Linking Return on Training Investment with Defects Causal Analysis

Santiago Matalonga
Universidad ORT Uruguay
smatalonga@uni.ort.edu.uy

Tomás San Feliu Gilabert
Universidad Politécnica de Madrid
tsanfe@mpsei.fi.upm.es

Abstract

In this paper, we present a process for linking organizational training efforts with defects causal analysis in software development organizations. The process is being implemented in a CMMI maturity level 3 organization. Since causal analysis is not an expected process area at maturity level three, key success factors for the implementation of the process are identified and analyzed. The conclusions were tested in this software development organization. In order to do that, a pilot project was selected and training was implemented to support the process. The training results are analyzed in order to validate the overall approach. The resulting work provides a guideline for implementing causal analysis in lower maturity organizations and establishes that the implementation is viable in the target organization.

1. Introduction

Organizations nowadays invest large amounts of money in their training programs. Just in the United States, organizations invest an average of U$S 100 billion annually [1]. Organizations invest in training under the assumption that higher trained employees will result in higher quality products and reduced costs.

Nevertheless, in spite of all this investment and effort, organizations have not been so successful at providing consistent data that will link the investment on organizational training to organizational results.

Kirkpatrick's four levels of training evaluation establishes a framework against which organizations can measure up its investment in training. Kirkpatrick's four levels are called: *Reaction,* where trainees feelings towards the training are measured; *learning,* were trainees acquired knowledge is measured; *Transfer,* in which a measure of the amount of the Learned knowledge as actually put to use in the work; and *Results,* were impact to the organization's bottom line results is measured.

In this paper we propose a process for an organizational training department within a CMMI [2] maturity level 3 (CMMI L3), that addresses this problem. Our objective is to design a training process that is able to present results at Kirkpatrick's "Results" level [3].

The key aspect for the successful implementation of our process is the capability of the organization in defects causal analysis. Since the Causal Analysis and Resolution (CAR) process area belongs to maturity level 5 in the staged representation of the CMMI model, we want to make sure that the implementation of some causal analysis specific practices is possible in a level 3 organization.

In the following section, we provide an overview of the target organization, followed by our process proposal for the training department. Then in section 4, , we provide an analysis of the current state of causal analysis research. We specifically focus on linking causal analysis to organizational training. We finally provide data on the implementation and validation of causal analysis sessions within a maturity level 3 organization.

2. Brief description of the target organization and its infrastructure

The organization we are working with has recently obtained a maturity level 3 rating. It is a software factory that provides customized software solutions to in-shore and off-shore customers. Its development offices are located in Uruguay, and it has sales offices in the Caribbean and Mexico.

In the past two years, the organizational training department has invested - in US dollars - the equivalent to 7% of billable working time of their software developers. As a result, the training department needs to show the organization the return of its investment.

A Microsoft Office SharePoint Server supports the organizational measurement system. The SharePoint Server allows for the interoperation with other Office tools, for instance Microsoft Access is used for data analysis needs and the front end for data recollection is usually a Microsoft InfoPath form. The organization defect tracking system is one example.

Currently, the training department is using the described tools for its measurement needs. However, its data repository is not yet linked to the defect tracking system.

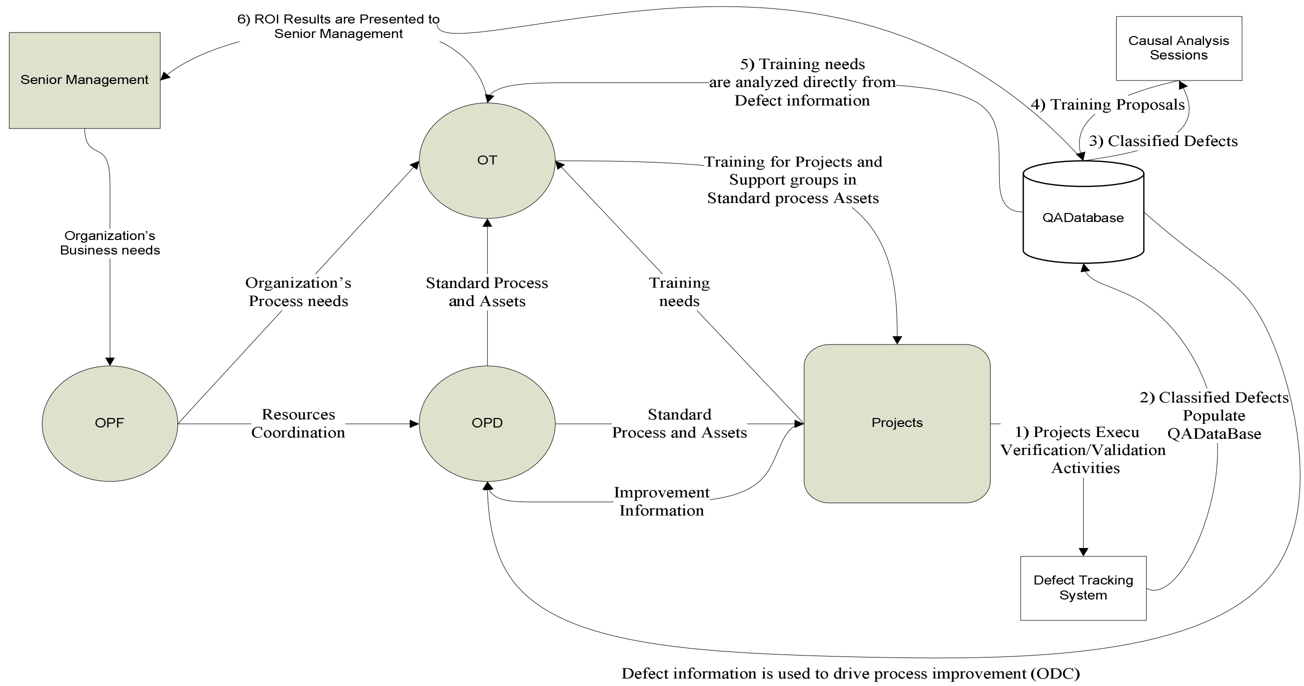

Figure 1 Overview of the Proposed Process

3. A description of our proposed process

This section describes the process we are deploying in order to link organizational training results to defects causal analysis. The idea is to make use of the information available at a CMMI L3 organization. Specifically, we want the training department to take advantage of the data available in the organization's defect tracking system. The purpose is that the training department will monitor the defect data in the system and will interpret the data as the training needs of the organization's development projects. The training department will use this information to plan training interventions specifically tailored to the development projects needs. We expect that those interventions will have positive results in the projects quality, and that those results will be noticeable in a reduction of the number of defects.

Figure 1 uses the diagram from the CMMI basic process management areas [2] in order to help us illustrate our process (items which come from the original CMMI diagram are shown in grey color, items from our proposal are left in white).

The first requirement of our process is that the defect tracking system must support a classification scheme that will allow for easier data analysis. IBM's Orthogonal Defect Classification (ODC) [4] is an example of one of such schemes, and the one we have used as reference. ODC is described in section 5.1.

In this process, when developers execute Verification and Validation activities they are required to classify the defects they log into the defect tracking system (arrow **1**). During the course of the projects life cycle, the execution of Validation and Verification activities will populate the Defect Tracking System database (QADataBase) with classified defects (arrow **2**).

At every project's milestone, the development team (arrows **3**) holds causal analysis meetings. The objective of these meetings is to provide the training department with training proposals. The idea is that developers will take the classified defect data (arrow **3**) from the QADatabase and identify which training they would have needed in order to prevent some of the mistakes that triggered the defects (arrow **4**).

The training department will use both the classified defects from the QADatabase and the outputs from the causal analysis sessions in order to plan the training interventions (arrow **5**).

Arrow **6** represents the ultimate goal of this proposal, which is the ability to show results in terms of Return on Investment (ROI) [5]. Return on Investment is interpreted as a Kirkpatrick's "Results" level measurement. Return on investment can be calculated by taking into account the defect reduction that should be noticed after the training department intervention. We expect to use the ROI classical formula to calculate ROI:

$$ROI = \frac{100 * (Benefits - Costs)}{Costs} \text{ [5].}$$

As a first step into the implementation of this process, we will turn our attention to validate arrows **3** and **4** of this cycle. First, we will present a systematic review of available literature about the implementation of causal analysis process area in lower maturity organizations. Our objective is to see if there have been previous attempts at implementing causal analysis in a context similar to the one we have described. If the review is successful, we will turn our attention into developing and validating a consistent classification scheme within the organization.

4. A review of causal analysis methodologies

Since Causal Analysis and Resolution (CAR) is a CMMI L5 process area, we cannot expect to see a CAR process deployed in a level 3 organization. Yet, the causal analysis meeting is the critical success factor of our process since the output of the causal analysis meeting (arrow **3**) is one of the inputs that the training department will need in order to plan and deploy the training interventions.

A systematic review [6] of the causal analysis bibliography was carried out. Our objective was to answer these two questions: **A)** Has anybody else tried to implement CAR at a CMMI L3 organization? **B)** What are people doing with the results of the CAR meeting?

Table 1 shows a summary of the results of the systematic review on the implementation of causal analysis in lower maturity organizations.

Table 1 Summary of the reviewed literature

Authors	Is it a Lower Maturity (L2 or L3) Organization?	Purpose of CAR analysis
Buglione, L et al. [7]	Yes	N/A just the challenge of implementing it at lower maturity organizations
Bhandari, I et al [8]	No	Process Improvement
Card D. N. [9]	No	Process Improvement
Card D. N.[10]	No	Project Defect Profile
Card D. N.[11]	N/A	Cost Saving by Defect Reduction
Mitzukami D.[12]	No	Project Defect Profile
Lezak M et al. [13]	No	Process Improvement
Fredericks M et al[14]	No	Process Improvement
Norman E.F. [15]	N/A	Defect Prediction
Bibi S. et al [16]	N/A	Project Defect Profile and Defect Prediction
Jacobs J. C. et al [17]	N/A	Project Defect Profile and Defect Prediction

In reference to our first question, we only found one reference [7] which shows that CAR can be implemented at a maturity level 3 organization (question **A**). The research done by Buglione and Abram [7], describes how it is possible to implement a CAR process area in organizations that have not yet achieved higher maturity rating. They base their implementation of the causal analysis meeting using Ishikawa (or Fishbones) diagrams [18] and defects have been classified using IBM's Orthogonal Defect Classification [4].

Finally, Table 1 shows that we have classified the results of Causal Analysis in three categories (question **B**). "Process Improvement" is the category that reports using Causal Analysis results as an input to process improvement initiatives. This use is aligned with [4] first proposal of ODC. The "Project Defect Profile" category represents initiatives that use defect data to compare the actual project to the historical projects database of the organization. They use Causal Analysis to understand deviations from the historical data and to implement corrective actions to their project and to the organization's process. In the "Defect Prediction", we have seen attempts to use artificial intelligence techniques to profile project's defects and to predict the number of remaining defects. Finally, "Cost Saving by Defect Reduction", means that the implementation of causal analysis can provide cost savings. All the results show the intended uses of causal analysis as they are recommended in the CMMI model. Unlike our research objective, none of the reviewed authors have linked causal analysis to training needs.

5. Development and Validation of the classification scheme

For a successful implementation of causal analysis session, the organization must develop a classification scheme that will enable developers to consistently classify defects. This section starts by presenting an overview of IBM's ODC taxonomy, which has served as the groundwork for our classification scheme. We then describe our implementation and our results.

5.1. Reference classification scheme: Orthogonal Defect Classification

At IBM [4, 20, 21] Defect Data classification has been used to drive Software Process Improvement initiatives. They call their process Orthogonal Defect Classification (ODC). Within the ODC context, developers classify defect data in orthogonal classifications. IBM proposes a taxonomy for defect classification based on the source of the defects:

- *Education*, in this category the developer did not understand some aspect of the product or the process. This category is further divided into education in base code, education in new function, and other education, depending on what was not understood.

- *Communication*, in this category the developer did not receive the required information or the information was incorrect.
- *Oversight*, in this category the developer failed to consider all cases and conditions. Usually some detail of the problem or process was overlooked. The developer forgot something, had difficulty checking thoroughly, or did not have enough time to be thorough.
- *Transcript*. In this category the developer knew what to do and understood the item thoroughly but simply made a mistake. Transcription errors are typically caused by some problem in a procedure, for example, typing or copying a list of items.

In relation to ODC we have reviewed the work by [19], where they recognize the importance of having a classification scheme that is consistent with the organizations' tools and culture.

5.2. Classification Scheme implementation plan

The classification scheme was developed based on the advice in [19]. In their work, the authors recognize the benefits that the organizations can obtain if they take the time to develop their own classification scheme consistent with their information needs. As a result of this work, it was decided to develop a classification scheme that was suitable for our target organization. We divided the deployment plan in four major stages (as shown in Table 2). In the first stage, we developed the custom classification scheme. Secondly, the classification scheme was deployed to the organization through its defect tracking system (this is described in this section). Then, training in developing classification ability had to be implemented in order to achieve a reliable classification capability within the organization (described in the following section).

Table 2 Classification scheme deployment plan

Stage	Activity
Organizations Capability	Design and Validate classification scheme
Deploy Classification scheme	Modify defect tracking system
Training	Design and deploy classification training Evaluate training results

As we mentioned earlier, to minimize the rejection risk of the classification deployment, it is important that the classification taxonomy was consistent with its information needs and culture[19]. A custom classification was developed with the help of the practitioners. Furthermore, the researchers established equivalence between the organization's classification and ODC (see Table 3[1]).

The organization was already using a defect tracking system whose defect states were the ones proposed by the Microsoft Solution Framework for CMMI Process Improvement [20]. Therefore, implementing the ODClike classification went seamless in the developer's culture. The new classification scheme had to be implemented into the defect tracking system front end (a Microsoft InfoPath form template), which required only 4 hours work.

Table 3 Example of the organization's defect classification scheme

Defect Category	Classification Criteria	ODC Category Equivalence
Product Integration Error	Interface implementation does not match specification. Changes are not transferred to lower layers. Data not transferred from lower layers.	Communication
Error in use or configuration of user interface controls	Controls 'Freeze'. Paging not working in DataGrids. DataGrids missing headers.	Education
…	…	…

5.3. Description of the classification training and its results

For a successful implementation of the causal analysis meetings the developers must be consistent in their defect classification across the organization. A training event was prepared for the pilot project's developers.

Three of the organization's developers attended training. One of the trainees (C1) was a senior developer of the organization. The other two developers had experience in the pilot project's product line. Moreover, since we wanted to simulate the turnover rate of the organization, we included a fourth individual (O) to our experiment. This fourth subject had no contact with the organization and was given no training in the classification scheme.

The training consisted on a 4-hour seminar split into 2 days. The seminar was prepared and given by one of the

[1] Since defect data can be sensitive to the organization, we were asked to show only what was strictly necessary to communicate our results. Hence, we decided not to include the full classification scheme here.

researchers. This kind of training is the standard training effort that this organization invests on a given process. The organization's training events of this kind are similar to the ones described in [21]. During the seminar, an explanation of the causal analysis meetings purposes was given to the trainees. The classification scheme was presented with a sample defect for each one of the 12 classification categories.

For training evaluation, we provided the trainees with a set of 31 defects which we asked them to classify. We wanted a measure of how reliable the classification was between the trained developers. For this objective we applied Cohen's Kappa [22] between each of the subjects following a process similar to the one shown in [19, 23]. Cohen's Kappa is a statistical measure of inter-rater reliability, it is used to compare the level of agreement between two subjects who are classifying the same data set. We also applied the Kappa correlation to the outsider (O), and to an expert classifier (E). Our intention was to measure how the untrained individual rated against the other subjects. Furthermore, we applied a Fleiss' Kappa [24] to the trained group, and to the trained group plus the untrained subject. Fleiss' Kappa is a statistical measure for assessing the agreement of a number of subjects classifying a fixed data set. We expect that the examination of the results will show that trained individuals score higher correlation values than untrained individuals.

Table 4 shows the results for every pair: E represents the expert rater (one of the analysts who helped develop the classification). C1 – C3 represent the trained developers and O represents the outsider. While Table 5 presents the Fleiss' Kappa results with and without the Outsider.

Table 4 Cohen's Kappa between subjects

Subjects	**Cohen's Kappa significance**	**Subjects**	**Cohen's Kappa significance**
E – C1	0,73	C1 – C3	0,47
E – C2	0,55	C1 – O	0,36
E – C3	0,52	C2 – C3	0,52
E - O	0,34	C2 – O	0,37
C1 – C2	0,62	C3 – O	0,30

Table 5 Fleiss' Kappa calculation results

Subject Group	K agreement
E-C1-C2-C3	0,53
E-C1-C2-C3-O	0,44

Table 6 presents the significance agreement intervals for both Cohen's and Fleiss' Kappas proposed by [25]. Based on this table, we set the objective for accepting developers' classification agreement in the "Substantial Agreement" interval or above.

Table 6 Kappa significance table for Cohen and Fleiss

Cohen's Kappa	**Significance**
< 0	Poor Agreement
0.00 – 0,20	Slight Agreement
0,21 – 0,40	Fair Agreement
0,41 – 0,60	Moderate Agreement
0,61 – 0,80	Substantial Agreement
0,81 – 1	Almost Perfect Agreement

Table 5 shows that the group did not achieve the target score, which is an indication of the effectiveness of the training. Nevertheless, the most remarkable result is that the inclusion of an untrained individual does not reduce the overall agreement significance.

The result on the exercise provides enough confidence in that the developers (C1-C3) are able to consistently classify defects according to the organization's classification scheme. Such results will enable the training department to process the causal analysis meetings output as input for specific training to the projects. In addition to this, results also show the importance of training for supporting the classification scheme.

The results in Table 4 show that training has effect on the subjects' ability to classify. This conclusion is drawn from Table 4 where we can see that trained subjects score higher agreement among themselves than when there are compared with the untrained subject. On the other hand, some of the trained subjects have failed to meet the 0,61 expected rate. We interpret this as an indication that training can be improved. This conclusion is supported by the fact that all three trained subjects scored correlations in the same significance category (between 0,47 and 0,62). In any event, the results have shown that with a 4 hour training regular developers classify 50% better that junior developers. And senior developers achieve a classification ability that scores substantial agreement in the correlation table. Finally, the comparison of the results with the untrained subject proves that training has impact on the classification ability of the developers. Taking into consideration the inclusion of the untrained subject shown in Table 5, it seems that the organization would do better to improve the efficiency of the training given, rather than achieving 100% training coverage of their developers.

6. Conclusions and future work

In this article, we have proposed a process that links defect causal analysis to the training department. The process enables an organizational training department to show its contribution to the bottom line results of the organization. We have taken the first steps into implementing the process and we have also conducted a verification of these first steps.

It was determined that the key point for the success of the deployment of the process was the causal analysis session. An effort was made to reviewing the current state of the implementation of causal analysis in lower maturity organizations. The result was that we were able to find research that shows implementing causal analysis in lower maturity organizations is possible.

We developed a custom classification scheme for the organization, and cost-effectively modified the defect tracking system to deploy it.

Training in that classification scheme was given to a pilot project. The results show that training improves the developers' ability to classify defects.

In conclusion, causal analysis meetings have been successfully implemented in a maturity level 3 organization. Previous experiences encourage us to affirm that an implementation of the Causal Analysis and Resolution process area can be achieved at a lower maturity organization. We are now developing the following steps to achieve a full implementation of the proposed process.

7. References

1 Salas, E., and Cannon-Bowers, J.A.: 'THE SCIENCE OF TRAINING: A Decade of Progress', Annual Review of Psychology, 2001, 52, pp. 471-499

2 Chrissis, M.B., Konrad, M., and Shrum, S.: 'CMMI : guidelines for process integration and product improvement' Addison-Wesley, 2007.

3 Kirkpatrick, D.L., and Kirkpatrick, J.D.: 'Evaluating Training Programs: The Four Levels' Berrett-Koehler Publishers, 2006.

4 Chillarege, R.: 'Orthogonal Defect Classification-A Concept for In-Process Measurements', 1992

5 Phillips, J.J.: 'Return on Investment in Training and Performance Improvement Programs, Second Edition (Improving Human Performance)' Butterworth-Heinemann, 2003.

6 Kitchenham, B.: 'Procedures for Performing Systematic Reviews', Technical Report TR/SE-0401-ISSN, Keele University, UK, 2004, pp. 1353-7776

7 Buglione, L., and Abran, A.: 'Introducing Root-Cause Analysis and Orthogonal Defect Classification at Lower CMMI Maturity Levels ', 2006

8 Bhandari, I., Halliday, M.J., Chaar, J., Chillarege, R., Jones, K., Atkinson, J.S., Lepori-Costello, C., Jasper, P.Y., Tarver, E.D., Lewis, C.C., and Yonezawa, M.: 'In-process improvement through defect data interpretation', IBM Syst. J, 1994, 33, (1), pp. 182-214

9 Card, D.N.: 'Defect-causal analysis drives down error rates', 1993, 10, (4), pp. 98-99

10 Card, D.N.: 'Managing Software Quality with Defects'. Proc. Proceedings of the 26th International Computer Software and Applications Conference on Prolonging Software Life: Development and Redevelopment, Year pp. 472-474

11 Card, D.N.: 'Learning from Our Mistakes with Defect Causal Analysis', 1998

12 Mizukami, D.: 'Analyzing Defects Can Tell a LOT About a Company'. Proc. SEPG Conference 2007, March 26 - 29, 2007 2007 pp. Pages

13 Leszak, M., Perry, D.E., and Stoll, D.: 'A case study in root cause defect analysis', Proceedings of the 22nd international conference on Software engineering, 2000, pp. 428-437

14 Fredericks, M., and Basili, V.: 'Using Defect Tracking and Analysis to Improve Software Quality ', Crosstalk, 1999

15 Norman, E.F.: 'A Critique of Software Defect Prediction Models', IEEE Transactions on Software Engineering, 1999

16 Bibi, S., Tsoumakas, G., Stamelos, I., and Vlahvas, I.: 'Software Defect Prediction Using Regression via Classification', Computer Systems and Applications, 2006. IEEE International Conference on., 2006, pp. 330-336

17 Jacobs, J.C., van Moll, J.H., Krause, P.J., Kusters, R.J., Trienekens, J.J.M., and Semiconductors, P.: 'Effects of Virtual Development on Product Quality: Exploring Defect Causes', Software Technology and Engineering Practice, 2003. Eleventh Annual International Workshop on, 2003, pp. 6-15

18 Ishikawa, K.: 'Guide to Quality Control' Asian Productivity Organization, 1986.

19 Freimut, B., Denger, C., and Ketterer, M.: 'An industrial case study of implementing and validating defect classification for process improvement and quality management', Software Metrics, 2005. 11th IEEE International Symposium, 2005, pp. 10 pp. %@ 1530-1435

20 Microsoft Solution Framework for CMMI Process Guideance Available at http://msdn2.microsoft.com/en-us/teamsystem/aa718802.aspx

21 Smith, C.: 'Achieving organizational training objectives with short course development', 13th Conference on Software Engineering Education & Training, 2000. Proceedings. , 2000, pp. 32-38

22 Cohen, J.: 'A Coefficient of Agreement for Nominal Scales', Educational and Psychological Measurement, 1960, 20, (1), pp. 37

23 Henningsson, K., and Wohlin, C.: 'Assuring fault classification agreement-an empirical evaluation', Empirical Software Engineering, 2004. ISESE'04. Proceedings. 2004 International Symposium on, 2004, pp. 95-104

24 Fleiss, J.L.: 'Measuring nominal scale agreement among many raters', Psychological Bulletin, 1971, 76, (5), pp. 378-382

25 Landis, J.R., and Koch, G.G.: 'The Measurement of Observer Agreement for Categorical Data', 1977, 33, (1), pp. 159-174