Investigating the effect of Process Experience on Inspection Effectiveness

Jeffrey Carver[1,2]
carver@cs.umd.edu

Forrest Shull[3]
fshull@fc-md.umd.edu

Victor Basili[1,2,3]
basili@cs.umd.edu

[1]University of Maryland
Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742
USA

[2]Experimental Software Engineering Group
Department of Computer Science
University of Maryland
A.V. Williams Building
College Park, MD 20742
USA

[3] Fraunhofer Center - Maryland
University of Maryland
4321 Hartwick Road
Suite 500
College Park, MD 20742
USA

ABSTRACT

Requirements inspections are a process for improving the quality of software by allowing software developers to detect defects early in the lifecycle when they are cheaper to fix. One issue that arises is the experience level with a particular inspection technique an inspector needs before he or she is effective and efficient in using that technique. This technical report describes a study run in CMSC735 in Fall 2001. The goal of this study was to begin to understand the impact of process experience on a software inspection. Some of the subjects were given a chance to observe an inspection using the Perspective Based Reading (PBR) techniques before they had to use these techniques themselves. This report discusses how the particular experience with process was evaluated and how the efficiency and effectiveness of these subjects compared with that of the subjects who did not get the opportunity to observe someone else using PBR prior to their own use of it.

General Terms

Measurement, Requirements, Experimentation, Verification

Keywords

Empirical studies, requirements inspections, software process, experimental process, software quality

1. INTRODUCTION

Software inspections have been shown to be a practical process for ensuring that artifacts created throughout the software lifecycle possess the required quality characteristics. For instance, inspections have been used to improve the quality of design and code by helping to detect and remove defects during development [Fagan76]. In this way, inspections help reduce the number of defects in a software system by ensuring that its artifacts correctly reflect the desired quality properties.

This technical report describes an empirical study aimed at improving the effectiveness and efficiency of inspections. Software product quality can be measured in a number of different ways. We chose the defect density of key software artifacts as our measure. Because a defect is an instance where a software artifact does not correctly translate information from the previous lifecycle phase, it is a sensible measure for evaluating software quality. In this paradigm, a lower number of defects indicates an artifact of higher quality. An effective process for reducing defects, and therefore increasing correctness, has been inspections. This study examined a specific inspection technique that has the goal of improving the correctness of a requirements document.

Previous studies of requirements inspection techniques have focused on the details of requirements inspection technique itself, e.g. [Porter97b]. While the goal of these earlier studies was to improve the steps of the technique, the inspectors who are executing the technique are just as important and have been neglected in terms of study. Because the requirements inspection process is human based, the variations among the individual inspectors involved will likely have an impact on the results of the inspection; therefore study of the characteristics of the inspectors is an important task. Specifically, we studied the inspector's experience with the technique, in terms of familiarity with its use.

2. BACKGROUND

Before the specific details of this study are discussed, some background information necessary to understand the study will be presented. Sections 2.1-2.3 provide some definitions of the major pieces of the work to be described in this report. Section 2.4 gives the high level goals of the study. Finally Section 2.5 discusses the impact the results of this work can have on practice.

2.1 Inspections

Because of the multiple types of inspections, the type that is used in this work needs to be defined. One of the earliest and most widely known inspection methods is the Fagan-style inspection [Fagan76]. In the Fagan inspection, each team member is assigned a role. Based on their role, the team members do some individual preparation to familiarize themselves with the document to be inspected. The inspectors are generally given little guidance on how to do the individual preparation effectively. After the individual preparation the team members meet to detect defects. In the Fagan inspection, the team meeting is the central activity. A series of studies have been conducted showing that the team meeting may not be a necessary part of the inspection process in terms of defects detected [Votta93]. Because the research has shown that meetings may not be as critical as once thought, and because the individuals are given little guidance for individually reviewing a software artifact, we focus on the individual inspector and his or her activities.

2.2 Reading Techniques (PBR)

In order to help inspectors find requirements defects during their individual preparation or inspection time, a technique called Perspective Based Reading (PBR) was developed [Basili96]. PBR has three important aspects. First, because a requirements document will potentially be used by a number of different stakeholders, PBR asks each inspector to assume the perspective of one of those stakeholders. By taking a perspective, the inspector is focused on ensuring that the information present is sufficient for that stakeholder to do his or her job. The perspectives initially identified in the original study were *tester, designer,* and *user*.

The second aspect of PBR is the procedure that is followed for each perspective. For each of the identified perspectives, an abstraction or model of the requirements is chosen. For example, the *tester* could create a set of test cases; the *user* could create a user manual; and the *designer* could create a design providing high-level details of potential classes, attributes and methods. The PBR procedure then provides the inspector with a series of steps to create that model based on the requirements document.

The third aspect of the PBR procedure is the defect taxonomy. First, the important classes of defects have to be identified. After identifying the classes of defects (omitted information, incorrect information, inconsistent information, ambiguous information, and extraneous information), a series of questions is inserted into the above procedure to help the inspector identify each relevant defect type. After each step of creating the model, questions asks the inspector to look for defects from the specific defect classes. Based on these three aspects a procedure is created for each reviewer to follow. For more information on PBR see [Shull00].

2.3 Background and Experiences of Inspectors

In addition to the specific methods and techniques being used during software development, the individual differences among the people performing these methods and techniques can have an impact. Researchers have suggested that the selection of inspectors based on their characteristics can impact the defects found during the inspection process [Parnas85], [Porter97a], [Sauer00]. These characteristics can include, for example, software development experience and domain experience. Some of the characteristics that have been investigated in the past concern the experience of the inspectors in different tasks, such as writing requirements, using requirements, testing software, writing use cases, and so forth. Because software developers create many software products and will use the same or similar development and inspection techniques on each piece of software created, another characteristic worth studying is the experience the inspector has with the specific modeling technique or process used during the inspection. This investigation aims to understand if there is any difference between inspectors who are familiar with a technique and those who are using the technique for the first time.

2.4 High level goals of the study

This study investigates the effect that experience with an inspection technique will have on the use of that technique. A better understanding of this type of knowledge is useful because the effects of experience with an inspection process will help determine whether a novice or an expert is more effective. This knowledge provides some guidance as to the type of knowledge or experience potential inspection team members should have.

This study is concerned with two measures, effectiveness and efficiency. Effectiveness can be measured as the percentage of known defects found in a software artifact, while efficiency can be measured as the effort required in finding a defect. With the constraints of a study being run in a graduate-level software engineering class in mind, a design was created to examine the impact of experience with the inspection technique on the use of that technique as compared to use without experience.

2.5 Impact of the Results on practice

The results of this type of study will have a practical impact on the planning of an inspection. For an inspection team leader who has some flexibility in choosing his team, some guidance on the choice of team members will be provided. The results here will either indicate that well trained and experienced inspectors are desired, or that this experience with the inspection technique does not matter in choosing the inspectors. In addition, the results of this study will provide some initial ideas of the learning curve on inspections and on PBR and type and amount of training necessary for a new technique to be effectively learned and used by inspectors.

3. GOALS AND OVERVIEW OF EXPERIMENT

3.1 Goals

G1 Technique Experience

To analyze **PBR** for the purpose of characterizing and understanding the impact of process experience with respect to **effectiveness and efficiency** from the point of view of the researcher.

PBR = The Perspective Based Reading procedures (User perspective only)

Effectiveness = the percentage of known defects found.

Efficiency = the amount of effort required to find each defect.

Process Experience = whether or not inspectors had used PBR before.

G2 Domain Knowledge

To analyze **PBR** for the purpose of characterizing the effect of domain knowledge with respect to **effectiveness and efficiency** from the point of view of the researcher.

Effectiveness = the percentage of known defects found.

Efficiency = the amount of effort required to find each defect.

Domain knowledge = self-reported level of familiarity with the general domain of the application (expert, somewhat familiar, not at all familiar).

G3 Software Development Experience

To analyze **PBR** for the purpose of characterizing the effect of software development experience with respect to **effectiveness** from the point of view of the researcher.

Effectiveness = the percentage of known defects found.

Software Development Experience = level of experience (none, classroom, industrial) in software development related tasks including requirements elicitation and writing, general software development experience, experience in testing, and experience in software inspections.

G4 Improvement of the technique

To analyze **PBR** for the purpose of improving it with respect to **effectiveness** from the point of view of inspectors.

Effectiveness = the percentage of known defects found.

G5 Research Instrumentation

To analyze **PBR** for the purpose of characterizing and understanding the impact of the participation of an observer in the review process with respect to **effectiveness and efficiency** from the point of view of the researcher.

Effectiveness = the percentage of known defects found.

Efficiency = the amount of effort required to find each defect.

Observer = a researcher who watches a subject perform a task and records notes about how the task was performed, including any problems encountered.

3.2 Questions

G1 Technique Experience

Q1 How does it affect the effectiveness of a subject to observe the use of PBR by someone else prior to using PBR himself or herself?

Q2 How does it affect the efficiency of a subject to observe the use of PBR by someone else prior to using PBR himself or herself compared with an inspector who does not observe the use of PBR first?

G2 Domain Knowledge

Q3 How does it affect the effectiveness of a subject to have experience in the application domain during an inspection?

Q4 How does it affect the efficiency of a subject to have experience in a domain affect the efficiency of an inspector during an inspection?

G3 Software Development Experience

Q5 How does software development experience affect the effectiveness of an inspector during an inspection?

Q6 How does software development experience affect the efficiency of an inspector during an inspection?

G4 Improvement of the technique

Q7 How can the effectiveness of the PBR techniques be improved?

G5 Research Instrumentation

Q9 How does having a process observer affect the effectiveness of the inspection?

Q9 How does having a process observer affect the efficiency of the inspection?

Q10 How does having a process observer affect the process conformance of the inspector?

3.3 Metrics

M1 True defects detected by the inspector in the requirements document
- Measured as a percentage of the known defects

M2 Effort spent during inspection
- Measured as number of hours

M3 Software Development Experience in practice
- Measured on 4 point scale:
    - Never developed software
    - Developed software on my own
    - Developed software as part of a team in a course
    - Developed software as part of a team in industry

M4 Experience Writing Requirements
- Measured on a 5 point scale
    - None
    - Studied in class or from a book
    - Practiced on a class project
    - Done once in industry
    - Done multiple times in industry

M5 Experience Writing Use Cases
- Measured on same 5 point scale as M4

M6 Experience Reviewing Requirements
- Measured on same 5 point scale as M4

M7 Experience Reviewing Use Cases
- Measured on same 5 point scale as M4

M8 Experience with Software Inspections
- Measured on same 5 point scale as M4

M9 Domain Knowledge about applying for a loan
- Measured on a 3 point scale:
  - Unfamiliar, never done it
  - Done it a few times, but not an expert
  - Very familiar, would be comfortable doing this

M10 Domain Knowledge about Mortgages
- Measured on same 3 point scale as M9

M11 Domain Knowledge about Parking Garages
- Measured on same 3 point scales as M9

M12 Qualitative feedback from the subjects
- Measured by asking:
  - How the effectiveness of PBR could be improved
  - What affect the process observer has on process conformance

M13 True defects detected by the observer in the requirements document
- Measured as a percentage of the known defects

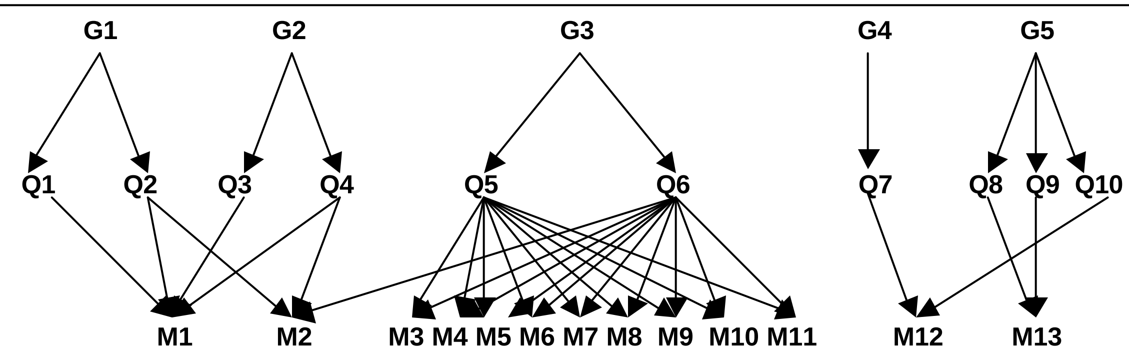

Figure 1 – Goals, Questions, Metrics Relationships

4. THE EXPERIMENT

The design from a previous experiment [Shull99, Shull01] with some modifications was used as the design for this study.

4.1 Experimenters

Researchers at the University of Maryland and the Fraunhofer Center-Maryland ran this experiment. This was the same group of researchers who had created the PBR techniques and therefore had a high level of expertise in their use.

4.2 Subjects

The subjects were graduate students at the University of Maryland enrolled in a graduate level Software Engineering class in Fall 2001. The subjects were paired up, with one subject acting as the *executor* (responsible for applying the procedure) and the other as the *observer* (responsible for recording observations about the application). More details about these roles are provided in Section 4.4.

In this study, there were 26 subjects grouped into 13 pairs. As will be described in Section 4.4, each pair performed two inspections, switching roles in between. This setup allowed all 26 subjects to perform a requirements inspection. The experience levels of the subjects was as follows:

- Experience Writing Requirements
  - 35% had industry experience writing requirements,
  - 39% had classroom experience
- Experience Writing Use Cases
  - 19% had industrial experience
  - 50% had classroom experience
- Experience Reviewing Requirements
  - 38% had industrial experience
  - 46% had classroom experience
- Domain Knowledge
  - 50% had high application domain knowledge
  - 50% had low application domain knowledge

4.3 Materials

The User perspective of the PBR reading techniques was applied to the requirements documents from two different systems: one for a Loan Arranger (LA) system, and one for an automated parking garage control system (PGCS). The LA system was responsible for organizing the loans held by a financial institution and bundling them for resale to investors. The PGCS was responsible for keeping track of how many open spaces there were in a parking garage and for keeping track of sales of both reserved (monthly) tickets as well as non-reserved (daily) tickets. The LA requirements had 8 pages, 26 functional requirements and 4 non-functional requirements, and 18 seeded defects. The PGCS requirements had 17 pages, 21 functional requirements, and 9 non-functional requirements, and 32 seeded defects.

4.4 Procedure

4.4.1 Overview

The goal of the experiment was to evaluate PBR through both quantitative and qualitative data collected about the process. In order to get a sufficient level of detail, subjects were paired up so that one could observe the other and take notes during the execution of the process. This approach, an *observational approach,* helps the experimenters understand not only the results of a process, but also how the process was applied. These types of studies provide a level of detail about individual process steps and their usefulness that is difficult to collect using traditional post-experiment questionnaires [Singer96]. Another goal of using the observational approach was to allow one team member to act as a "process guide" to keep the inspector on track and following the procedure.

In order to investigate the effects of observing PBR before using it, two requirements inspections were performed. During the first inspection, one team member performed the inspection while being observed by his or her partner. More information on the observation process will be provided in Section 4.4.4. This inspection gave the observer a chance to see PBR being used by someone else. After this first inspection was complete, the partners switched roles and were given a new requirements document to inspect. In this second inspection, the team member now performing PBR had already observed its use once. After completion of both inspections, the teams wrote a report detailing their experiences. This report was the source of much of the qualitative data collected during the study. More details about the report are in Section 4.5.

4.4.2 Experimenter's Procedure

For the design of this study it was important to characterize the subjects and make the team assignments based on subject characterization. To meet the design assumption, there were two constraints placed on each team. First, each team should consist of two subjects that were either both highly experienced in software development or both inexperienced in software development. This constraint was put in place for two reasons. The first reason was so that we could study the effect of observation on both experienced and inexperienced teams. The second reason was to eliminate one potential threat to validity. That threat being that when a low experienced subject being observed and guided by high experienced subject, there is a potentially confounding influence on the results.

The second constraint dealt with application domain knowledge. In order to study the effects of application domain knowledge, it was assumed that the Loan Arranger (LA) domain was unfamiliar to the subject population and the Parking Garage Control System (PGCS) domain was much more familiar. Therefore, at least one member of each pair was knowledgeable in the PGCS domain and at least one member was not knowledgeable in the LA domain, and each member of the team was assigned to review a document to satisfy that constraint.

Subjects were categorized based on their development experience and domain knowledge. Random pairings of the subjects were made to satisfy the above constraints. See Figure 2 below for more details.

4.4.3 Training

Before the study, subjects received training in the PBR techniques to be applied and the observational methods. Training in PBR was accomplished in a 60-minute class lecture. First the theory behind PBR was given, then the history and evolution of the techniques and finally the PBR techniques were presented along with some examples of its use. The students were then given a chance to practice and then ask questions.

Training in observational methods was done by presenting the roles of *Process executor* and *Process observer* and defining their specific responsibilities. During the 30-minute class lecture, a short example was performed. Subjects were instructed that when they performed the inspections, they were to come up with their own set of questions for eliciting information about the overall effectiveness of the techniques and the way in which the process was applied (e.g. if the procedure was too detailed or missing key information).

Finally, after the in class training was done, each pair of subjects spent 45 minutes with one of the experimenters. During this time, the subjects performed an inspection on a sample requirements document. Each subject spent part of this time as the observer and part of the time as the executor. This time gave the experimenter a chance to watch both the execution of PBR and the Observer-Executor method to help the subjects understand if they were behaving properly. The subjects were also given a chance to ask the experimenter questions about either PBR or the observation procedure.

4.4.4 Execution

A quasi-experimental, factorial design with two treatments was used [Campbell63]. In the first treatment, roughly half of the teams inspected the LA requirements and the other half the PGCS requirements. After this inspection was complete, the team members switched roles, i.e. the process observer in the first inspection became the process executor for the second inspection. The teams also switched requirements documents, from LA to PGCS or vice-versa. All subjects used the User perspective of PBR in both treatments. Figure 2 illustrates the experimental design.

After performing both treatments, each team wrote a report discussing their experiences and evaluating the PBR procedure. They were told to address at least the following issues in their reports:

- The methods they used to understand the PBR procedure
- The feasibility of PBR
- Whether or not PBR was useful for the task it was designed to accomplish
- Any suggested improvements to PBR
- How the experience of observing first affected the second executor

All conclusions drawn in the reports were to be backed up by specific observations made during the execution of PBR.

Treatments	Group 1	Group 2
Review #1	**LA**	**PGCS**
	Switch Roles	*Switch Roles*
Review #2	**PGCS**	**LA**
	4 Low Experience Teams 3 High Experience Teams	3 Low Experience Teams 3 High Experience Teams

Figure 2 - Experimental Design

4.4.5 Data Collection

Quantitative data was collected, such as the time required to perform the inspection using PBR, and the number and type of defects detected. By using the observational techniques the subjects collected the qualitative data that went into the report. The report included both direct observations taken during the inspections as well as retrospective, or post-hoc, information. The observational data included:

- Subjective evaluation of the effectiveness of the technique.
- Specific problems with steps in the technique

The retrospective data included:

- Usefulness of the different perspectives
- Practicality of the techniques and whether they would be used again
- High-level problems with the techniques.

4.5 Results

Because this study's design was based on that of a previous study, a comparison of the quantitative data from this study to the previous study was possible to provide a sanity check on the data.

Raw Data

Because of the relatively small number of subjects and exploratory nature of this study, an a value of a = 0.1 was chosen for the analysis that follows. Figures 3 through 5 summarize the raw data collected during the experiment. The complete raw data appears in Appendix B. The sections that follow will discuss the specific results relative to each experimental goal.

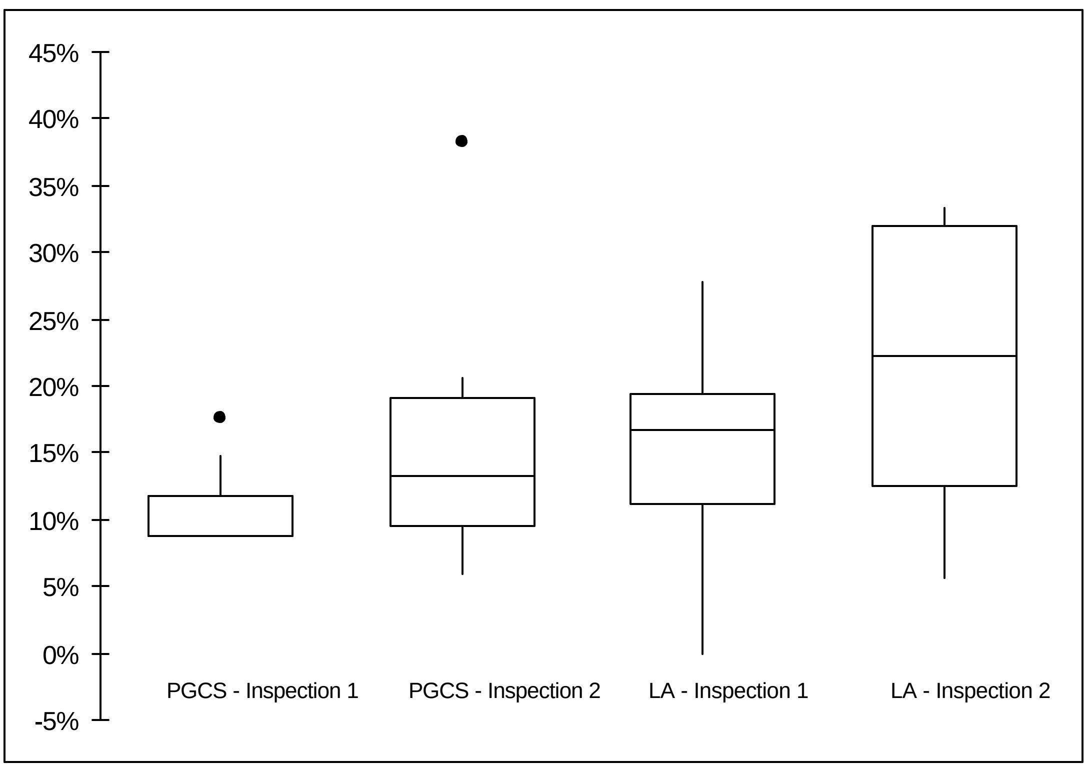

Figure 3 – Percentage of Defects Found

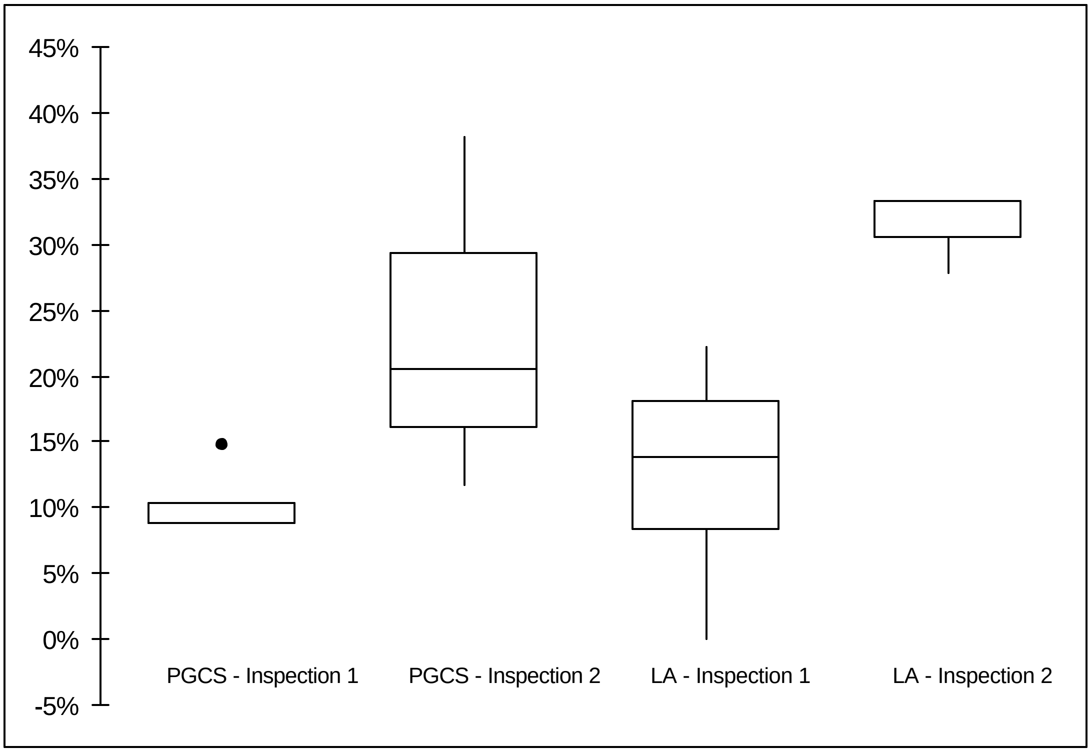

Figure 4 – Percentage of Defects Found by Low Experience Inspectors

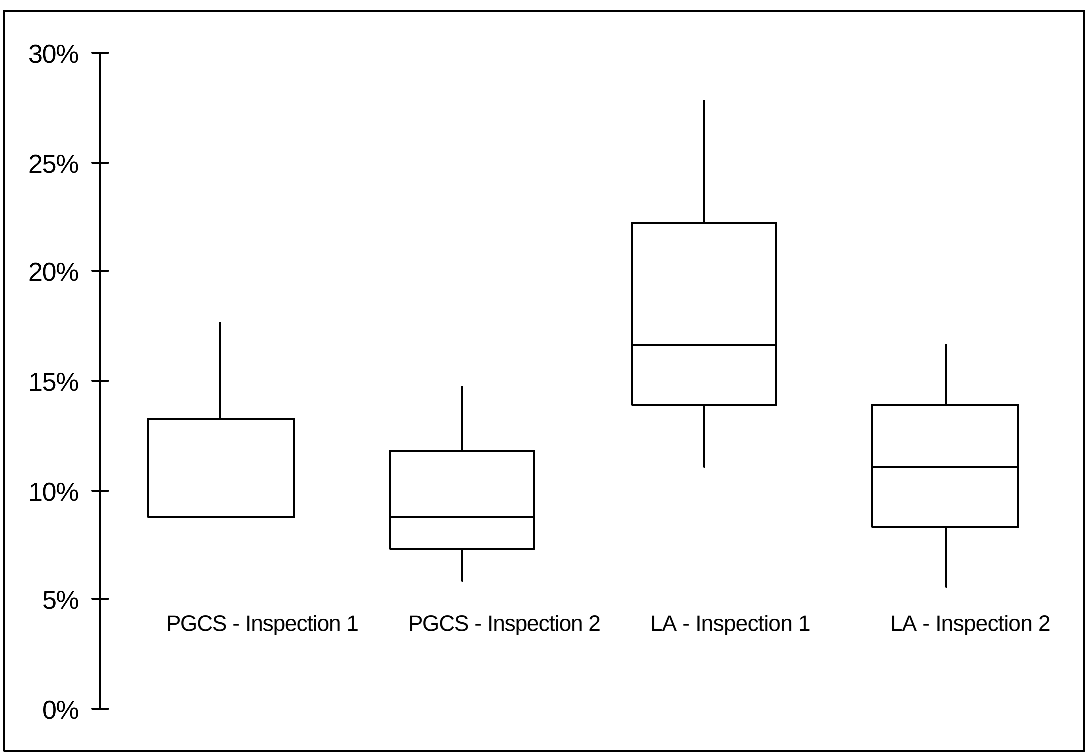

Figure 5 – Percentage of Defects Found by High Experience Inspectors

Technique Experience

Goal G1 was concerned with understanding the effect of experience with the technique on the effectiveness and efficiency of a requirements inspection. This information can be found by comparing the percentage of defects found and the time per defect by the second group of inspectors (who observed the use of PBR first) to that of the first group (who did not observe the use of PBR first).

Question Q1 asked whether inspectors with different levels of process experience had different levels of effectiveness. The qualitative data from the subject reports shows:

- Teams 1, 10, 11, and 13 (4 of the 13 total teams) thought that they understood the process better the second time
- Teams 2, 3, 5, 6, 7, 8, and 13 (7 of the 13 teams) thought they were better able to understand and perform the individual steps in the process the second time
- Teams 1, 2, 4, 5, and 11 (5 of the 13 teams) thought they were either more confident or efficient the second time
- Teams 9 and 12 (2 of the 13 teams) made no comments on this issue.
- None of the teams said that the teams stated that process experience was not helpful.

Table 1 - Percentage of Defects Found

Requirements Experience	First Inspection	Second Inspection	p-value
All Subjects			
PGCS	16.7 %	10.92 %	.146
LA	15.0 %	21.3 %	.159
Both	15.8 %	15.7 %	.979
Low Experience			
PGCS	23.5%	10.3%	.118
LA	12.5%	31.5%	.010
Both	17.2%	19.4%	.369
High Experience			
PGCS	9.8%	11.2%	.322
LA	18.5%	11.1%	.148
Both	14.2%	11.4%	.242

While in most cases there was not a statistically significant difference in the quantitative data between the subjects in the first inspection and those in the second inspection, there is some indication that this distinction is useful and should be studied further. From Table 1, the only statistically significant improvement from inspection 1 to inspection 2 was for subjects with low requirements experience inspecting the Loan Arranger ($p = .01$). But, based on the qualitative data, there is an indication that the subjects performing the second inspection felt more comfortable with the technique and thought that they better understood the assigned procedure.

Additionally, based on the design of the study, those subjects who inspected LA second observed PGCS first, and vice versa. The data in Table 1 shows that subjects who observed PGCS first seemed to have an improvement during the second inspection, while those who observed LA did not see an improvement. We can hypothesize that for the observation of an inspection of a requirements document to be helpful, that inspection must be performed on an artifact of high domain knowledge.

Table 2 – Effort per Defect (in minutes)

Requirements Experience	First Inspection	Second Inspection	p-value
All Subjects			
PGCS	29.2	35.9	.15
LA	54.2	61.5	.36
Both	41.7	47.7	.29
Low Experience			
PGCS	19.4	39	.007
LA	52.9	40	.14
Both	36.2	39.4	.36
High Experience			
PGCS	39.1	31.8	.17
LA	55.5	83.1	.27
Both	47.3	57.4	.33

Question Q2 asked if the efficiency of the inspector would change based on their process knowledge. Table 2 presents the average efficiency, in terms of minutes per defect found, for each set of subjects. As the data shows, it appears that in most cases, the subjects in the second inspection were less efficient than those in the first inspection. The only cases where the efficiency improved were for low experience subjects who were inspecting the LA requirements and for high experience subjects inspecting the PGCS requirements. Both of these cases of improved efficiency were cases where, based on Table 2, the second inspectors also found more defects than the first inspectors. So, technique experience did not show an effect on efficiency.

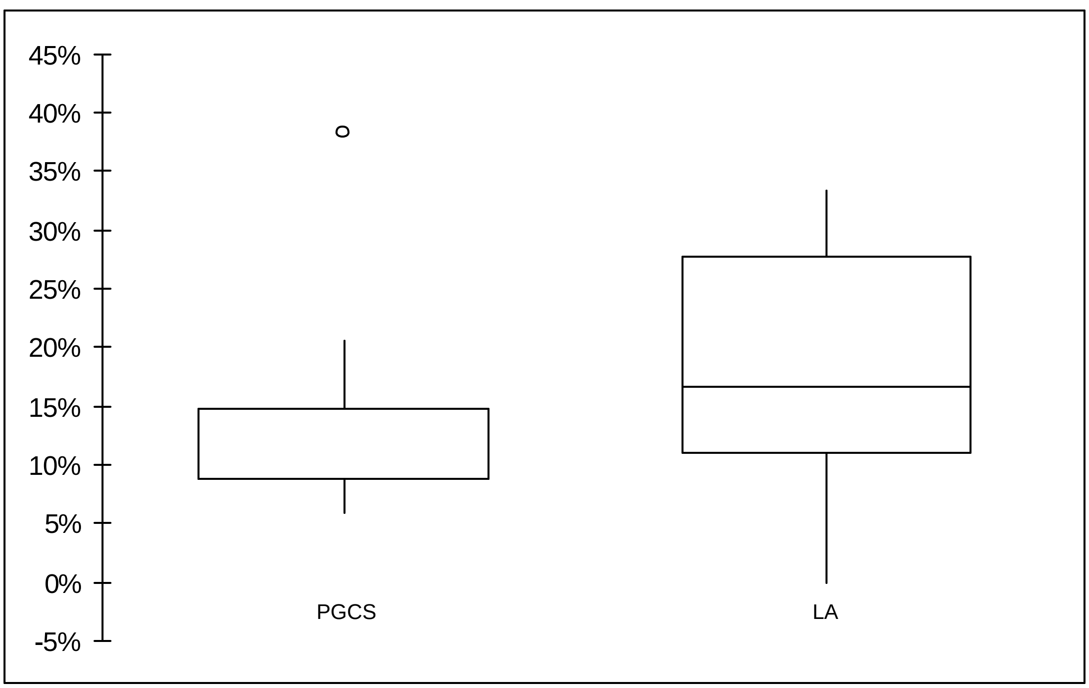

Figure 6 – Percentage of Defects Found

Domain Knowledge

Goal G2 was concerned with the effect domain experience had on the effectiveness and efficiency of a requirements inspection. To address Question Q3 dealing with effectiveness, the average percentage of defects found by the subjects in the different application domains was measured. The results, shown in Figure 6, were that the subjects inspecting the Parking Garage requirements found 13.5% of the defects on average, while the subjects inspecting the Loan Arranger requirements found 17.5% of the defects on average, counter to what might have been expected. However, this difference between the two domains is not statistically significant. Also, from Figure 6 it can be seen that the variance is higher in the LA than in the PGCS.

Table 3 – Defect Percentage Comparison with Historical Data

	This Study	735 Fall 99	435 Fall 98	735 Fall 97	NASA 95	NASA 94
Parking Garage	12.5%	22.9%	N/A	29.6%	33.93%	20.57%
Loan Arranger	17.5%	17.36%	11%	N/A	N/A	N/A
Observer?	Yes	Yes	No	No	No	No

But, based on Table 3, the subjects in this study had considerable lower defect detection rates on PGCS than did subjects in previous studies, while the LA inspectors did about the same as a previous study using graduate students (735 Fall 1999), and better than a previous study using undergraduate students (435 Fall 1998). This discrepancy could account for the poor performance of the subject on PGCS compared to the subjects on LA in this study.

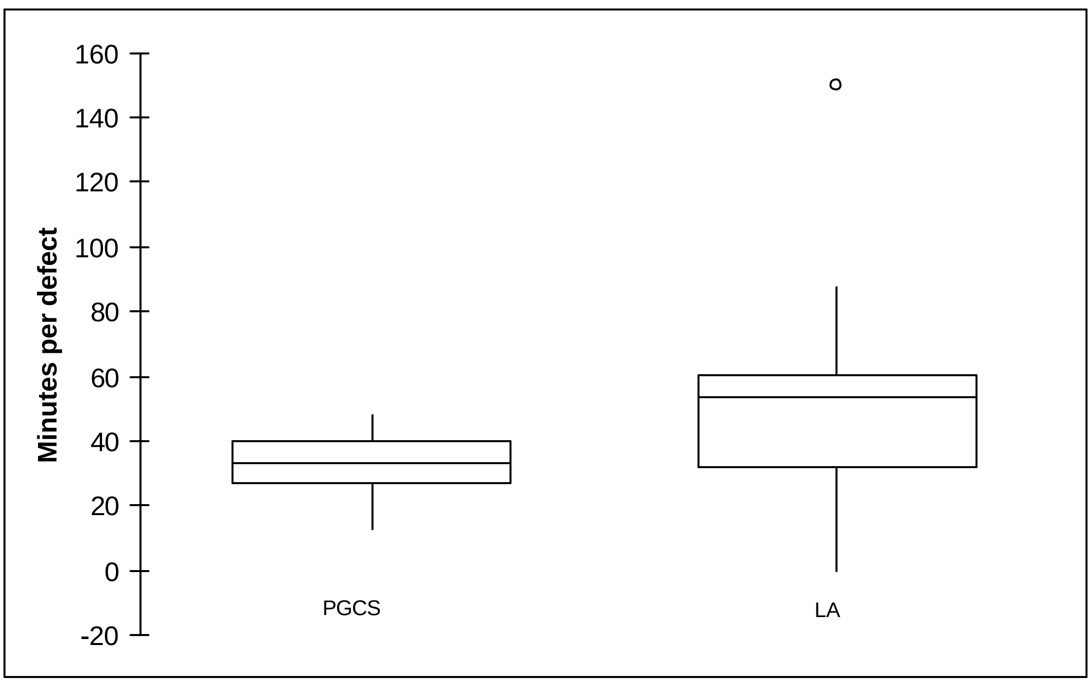

Figure 7 – Effort per Defect Found

Question Q4 address the issue of efficiency with relation to domain knowledge. Figure 7 shows that the average effort per defect for the PGCS requirements, where the subjects had high domain knowledge was 32.8 minutes, while for the LA, where the subjects had low domain knowledge, it was 53.4 minutes. This difference is statistically significant (t-test p-value= .04). So, while domain knowledge did not really improve the effectiveness of the inspection, it does appear to improve the efficiency.

Software Development Experience

Goal G3 addresses the impact of software development experience on the inspection process. Question Q5 deals specifically with the impact that experience has on the effectiveness of the inspection. Based on the data presented in Table 4, it can be seen that Software Development experience did not help the subjects' performance during the inspection. In fact, for the second inspection and overall (both inspections together), the highly experienced subjects did significantly worse than the low experienced subjects. One possible explanation for this result is that PBR neutralized the effect of software development experience, i.e., the new technique improved the performance of the inexperienced subjects while hurting the performance of the more experienced subjects.

Table 4 – Defects Found in First Inspection vs. Second Inspection

Experience	First Inspection	Second Inspection	p-value	Both Inspections
Low	17.2%	19.4%	.369	18.3%
High	14.2%	11.4%	.242	12.8%
p-value	.293	.0691	--	.067

Question Q6 deals with the impact of software development experience on efficiency of the inspection. Based on the data presented in Table 5, it can be seen that software development experience appears to make subjects less efficient during the inspection. In inspection 1 alone, inspection 2 alone, and both inspections taken together, the low experience subjects spent less effort to find each defect. Also based on Table 4 above, the number of defects found by the low experience subjects was higher than that of the high experience subjects. So, it appears that in spending about the same effort, the low experience subjects found more defects.

Table 5 – Minutes per Defect Found in First Inspection vs. Second Inspection

Experience	First Inspection	Second Inspection	p-value	Both Inspections
Low	36.2	39.4	.36	37.9
High	47.3	57.4	.33	52.4
p-value	.19	.2	--	.11

Improvement of Technique

Goal G4 was focused on determining how the techniques could be improved. Question Q7 addressed whether the effectiveness of the techniques could be improved. The subjects' reports were helpful in discovering this information. The following suggestion was made to improve PBR:

- Teams 2, 3, 6, 9, and 11 (5 of the 13 teams) thought that the questions could be improved by adding more questions to deal with defect types, or project specific issues, or historical problems

A positive result was that the suggestions that were made by the subjects were mainly dealing with the questions in the technique, the easiest piece to tailor. As discussed in Section 2.2, PBR consists of three major pieces, a perspective, a model, and a set of questions. While all three of these pieces are important, the effort required to tailor each part varies. Creating whole new perspectives requires the greatest amount of effort, while simply tailoring the questions requires the least amount of effort. These results show that PBR is useful and may require only some small tailoring to the questions within the procedure.

Using a Process Guide

Goal G5 was concerned with the effect of the process observer on the requirements inspection. Question Q8 dealt with the relationship between the process guide and the effectiveness of the techniques. Quantitative data collected was used to measure the percentage of the known defects found by the teams reviewing the PGCS requirements and by the teams reviewing the LA requirements. This data was compared with historical data on the same requirements documents to see how the subjects compare.

The qualitative data from the reports submitted by the subjects and the post experiment discussion showed:

- Teams 1, 2, and 9 (3 of the 13) teams said they thought observer either was helpful, or would be helpful in an industrial setting.
- Teams 4, 9, and 10 (3 of the 13) teams said that instead of staying in their roles as observer and executor, it would have been better if they performed a "team-inspection" where they both were responsible for finding defects. In fact, some of the teams ended up doing this type of inspection rather than exactly following the assigned procedure.
  - Those three teams all found a lower percentage of defects during the second inspection than they did in the first one.
  - If we combine the number of defects found in the first inspection and the second inspection for each team, these three teams performed near the overall average. Therefore, working as a team did not give them any consistent advantage or disadvantage compared with the other teams.

As mentioned earlier, based on Table 3, the subjects did not do considerably better with the process observer/guide present. The LA inspectors found more defects than a set of undergraduate subjects inspecting a slightly longer and more complex version of the LA requirements. Therefore, it cannot be said that the process observer helped effectiveness or efficiency in the LA requirements document. Also, for the PGCS requirements document the inspectors found fewer defects than in any of the previous studies. One potential explanation for this result is that the subjects in this study were overall less experienced in working with requirements than subjects in previous studies were. In this study, only 35% of the subjects had industrial experience writing requirements, while in the CMSC 735 Fall 1997 study 56% had industrial experience and in the two NASA studies all subjects were all industrial professionals. Based on these results, there is no support for the use of a process guide as a process improvement tool.

Question Q9 dealt with the interaction between the process guide and the efficiency of the technique. Based on the data in Table 6, it appears that the process guide did not affect the efficiency when comparing this study and the 735 Fall 1999 study (also with the process guide) to the 735 Fall 1997 study when looking at PGCS. When comparing this study to the NASA studies, which were done by more experience subjects, we see that the efficiency of the NASA studies was better. This improved efficiency could be due to the lack of the process guide, or it could be do to the increased level of experience these subjects possessed.

Table 6 – Effort (in Minutes) Comparison with Historical Data

	This Study	**735 Fall 99**	**735 Fall 97**	**NASA 95**	**NASA 94**
Parking Garage	32.8	37.1	32.1	9.4	1634
Loan Arranger	53.4	81.6	N/A	N/A	N/A
Observer?	Yes	Yes	No	No	No

Question Q10 dealt with the interaction between the process guide and process conformance of the executor. The subjects did not comment on process conformance in their reports, so we have no data to report on this question.

5. CONCLUSIONS

Further study needs to be done to better understand the effect of process experience on the effectiveness and efficiency of an inspector during an inspection. This study allowed the subjects to acquire process experience only by observing another subject using PBR. The qualitative data indicated that the subjects found this observation helpful, but the quantitative results did not show any difference. It is recommended that a future study in this area should redesign this study such that the same subject performs two or more inspections using PBR to truly measure the effect of process experience.

6. ACKNOWLEDGEMENTS

This work partially supported by NSF grants CCR9706151 and CCR9900307 and the NSF CeBASE project, grant CCR0086078.

We would like to thank the students of the CMSC735 Fall 2001 class for participating as subjects in the study.

7. REFERENCES

[Basili96] Basili, V.R., Green, S., Laitenberger, O., Lanubile, F., Shull, F., Sorumgard, S., Zelkowitz, M.V., "The Empirical Investigation of Perspective-Based Reading," *Empirical Software Engineering – An International Journal*, vol. 1, no. 2, 1996.

[Campbell63] Campbell, D., and Stanley, J. *Experimental and Quasi-Experimental Designs for Research.* Houghton Mifflin Company, Boston, 1963.

[Fagan76] Fagan, M.E., "Design and code inspections to reduce errors in program development." *IBM Systems Journal.* 15(3). 1976. 182-211.

[Parnas85] Parnas, D.L., and Weiss, D.M. "Active Design Reviews: Principles and Practice." *Proc of 8th International Conference on Software Engineering*, 1985, p132-136.

[Porter97a] Porter, A.A., and Johnson, P.M. "Assessing Software Review Meetings: Results of a Comparative Analysis of Two Experimental Studies." *IEEE Transactions on Software Engineering*. Vol. 23, no. 3. March 1997. pp 129-145.

[Porter97b] Porter, A.A., Siy, H.P. and Votta, L.G. "Understanding the Effects of Developer Activities on Inspection Interval." In *Proceedings of the 19th International Conference on Software Engineering*. 1997. Boston, MA.

[Sauer00] Sauer, C., Jeffery, D.R., Land, L., and Yetton, P. "The Effectiveness of Software Development Technical Reviews: A Behaviorally Motivated Program of Research." *IEEE Transactions on Software Engineering*. SE-26(1) : 1-14.

[Shull99] Shull, F., Travassos, G.H., Carver, J., and Basili, V.R. "Evolving a Set of Techniques for OO Inspections." University of Maryland Technical Report CS-TR-4070, October 1999.

[Shull00] Shull, F., Rus, I., and Basili, V. "How Perspective-Based Reading Can Improve Requirements Inspections." *IEEE Computer*, 33(7), July 2000, p. 73-79.

[Shull01] Shull, F., Carver, J., and Travassos, G.H. "An Empirical Methodology for Introducing Software Processes." In *Proceedings of the 8th European Software Engineering Conference*, Vienna, Austria, Sept. 10-14, 2001. p. 288-296.

[Singer96] Singer, J. and Lethbridge, T.C., "Methods for Studying Maintenance Activities." In *Proceedings of the 1st International Workshop on Empirical Studies of Software Maintenance*, Monterey, CA, 1996.

[Votta93] Votta, L.G., "Does Every Inspection Need a Meeting?" *Proceedings of ACM SIGSOFT'93 Symp. Foundations of Software Engineering*, Assoc. for Computing Machinery, Dec. 1993.

Appendix A – Experience Questionnaire

Name ____

General Background

Please estimate your English-language background:
__ I am a native speaker.
__ English is my second language. [Please complete both of the following.]

My reading comprehension skills are:
__ low
__ medium
__ high

My listening and speaking skills are:
__ low
__ medium
__ high

What is your previous experience with software development in practice? (Check the bottom-most item that applies.)
__ I have never developed software.
__ I have developed software on my own.
__ I have developed software as a part of a team, as part of a course.
__ I have developed software as a part of a team, in industry.

Please explain your answer. Include the number of semesters or number of years of relevant experience. (E.g. "I worked for 10 years as a programmer in industry.")

Software Development Experience

Please rate your experience in this section with respect to the following 5-point scale:

1 = none
2 = studied in class or from book
3 = practiced in a class project
4 = used on one project in industry
5 = used on multiple projects in industry

Experience with Requirements

Experience writing requirements	1	2	3	4	5
▪ Experience writing use cases	1	2	3	4	5
▪ Experience reviewing requirements	1	2	3	4	5
▪ Experience reviewing use cases	1	2	3	4	5
▪ Experience changing requirements for maintenance	1	2	3	4	5

Experience in Coding

▪ Experience in coding, based on requirements/use cases	1	2	3	4	5
▪ Experience in coding, based on design	1	2	3	4	5
▪ Experience in coding, based on OO design	1	2	3	4	5
▪ Experience in maintenance of code	1	2	3	4	5

Appendix A – Experience Questionnaire

Experience in Testing

Experience in testing software	1	2	3	4	5
Experience in testing, based on requirements/use cases	1	2	3	4	5
Experience with equivalence-partition testing	1	2	3	4	5

Other Experience

Experience with software project management?	1	2	3	4	5
Experience with software inspections?	1	2	3	4	5

Experience in Problem Domains

We will use answers in this section to understand how familiar you are with various systems we may use as examples or for assignments during the class.

Please rate your experience in this section with respect to the following 3-point scale:
1 = I'm really unfamiliar with the concept. I've never done it.
3 = I've done this a few times, but I'm no expert.
5 = I'm very familiar with this area. I would be very comfortable doing this.

How much do you know about:

Applying for a loan?	1	3	5
Applying for a mortgage?	1	3	5
Using a parking garage?	1	3	5
Using an ATM?	1	3	5
Renting movies from a video rental store (e.g. Blockbusters)?	1	3	5

Appendix B – Raw Data

The table below presents the raw data from this study. The subject ID identifies which subjects were paired together. "Effort" is the number of minutes reported by the subject to perform the inspection. "Observed First?" indicates whether or not the subject observed their partner performing the PBR inspection prior to performing it himself or herself. "Artifact Inspected" indicates which artifact, the Parking Garage Control System (PGCS) or the Loan Arranger (LA), was inspected. "Defect Rate" indicates the percentage of the known defects that the subject found

Subject ID	Effort (Minutes)	Observed First?	Artifact Inspected	Defect Rate	Effort per Defect
1_A	165	No	PGCS	38.2%	12.7
1_B	330	Yes	LA	33.3%	55.0
2_A	100	No	PGCS	11.8%	25.0
2_B	150	Yes	LA	33.3%	25.0
3_A	143	No	PGCS	20.6%	20.4
3_B	200	Yes	LA	27.8%	40.0
4_A	180	No	LA	22.2%	45.0
4_B	107	Yes	PGCS	8.8%	35.7
5_A	121	No	LA	11.1%	47.3
5_B	142	Yes	PGCS	8.8%	60.5
6_A	160	No	LA	16.7%	33.0
6_B	165	Yes	PGCS	14.7%	53.3
7_A	118	No	LA	0.0%	40.0
7_B	120	Yes	PGCS	8.8%	0.0
8_A	145	No	PGCS	8.8%	48.3
8_B	95	Yes	LA	16.7%	31.7
9_A	145	No	PGCS	14.7%	29.0
9_B	150	Yes	LA	5.6%	150.0
10_A	270	No	LA	27.8%	54.0
10_B	81	Yes	PGCS	8.8%	27.0
11_A	75	No	LA	16.7%	25.0
11_B	90	Yes	PGCS	8.8%	30.0
12_A	80	No	PGCS	5.9%	40.0
12_B	135	Yes	LA	11.1%	67.5
13_A	175	No	LA	11.1%	38.3
13_B	230	Yes	PGCS	17.6%	87.5