description

The effectiveness of the overall verification and validation approach can be characterized by determining for each defect detection activity the percentage of defects that are detected. Ideally, 100% of all defects created in one phase (e.g., analysis or design) should be detected in the corresponding detection activity (analysis inspection, design inspection). Depending on the selection of documents to be inspected, the effectiveness of the individual inspections, and the detectability of defects, this percentage can be lower than 100%.

Object

Overall Verification and Validation Approach

Quality Aspect

Defect Slippage

Situation

In both projects, a high percentage of defects with origin analysis are detected in analysis inspections as shown below. In Project A the percentage of design defects detected in design inspections is comparably low at 25%.

Project A		Project B	
Analysis-Phase	Design-Phase	Analysis-Phase	Design-Phase
72%	25%	100%	57.2%

Interpretation

In Project A most of the defects of origin analysis were indeed detected in analysis inspections. Thus, inspections helped to assure the quality of the requirements specification early on, which resulted in fewer defects and thus rework effort in the testing phase. The comparably low effectiveness value of 25% for design inspections was explained by the time pressure of the project. Due to the time pressure, design, implementation, and test activities had to be performed in parallel. Thus, testing was often given higher priority compared to inspections, resulting in many slipped design defects. The developers felt, that in future projects more effort should be scheduled for and invested in to check the quality of the design as early as possible in the life cycle.

In Project B the inspected analysis documents provided the concept for the implementation from the viewpoint of the application domain. The design documents contained the technical aspects of the system. In later testing phases defects were only tracked back to this technical concept. Therefore, no defects of origin analysis are detected in later phases, explaining the value of 100% for analysis inspections. The value of 57.2% for design inspections was explained by the implementation strategy of the development team. Due to the complexity of the implemented requirements, the developers first developed and implemented a strategy suitable for general and normal aspects of the re-

Copyright © Fraunhofer IESE 2000

quirements. The objective of the inspection was to find defects in these general and normal aspects of the requirements. Specific aspects of the requirements (special cases) were not the objective of the inspections. Defects in these requirements were intentionally targeted to be found in testing.

Improvement

The developers of Project A proposed to invest in future projects more time in design inspections.

Copyright © Fraunhofer IESE 2000

3.3.3 What is the severity of analysis/design defects that slipped through analysis and design inspections (per detection activity)?

Description

Defects in inspected documents can differ in the severity of their impact on the customer, other systems, or the system under inspection. Depending on this impact, defects can be classified as *very critical*, *critical*, or *interesting*. Similar to the severity of detected defects (see Section 3.2.5) it is also possible to classify the severity of defects that were not detected by inspections but in later (testing) activities.

Object

PBR inspections

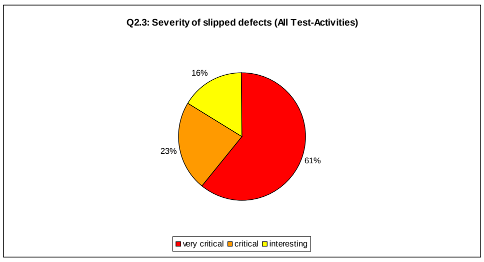
Quality Aspect

Severity of defects that were not detected by inspections

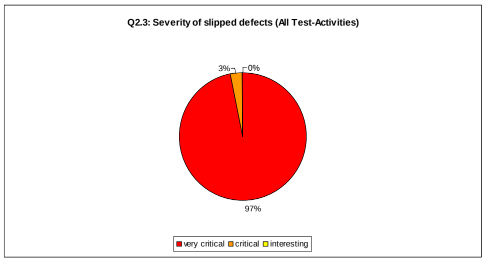
Situation

In both projects a high percentage (84% resp. 97%) of defects slipping through analysis ands design inspections are of severity *very critical* or *critical* as shown below.

Project A



Project B



Interpretation

In Project A the emphasis was on user-output descriptions such as screen definitions and letters to be sent to Allianz' insurance clients. Since user-friendliness was the focus of the project, defects with an impact on the user or customer were classified *very critical*, whereas other important defects were classified as *critical*. All these defects had to be corrected immediately.

In Project B all defects that had to be corrected immediately were classified as *very critical*.

Thus, the majority of defects that slip through inspections have a high severity. This, however, does not mean, that the inspections were ineffective as high severity defects slipped. Rather, its is the nature of defects detected in testing that they have to be corrected immediately.

Copyright © Fraunhofer IESE 2000

3.3.4 What is the impact of those defects slipping through analysis and design inspections?

Description

Defects detected in defect detection activities can have a different impact on the customer if they had not been found and had been detected in the final system. A standard classification scheme to capture the impact of defects is given in [6]. This scheme defines the impact of defects on the customer as *user-friendliness*, *documentation*, *functionality*, *integrity*, *performance*, *standards*, *maintainability*, or *reliability.* Similar to the impact of detected defects (see Section 3.2.7) it is also possible to classify the impact of defects that were not detected by inspections but in later (testing) activities.

Object

PBR inspections

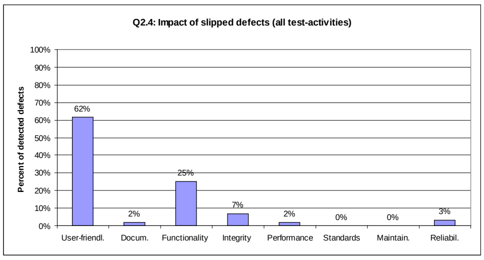
Quality Aspect

Impact of defects that were not detected by PBR inspections

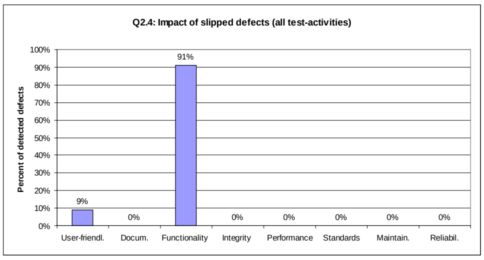
Situation

In Project A, the majority of defects slipping through analysis ands design inspections impact the *user-friendliness* of the system, whereas in Project B the majority of defects have an impact on the *functionality* of the system as shown below.

Project A



Project B



Interpretation

The results can be explained by the nature of the projects. In Project A the focus was on user-friendliness in user-output descriptions (i.e., letters to be sent to Allianz' insurance clients, screen definitions targeted to people working in call-centers) explaining naturally the high proportion of *user-friendliness* defects. The focus of Project B was the definition and implementation of (a part of) the functionality for the entire system. This fact explains the high proportion of *functionality* defects.

In Project A the proportion of defects of impact *functionality* could be traced back to stakeholders not represented as perspectives in the inspection whereas in Project B the proportion of defects of impact *functionality* could be traced back to special aspects of the requirements that were not targeted to be addressed in inspections.

Copyright © Fraunhofer IESE 2000

Improvement

In future PBR inspections a careful selection of the stakeholders of the inspected documents has to be performed. In addition to the selection strategy in these projects, stakeholder from interfacing systems have to be considered more carefully.

Copyright © Fraunhofer IESE 2000

3.3.5 What is the type of defects slipping through analysis and design inspections?

Description

The defects in an inspected document can be due to the fact that the inspected document is inconsistent in itself (*inconsistent*), that the requirements or design is ambiguous and allows different interpretations (*ambiguous*), that the document is not concrete enough to allow an implementation (*not concrete*), that an important aspect has been omitted (*forgotten*), or that a mistake has been made (*wrong*). Additionally, new aspects can be defined that should be put into the requirements or design (*new*) or different ideas for implementation can be provided (*alternative*). Finally, unimportant aspects leading to gold-plating can be identified and removed (*extraneous*). Similar to the type of detected defects (see Section 3.2.6) it is also possible to classify the type of defects that were not detected by inspections but in later (testing) activities.

Object

PBR inspections

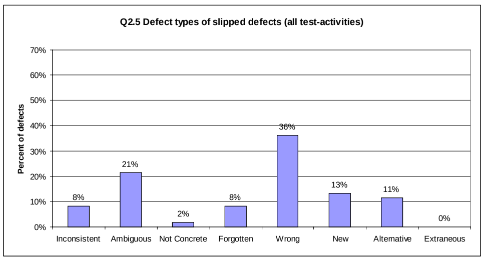
Quality Aspect

Type of defects not detected by inspections.

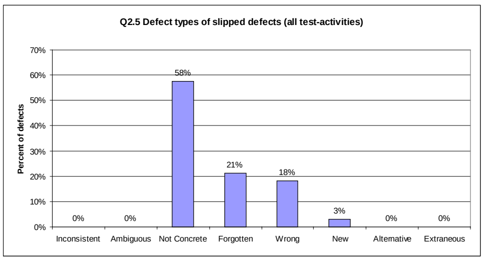
Situation

In Project A the majority of defects slipping through analysis and design inspections is of type *wrong*, *ambiguous*, and *new*. In Project B the defects were of type *not concrete, forgotten,* and *wrong*.

Project A



Project B



Interpretation

In Project A the focus of inspections was on user-interface descriptions. The fact that many *wrong* defects were not detected was interpreted by the fact that the design inspections were afterwards considered as not thorough enough. The high proportion of *ambiguous* defects was explained by the nature of the project with the focus on user-output descriptions. User-output descriptions aim at presenting information to the user in an unambiguous way. Thus, user-output descriptions naturally contain a high proportion of *ambiguous* defects. The proportion of defects of type *new* was explained that an other project imposed new requirements to Project A in the course of the project.

Copyright © Fraunhofer IESE 2000

In Project B the developers contributed the *not concrete*, *forgotten* and *wrong* defects to the development strategy. In this strategy the inspectors focused first on usual and important aspects of the case to be implemented whereas seldom aspects of the case were intentionally left to be detected by testing. These specific aspects caused the *not concrete*, *forgotten* and *wrong* defects.

Improvement

The developers in Project A proposed to perform more thorough design inspections to detect technical aspects earlier.

Copyright © Fraunhofer IESE 2000

3.3.6 How many defects are detected during testing?

Description

In earlier measurement programs at Allianz, the defects were classified according to the phase in which the origin of the defect was. Following this classification scheme, a defect can originate in the analysis or design phase (*change requirements/design*) or in the coding phase (*change program)*. Additionally, the origin of a defect can be in the changing context of the system (*change context*). Finally, a defect reported in the defect tracking system can result in the decision that no change is necessary (*no change*) or that the project is not responsible for the reported defect (*other project responsible*).

In the previous measurement programs, about 50% -60%of defects originating in the analysis or design phase was determined as a baseline. It was one objective to reduce the number of defects from earlier phases. Thus, the proportion of analysis or design defects was expected to decrease in Project A and B compared to this baseline.

Object

Testing process

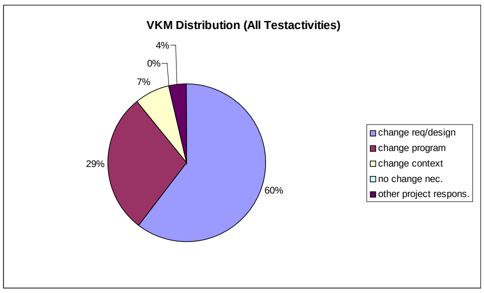
Quality Aspect

Origin of detected defects

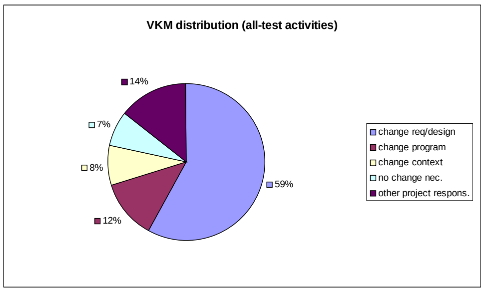
Situation

In both projects about 60% of the defects detected in testing originate from the analysis or design phase.

Project A



Project B



Interpretation

Although the percentage of defects originating in early phases did not decrease in comparison with the historical baseline, the developers did not contribute this result to the ineffectiveness of the inspections. In Project A the developers were rather convinced that without inspections the proportion of defects originating in early phases would have been much larger. Moreover, the testing activities naturally test from the viewpoint of the later users of the system. Therefore, the developers of Project B considered a high percentage of defects originating in analysis and design as natural for the testing process regardless of the presence or absence of inspections.

Copyright © Fraunhofer IESE 2000

3.4 Characterization of the development effort

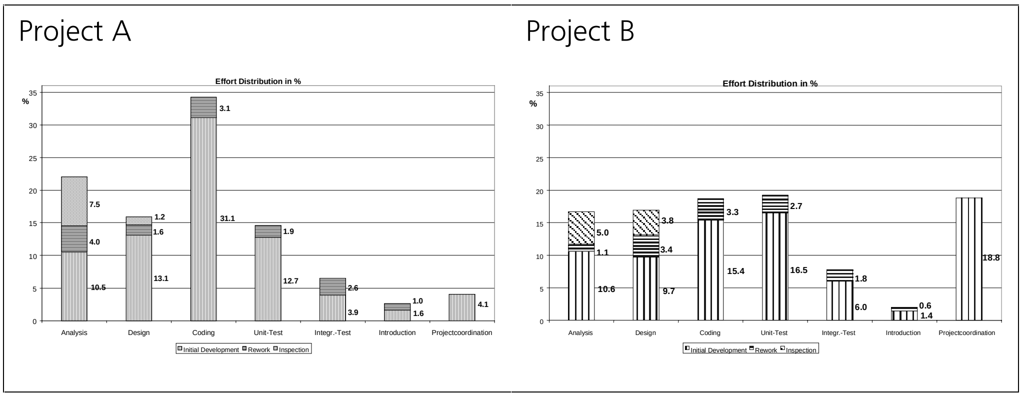
3.4.1 What is the effort distribution broken down by the development activities?

Description
In earlier measurement programs at Allianz, it was determined, that on two projects similar in size to Project A and B the IT Department spent 47% resp. 32% of the total development effort on testing activities (unit test, integration test, introduction/acceptance test). Since inspections find defects at a much earlier point in time than testing does, the testing effort is supposed to decrease since fewer defects have to be isolated and corrected during testing. Thus, the effort for the various development activities is measured and it is distinguished between effort for initial creation and rework.

Object
Development process

Quality Aspect
Effort breakdown for the development activities (IT effort)

Situation
In the two projects a testing effort – defined as the effort spent from the IT department for the activities unit test, integration test, introduction/acceptance test – accounts or 23.7% resp. 29% of the total development effort as shown below.



Interpretation
Many factors can contribute to the fact that the testing effort in both projects is lower than the baseline. For example in both projects a testing strategy differing from the usual strategy was applied. However, the developers of projects regarded the introduction of inspections as one major factor to this reduction of the testing effort.

This decreased testing effort contributed to an overall reduction of the development effort, as in Project A and B the effort for inspections and testing together account for 32.5% resp. 37.8% of the total development effort, which is on average still less than the testing effort of the baseline projects.

Copyright © Fraunhofer IESE 2000

3.5 Characterization of the Cost-Benefit of PBR inspections

3.5.1 What is the effort for performing PBR inspections? + What are the effort savings from inspections?

Description

Inspections are supposed to reduce the development cost as defects are detected at a time when they are easier and less costly to fix. On the other hand, they require an up-front investment in effort, since the training and the performance of inspections requires additional effort in the early phases of the project. Therefore, it has to be investigated, whether the invested cost pays off in terms of an overall reduced development effort.

The costs of inspections are determined by the effort spent on inspections (e.g., training, creating scenarios, planning, preparation, meeting, etc.). The benefit of inspections is subjectively assessed by the project members after the inspection meeting has taken place. For this purpose, the project members estimate the effort saved in later phases due to the early detection of defects.

Object

PBR inspections

Quality Aspect

Cost-Benefit

Situation

In both projects, the savings due to inspections exceeded the inspection costs as shown below.

	Costs of Inspections	Estimated Savings
Project A	52 person days	89 person days
Project B	44 person days	102 person days

Interpretation

Based on the results the developers of both projects considered analysis and high-level design inspections as beneficial and profitable from an economic point of view.

However, in both projects the developers pointed out, that for low-level design inspections cost and the estimated benefit were almost identical. The developers in Project A did not contribute this result to the fact that low-level design inspections are not cost-effective. Rather they were of the opinion, that due to time pressure were not optimally performed so that the cost-effectiveness could be increased by investing more effort in more thorough design inspections.

Copyright © Fraunhofer IESE 2000

3.5.2 What are additional and indirect benefits of inspections?

Description

Inspections do not only provide benefits in terms of saved correction effort. There are also indirect, qualitative benefits. Inspections can contribute to the learning of the inspections participants (*know how*), they can elicit ideas for process improvements (*process improvement*), they can improve the communication between different stakeholders involved in the development (*better communication*), they can be beneficial to keep important project deadlines (*keeping deadlines*) and they can provide benefits for follow-up projects (*follow-up project*). Additionally, the quality improvements achieved by inspections can be visible in various forms. Due to the higher quality of the inspected letters sent to Allianz' insurance clients, the clients receive letters that are easier to read and understand. This higher quality results in fewer phone calls to the call-center, whose workload is therefore reduced (*fewer questions in call center*). Finally, the quality of the system can be judged based on whether it is appropriate for the (Allianz-internal) clients (*better customer satisfaction)*.

After each inspection, the project leader assessed these benefits on an ordinal scale: (none = 0 , low = 1, medium = 2, high = 3).

Object

PBR inspections

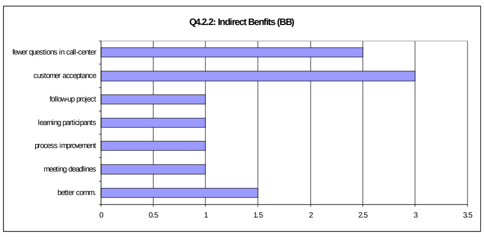
Quality Aspect

Cost-Benefits

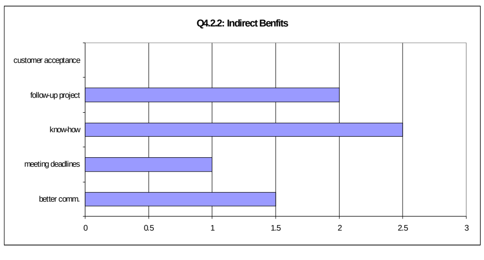
Situation

In Project A the benefits *fewer questions in call center* and *better customer satisfaction* were overall rated as high. In Project B the benefits *follow-up project* and *know how* were rated from medium to high. (Note: the values shown below were computed as the average (median) rating for each benefit in analysis and high-level design inspections).

Project A



Project B



Interpretation

The focus of Project A's inspections was on user-output descriptions such as letters to be send to customers and screen definitions to be targeted to people working in call-centers. Due to the definition of scenarios for this target group and the involvement of the future (internal) clients in early phases as inspectors, many defects regarding the user-friendliness could be detected. This led

Copyright © Fraunhofer IESE 2000

to the definition of a more appropriate system contributing to Allianz' business objective "Better customer satisfaction of delivered products" that motivated the introduction of inspections in the first place.

The emphasis of Project B was to design and implement a crucial and complex case for which many different aspects such as financial, actuarial, organizational, and implementation issues had to be taken into account. The involvement of experts in the respective domains as inspectors contributed to the learning of the developers since the document authors could gain insight into the various domains. As a result, the document authors considered this additional knowledge as valuable for future development activities.

Copyright © Fraunhofer IESE 2000

3.5.3 What is the average duration of each inspection step?

Description — The inspection process consists of several activities. Several working days can pass from the distribution of the documents to be inspected to the final completion of the inspection. In order to plan the inspection process, a rough estimate for the duration of the process should be available. Therefore, we determined the average duration between the distribution of the document to the inspection participants to the meeting and the duration from the meeting to the completion of the inspection.

Object — PBR inspections for analysis and high-level design documents

Quality Aspect — Average duration of the inspection process

Situation — In both projects, the average duration from the distribution of the documents to the inspection participants was 6 working days. The duration from the meeting to the completion was 9 resp. 20 working days.

Project A			Project B		
distribution to meeting	meeting to completion	total duration	distribution to meeting	meeting to completion	total duration
6 days	9 days	14 days	6 days	20 days	26 days

Interpretation — The duration of 6 days between the distribution of documents to the meeting is explained by the fact, that the documents are usually distributed about one week before the meeting takes place. This time was regarded as sufficient to allow the inspectors to inspect the document carefully.

The duration between the meeting and the final completion comprised activities like writing the inspection minutes, correcting the document, distributing the corrected document to the inspection participants, and the agreement of the inspection participants to the correction.

In Project A the inspection participants felt that 3-4 working days should be sufficient for a completion of the inspection after the meeting. In this project, however, the actual duration of 9 days was necessary to completely finalize the inspection minutes. This task also included the classification of defects according to several attributes (type, severity, impact, reference) for the measurement program. This additional task was largely responsible for the longer duration.

In Project B the major point contributing to the duration of 20 working days was – besides the data collection for the measurement program – the coordination with the inspection participants to agree on the corrected document. Absence times had also to be considered in this duration. Overall, the project

Copyright © Fraunhofer IESE 2000

participants felt that this duration was acceptable given the circumstances of the project.

Copyright © Fraunhofer IESE 2000

3.5.4 What is the average effort of each inspection step?

Description

In order to characterize the effort for PBR inspections, the average effort for each inspection step is useful. At Allianz the inspection process consists of the creation of Perspective-based scenarios, organization of the inspection, preparation of the inspectors (i.e., reading the document to find defects), preparation for the meeting by moderator and author, inspection meeting, writing the inspection minutes, agreement on the corrected document.

Object

PBR inspections

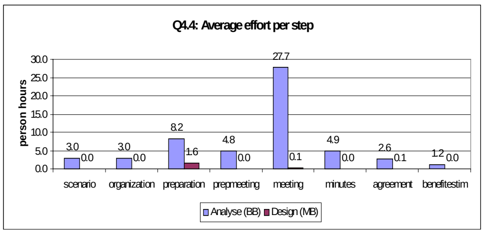
Quality Aspect

Average Effort for each inspection step

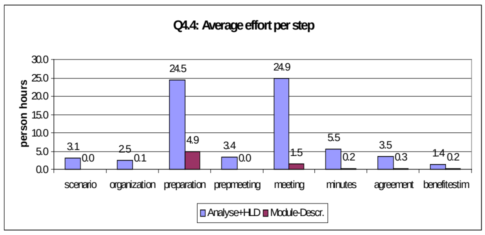
Situation

The average effort for the inspection steps is shown below.

Project A



Project B



Interpretation

In both projects the inspection participants felt that a thorough preparation of the inspectors (i.e., reading the document in order to find defects) is the prerequisite for a successful inspection. In Project A the inspection participants were of the opinion that in future inspections it should be ensured that sufficient effort is spent by the inspectors in the preparation.

Additionally, in Project A sometimes much effort was spent on explaining the background of the inspected document, especially if the inspectors were not familiar with the development project.

Improvement

It should be ensured that enough effort is spent by the inspectors for preparing the meeting and scrutinizing the document for defects. For example, an entry criterion for the inspection meeting could be defined. With this entry criterion the inspection meeting can only start, if the inspectors have spent a sufficient amount of time on scrutinizing the document for defects.

If the inspectors are not familiar with the document to be inspected or with the project at all, an introduction to the document should be given prior to the preparation step. This introduction could be organized in form of a meeting where the author explains the document, or – more informally – by having the

Copyright © Fraunhofer IESE 2000

author to distribute the document personally and give an introduction if requested by an individual inspector.

Copyright © Fraunhofer IESE 2000

3.6 Characterization of the impact of the inspectors' experience on effectiveness and efficiency

Among the factors that have an impact on inspection effectiveness and efficiency, the experience of the inspectors seems to be a major one. More experienced inspectors might be expected to find more defects than inexperienced inspectors during preparation as they might already know potential pitfalls and problems in the inspected document. However, this seemingly obvious statement should be investigated empirically. Therefore, HYPER investigated the impact of inspector experience on inspection effectiveness and efficiency.

Note: In the context of HYPER it was decided not to collect data from individual inspectors but from perspectives only (i.e., if two or more inspectors of perspective X were present, defects were recorded for perspective X making it impossible to distinguish who of the inspectors found the defect). Thus, to investigate the impact of inspector experience on inspection effectiveness and efficiency, only perspectives represented by exactly one inspector could be analyzed. Since important perspectives often were represented by two or more inspectors, these important perspectives could not be taken into account. This might have biased the results[4] presented in the sections below.

[4] The detailed test results are presented in Section 6

Copyright © Fraunhofer IESE 2000

3.6.1 What is the impact of domain experience on efficiency and effectiveness?

Description

There are several ways to describe the different skills of an inspector. One aspect is the experience in the application domain that the inspectors represents in an PBR inspection. It might be expected that more experienced inspectors might find more defects (of severity *very critical* and *critical*) per page than less experienced inspectors (effectiveness) or more defects (of severity *very critical* and *critical*) per person hour (efficiency).

The experience in the application domain was for each perspective/inspector subjectively assessed by the project leader on an ordinal scale as (low, medium, high). The effectiveness was defined as the percentage of defects (of severity *very critical* and *critical*) detected by an inspector. The efficiency was defined as the number of defects (of severity *very critical* and *critical*) detected by an inspector per person hour.

Object

Inspectors in PBR inspections

Quality Aspect

Impact of domain experience on effectiveness and efficiency

Situation

In order to test, whether any statistical difference exists in the effectiveness or efficiency for different levels of experience, the Kruskal-Wallis test was performed [9]. The result was that in both projects no statistical significant difference could be observed between the different levels of experience.

Interpretation

In Project A the developers contributed the fact that no difference between different levels of experience in the application domain for the respective perspective could be observed (both regarding effectiveness and efficiency) to the kind of inspected artifacts. In this project user-interface descriptions were inspected with the focus on customer- and user-friendliness. Thus, the experience in the different application domain was of minor importance, since no specific background was required to detect user-friendliness defects. If technical aspects had been the subject of inspection, the experience in the application domain might have shown a larger impact.

In Project B the developers had despite the negative test result the intuitive feeling that more experienced inspectors were more ‘effective’. The reason why this intuition did not match the test result was the way to measure the effectiveness: the number of defects detected (per page or as percentage) was deemed as inappropriate since more experienced inspectors might not necessarily find more defects than inexperienced inspectors. However, they could be expected to find more crucial and more subtle defects. Thus, in assessing the impact of domain experience in the future, more aspects of the detected defects have to be investigated than just their number.

Copyright © Fraunhofer IESE 2000

3.6.2 What is the impact of inspection experience on efficiency and effectiveness?

Description

There are several ways to describe the different skills of an inspector. One aspect is the experience with inspections. It might be expected that more experienced inspectors might find more defects (of severity *very critical* and *critical*) per page than less experienced inspectors (effectiveness) or more defects (of severity *very critical* and *critical*) per person hour (efficiency).

The experience of an inspector in PBR inspection was measured as the number of inspections the inspector had been participating in. The effectiveness was defined as the percentage of defects (of severity *very critical* and *critical*) detected by an inspector. The efficiency was defined as the number of defects (of severity *very critical* and *critical*) detected by an inspector per person hour.

Object

Inspectors in PBR inspections

Quality Aspect

Impact of domain experience on effectiveness and efficiency

Situation

In order to test, whether any statistical difference exists in the effectiveness or efficiency for different levels of experience, the Kruskal-Wallis test was performed [9]. In Project A no statistically significant difference between inspectors with different levels of inspections experience could be observed. In Project B a statistical difference in the inspection effectiveness and efficiency could be observed between inspectors participating for the first and second time

Interpretation

The developers in Project A considered the experience in terms of participated inspections less important than the experience in the application domain. The observable difference between inspectors with different levels of experience was that inspectors with higher inspection experience discussed and decided issues faster than inexperienced inspectors. Additionally, for inexperienced inspectors often introductions regarding the objectives of PBR were given, which had an impact on the duration of the meeting and hence the inspection efficiency.

In Project B the significant difference was explained by the fact that in a first inspection the new inspectors might have been not as thoroughly prepared as inspectors with more inspection experience and that new inspectors were a little cautious in discussions. Yet, once an inspector experienced the first inspection, there was no difference in the inspection participants.

Copyright © Fraunhofer IESE 2000

3.7 Accuracy of Defect Content Models

3.7.1 What is accuracy of the state-of-the-art defect content models for each analysis and design inspection?

Description

After an inspection meeting it has to be decided whether the inspected document is of sufficient quality. A useful metric for this decision would be the number of defects that were *not* detected in the inspection. This metric, however, is not available directly after the inspection meeting. Defect Content Estimation Models [2], [1] aim at estimating the number of remaining defects. In the context of Allianz it was investigated, whether these models provide accurate enough estimates to be applicable for controlling the inspection process at Allianz.

Object

Defect Content Estimation Models

Quality Aspect

Accuracy

Situation

Defect content models estimate the number of remaining defects based on the number of defects that are detected by several inspectors. Thus, they require collecting, which defect was detected by which inspector. In the context of HYPER, however, it was decided not to collect data from individual inspectors but from perspectives only (i.e., if two or more inspectors of perspective X were present, defects were recorded for perspective X making it impossible to distinguish who of the inspectors found the defect).

The overlap in detected defects between different perspectives was rather low. This could be explained by the fact that different perspectives look at different aspects of the system and, therefore, also for different populations of defects. This situation made the case for selecting the appropriate Defect Content Model: Generally, Defect Content Models can be divided into Capture-Recapture Models, which have their origin in biology to estimate the size of animal populations based on incomplete samples, and graphical approaches. Capture-Recapture Models assume that all inspectors (or perspectives) consider the same population of defects. Since it, however, seemed that different perspectives focussed on different defect populations, the application of Capture-Recapture Models seemed not justified in this context.

Therefore, a graphical approach, namely the Extended Detection Profile Method (EDPM) [2] was used. The results are shown in the table below.

Copyright © Fraunhofer IESE 2000

Inspection ID	Defects in Inspection[5]	Defects found in Test	Total Defects	EDPM	RE
1	16	10	26	25	-0.04
2	8	5	13	11	-0.15
3	6	0	6	6	0.00
4	17	16	33	32	-0.03
5	11	3	14	20	0.43
6	8	6	14	11	-0.21

Table 2 Accuracy of Defect Content Models for Documents of type user-output description in Project A

Interpretation

The table shows for each inspection of a user-output description in Project A the number of defects (of severity *very critical* and *critical*) detected in inspections, the number of defects detected during test with origin in the inspected document, the total number of defects (i.e., the sum of inspection and test defects), the estimate of the total number of defects according to the EDPM and the accuracy of this estimate in terms of the relative error[6].

It can be seen that the estimates are usually close to the actual values (with an median absolute relative error of 0.14). Although this accuracy is encouraging, further investigation is necessary. First, it has to be clarified, whether defects found in user-interface descriptions (i.e., issues wrt. user-friendliness) can be detected by a re-inspection called for by a defect content estimate. Second, it has to be clarified, what the impact of the fact is that defects are recorded per perspective and not per inspector. Third, the overall accuracy should be determined based on more inspections and their defect content estimates.

[5] Only defects of severity *very critical* or *critical*

[6] $RE = \frac{\hat{N} - N}{N}$ with $\hat{N}$ being the EDPM estimate and N the actual total number of defects

Copyright © Fraunhofer IESE 2000

4 Lessons Learned

In this section, we report the qualitative experiences we encountered in the course of implementing and performing inspections and their accompanying measurement program. These lessons learned might be of interest of others who plan to implement and perform inspections or perform a measurement program.

In the following to subsections both the lessons learned with respect to PBR inspections and the GQM measurement program are presented. For each lesson learned the associated measurement objects and its quality aspect are reported along with a description of the lesson learned.

4.1 Lessons Learned with respect to PBR Inspections

4.1.1 Lesson Learned: Create a positive atmosphere

Object	PBR inspections
Quality Aspect	Introduction and performance of inspections in a project
Description	A crucial point to the transfer of inspections is the creation of an open and constructive atmosphere during inspection meetings. This is due to the fact that the authors of the inspected artifacts must not get the feeling that they are blamed for the defects that are detected in "their" artifacts. Otherwise they will resist the distribution of "their" artifacts to other people. Therefore, it is the main task of the moderator to create and maintain this open and constructive atmosphere. For example, s/he has to be able to cope with possibly antagonistic participants. Special moderator training is therefore helpful for the people who are going to this responsibility. However, the inspection participants should also be reminded that they play an important role in ensuring a constructive atmosphere in inspection meetings.

Copyright © Fraunhofer IESE 2000

4.1.2 Lesson Learned: Use appropriate terminology

Object	PBR inspections
Quality Aspect	Introduction and performance of inspections in a project
Description	To promote a positive atmosphere, Allianz Life uses a special word for issues raised during inspections: "finding" (German: Erkenntnis). This word is not a synonym for "defect", because it denotes, on the one hand, defects in a narrower sense and on the other hand, questions, improvement proposals, and comments. Besides, the meaning of the word "finding" is entirely positive, so that negative associations do not arise at all. Thus, the objective of inspections was to "gain findings" and not to "detect defects." This was of psychological importance, since the term "gain findings" conveyed a positive meaning, which facilitates the acceptance of inspections.

4.1.3 Lesson Learned: Acceptance of Perspective-based Reading

Object	Reading technique Perspective-based Reading
Quality Aspect	Acceptance of Perspective-based Reading
Description	The inspectors found Perspective-based inspections to be easy to use and practical. The approach was very suitable for the Allianz' environment, where various stakeholders are interested in the project deliverables to be inspected. The scenarios for the perspectives were considered very helpful – especially for inspectors with little experience in the domain and Perspective-based inspections – since the scenarios guided the inspectors in scrutinizing the document for defects.

Copyright © Fraunhofer IESE 2000

4.2 Lessons Learned with respect to performing a GQM measurement program to measure PBR Inspections

4.2.1 Lesson Learned: Definition of defect classifications

Object	GQM measurement program
Quality Aspect	Definition of measures
Description	In the measurement program defects were classified according to several attributes such as type, impact, severity, etc. Although an initial definition for these attributes and their values was created in the planning of the measurement program, discussions evolved during data collection about the meaning of the attribute values. As a result of these discussions the definition of the classification scheme evolved and changed over time. Thus, it was hardly possible to compare the hypotheses of the experts given at the beginning of the project with the final actual results. Therefore, when defining classification schemes for measurement programs, the future data collectors should be involved in the definition of the scheme. Additionally, the definition of attribute values should augmented with example defects from similar projects.

4.2.2 Lesson Learned: Coaching of data collection

Object	GQM measurement program
Quality Aspect	Motivation of the data collection
Description	A very crucial aspect during the measurement program is to continuously support and motivate the project members to collect data, since measurement requires a constant, additional effort. In our case, the project team members have to specify their project effort and classify defects. Usually, these data collection tasks are regarded as additional burden. Thus, especially in the initial phase of a measurement program, a coaching team has to constantly demonstrate the benefits of measurement to project members.

Copyright © Fraunhofer IESE 2000

4.2.3 Lesson Learned: Tool-Support of data collection

Object	GQM measurement program
Quality Aspect	Data collection
Description	Additionally, data collection and the subsequent analysis should be assisted by appropriate tools. In this experiment, data collection is mainly performed manually on paper-based forms or by filling in WinWord templates on-line. This imposes problems when converting these raw data into a format being appropriate for a subsequent analysis: Paper-based forms have to be entered manually into the computer, from the WinWord templates the appropriate data have to be extracted. This proved to be a time-consuming process. Furthermore, when data collection is regarded a tedious activity, such as the classification of defects detected during testing, developers might feel more compelled to provide high quality data, when they are assisted by tools that are quick and easy to use.

Copyright © Fraunhofer IESE 2000

5 Summary and Conclusions

This report described the successful application of PBR inspections in an industrial setting. Based on two real-world projects of strategic importance for Allianz Life, it was shown in quantitative terms that the usage of innovative defect detection techniques in an inspection context throughout the early phases of the development life-cycle

- detect the defects more locally as 72% to 100% of analysis defects are detected in the analysis phase and 25% to 58% of design defects are detected in the design phase;
- have a cost-benefit ratio of about 1:2;
- reduce the testing effort from an average 39.5% to between 23.8% and 29%;
- reduce the overall development time as the additional effort for inspections is less than the saved testing effort.

From the Process Improvement Experiment we could also learn about the influences of the human factor on the success of the transfer initiative. The main success factors were :

- a careful and role-dependent motivation of all participants of inspections, including management;
- training in the basic technologies (PBR inspections and goal-oriented measurement) as an initial investment;
- creation of an open and constructive atmosphere during inspection meetings;
- usage of an appropriate terminology.

As a result of the case studies Allianz life decided to broaden the application of PBR inspections in future development projects. However, additional work is required to optimise the cost-effectiveness of inspections and to investigate further improvement opportunities regarding the overall verification and validation process.

Copyright © Fraunhofer IESE 2000

6 Appendix A – Test Results

One objective of HYPER was to investigate the impact of inspector experience on the inspection effectiveness and efficiency. Inspector experience was separated into domain experience (measured on an ordinal scale *low*, *medium*, *high* assessed for each inspector by the project leader) and inspection experience (measured as the number of inspections the inspector participated in). Inspection efficiency was defined as the number of defects of severity *very critical* and *critical* the inspector detected per person hour. Inspection effectiveness was defined as the proportion of defects found by an inspector compared to the total number of defects in the document. (Note: only inspectors who represented one perspective on their own were considered in this analysis. This might have biased the results).

In order to decide, whether differences in the inspectors' effectiveness or efficiency for different levels of experience are due to chance, a statistical test has to be performed. In this analysis we performed the Kruskal-Wallis-Test [9], a non-parametric alternative to the ANOVA).

The Kruskal-Wallis-Test tests, whether the differences in the median values of two or more populations are statistically significant. For this purpose the datapoints (here: the effectiveness or efficiency values of the individual inspectors) are sorted into an increasing order, their ranks are recorded and summed up. In this appendix we report for each level of experience the number of inspectors with that level of experience, the mean effectiveness or efficiency for that level of experience, the sum of ranks and the mean rank. The overall test result is reported as test significance. If this value is smaller than 0.10, the differences between the populations are statistically significant.

In the following two subsections the test reports are reported with respect to the analyses of domain experience and inspection experience. In each subsection, the tests concerning inspection effectiveness and efficiency for the two considered projects are presented.

Copyright © Fraunhofer IESE 2000

6.1 Impact of the domain experience on inspection effectiveness and efficiency

Level	No. of inspectors	Mean	Score Sum	Score Mean	Test Significance
Low	9	0.73	121.5	13.50	P=0.8009
Medium	11	0.90	173.5	15.77	
High	8	0.66	111	13.88	

Table 3 Test Results efficiency and domain experience for Project A

Level	No. of inspectors	Mean	Score Sum	Score Mean	Test Significance
Low	9	0.12	126	14.00	P=0.6669
Medium	11	0.15	177.5	16.14	
High	8	0.11	102.5	12.81	

Table 4 Test Results effectiveness and domain experience for Project A

Level	No. of inspectors	Mean	Score Sum	Score Mean	Test Significance
Low	0	--	--	--	P=0.2026
Medium	10	0.62	66	6.60	
High	4	1.42	36	9.75	

Table 5 Test Results efficiency and domain experience for Project B

Level	No. of inspectors	Mean	Score Sum	Score Mean	Test Significance
Low	0	--	--	--	P=0.3558
Medium	10	0.07	68.5	6.85	
High	4	0.12	36.5	9.13	

Table 6 Test Results effectiveness and domain experience for Project B

Copyright © Fraunhofer IESE 2000

6.2 Impact of the inspection experience on inspection effectiveness and efficiency

Level	No. of inspectors	Mean	Score Sum	Score Mean	Test Significance
1	11	0.66	147.5	13.41	P=0.4957
2	6	0.79	99	16.50	
3	7	0.69	87.5	12.50	
4	2	1.08	35	17.50	
5	1	2.40	28	28.00	
6	1	0.33	9	9.00	

Table 7 Test Results efficiency and inspection experience for Project A

Level	No. of inspectors	Mean	Score Sum	Score Mean	Test Significance
1	11	0.13	159.5	14.5	P=0.2161
2	6	0.18	118	19.67	
3	7	0.10	77.5	11.07	
4	2	0.09	21	10.50	
5	1	0.30	25	25.00	
6	1	0.03	5	5.00	

Table 8 Test Results effectiveness and inspection experience for Project A

Level	No. of inspectors	Mean	Score Sum	Score Mean	Test Significance
1	9	2.11	51.5	5.72	P=<u>0.0268</u>
2	4	7.00	48.5	12.13	
3	1	2.00	5	5.00	

Table 9 Test Results efficiency and inspection experience for Project B

Level	No. of inspectors	Mean	Score Sum	Score Mean	Test Significance
1	9	0.07	56.5	6.28	P=<u>0.0846</u>
2	4	0.19	45	11.25	
3	1	0.04	3.5	3.50	

Table 10 Test Results effectiveness and inspection experience for Project B

Copyright © Fraunhofer IESE 2000

7 References

[1] Lionel Briand, Khaled El Emam, Bernd Freimut, and Oliver Laitenberger, Quantitative Evaluation of Capture Recapture Models to Control Software Inspections, in Proceedings of the 8th International Symposium on Software Reliability Engineering, pp. 234-244, 1997.

[2] Lionel C. Briand, Khaled El Emam, and Bernd G. Freimut, A Comparison and Integration of Capture-Recapture Models and the Detection Profile Method, in Proceedings of the 9th International Symposium on Software Reliability Engineering, pp. 32-41, 1998.

[3] Lionel C. Briand, Christiane Differding, and H. Dieter Rombach, Practical Guidelines for Measurement-Based Process Improvement, Software Process Improvement and Practice Journal, vol. 2, no. 3, 1997.

[4] Bernd Freimut and Brigitte Klein, GQM and Measurement Plan Documentation, Document code: ESSI project # 27839 – Deliverable D3-1, D3-2 – Version 1.0, FhG IESE

[5] Bernd Freimut, Christiane Gresse von Wangenheim, Susanne Hartkopf, and Günther Ruhe, Training Tutorial: Goal-oriented measurement with the GQM Approach, Document code: ESSI project # 27839 – Deliverable D2-1 – Version 1.0, FhG IESE

[6] IBM Research; Center for Software Engineering; Orthogonal Defect Classification
http://www.research.ibm.com/softeng/ODC/ODC.HTM

[7] Oliver Laitenberger, Training Tutorial: Perspective-based Inspections, Document code: ESSI project # 27839 – Deliverable D2-2 – Version 1.0, FhG IESE

[8] Brigitte Klein, Manual of the Overall Verification and Validation Approach, Document code: ESSI project # 27839 – Deliverable D6-1 – Version 1.0, Allianz Lebensversicherungs-AG

[9] Sidney Siegel and N. John Castellan, Nonparametric Statistics for the Social Sciences, McGraw Hill, 1988

[10] Rini van Solingen and Egon Berghout, The Goal/Question/Metric Method. McGraw-Hill, 1999.

Copyright © Fraunhofer IESE 2000

Copyright © Fraunhofer IESE 2000

Document Information

Title:	Experience Package from the ESSI Process Improvement Experiment HYPER
Date:	February 2000
Report:	IESE-015.00/E
Status:	Final
Distribution:	Public

Copyright 2000, Fraunhofer IESE.
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means including, without limitation, photocopying, recording, or otherwise, without the prior written permission of the publisher. Written permission is not needed if this publication is distributed for non-commercial purposes.