# A Controlled Experiment to Evaluate the Effects of Mindfulness in Software Engineering

Beatriz Bernárdez
beat@us.es

Amador Durán
amador@us.es

José A. Parejo
japarejo@us.es

Antonio Ruiz–Cortés
aruiz@us.es

Department of Computer Languages and Systems
University of Seville, Spain

## ABSTRACT

**Context**. Many reports support the fact that some psycho–social aspects of software engineers are key factors for the quality of the software development process and its resulting products. Based on the experience of some of the authors after more than a year of practising *mindfulness*—a meditation technique aimed to increase clearness of mind and awareness—we guessed that it could be interesting to empirically evaluate whether mindfulness affects positively not only the behaviour but also the professional performance of software engineers.

**Goal**. In this paper, we present a quasi–experiment carried out at the University of Seville to evaluate whether Software Engineering & Information Systems students enhance their conceptual modelling skills after the continued daily practice of mindfulness during four weeks.

**Method**. Students were divided into two groups: one group practised mindfulness, and the other—the control group—were trained in *public speaking*. In order to study the possible cause–and–effect relationship, effectiveness (the rate of model elements correctly identified) and efficiency (the number of model elements correctly identified per unit of time) of the students developing conceptual modelling exercises were measured before and after taking the mind-fulness and public speaking sessions.

**Results**. The experiment results have revealed that the students who practised mindfulness have become more efficient in develop-ing conceptual models than those who attended the public speaking sessions. With respect to effectiveness, some enhancement have been observed, although not as significant as in the case of efficiency.

**Conclusions**. This rising trend in effectiveness suggests that the number of sessions could have been insufficient and that a longer period of sessions could have also enhanced effectiveness significantly.

## Keywords

Mindfulness, Quasi–experiment, Conceptual Modelling

## 1. INTRODUCTION

Software quality remains as one of the most important challenges in Software Engineering (SE). Both software industry and academia have made real efforts to improve software quality and are aware that the main issues are not only technical, but also organizational and psycho–social [15].

Since 1996, the Standish Group has been analysing critical factors in software development, published as the well–known CHAOS reports [31]. Although some factors have to do with project management and technology, many of them are strongly *social* such as user involvement, executive management support, clear statement of requirements, and hard–working, focused staff. In addition, *emotional maturity*—defined as the skill to be self–aware, socially aware, self–managed and to manage relationships—has recently become one of the top–ten critical factors in the last edition of the report [32]. Because of its importance, the Standish Group has developed a program to help software development organizations identify trouble areas and improve the emotional maturity of their employees [33]. In addition, recently published works [1] have positively related some aspects of the personality of software engineers such as extroversion, agreeableness, openness to new ideas, and conscientiousness, with the ability to integrate into a team and also on aspects of code quality.

It thus seems reasonable to educate software engineers in some psycho–social skills that promote the smooth running of the software development process. A possible approach for achieving this goal is *mindfulness* [4], a meditative practice that has influence on some aspects of personality which are relevant for the problem at hand. For example, some works have already recommended the practice of mindfulness to promote the ideal work environment in agile development teams, i.e. daily meetings, cooperation with customers and users (C&Us), and a cordial atmosphere between the members of the development team [36, 18].

### 1.1 Psycho–social aspects in SE

Psycho–social aspects in SE have a very special influence in the Requirements Engineering (RE) phase, since interaction with C&Us is more critical for the project success than in any other phase [7]. It is essential to put oneself in the shoes of C&Us in order to understand their position and the needs to be satisfied by

the software system to be developed. In her book about mindfulness [17], Langer includes a quote of *Analysis Terminable and Interminable* by Sigmund Freud (1937) that, relatively speaking, can be applied to the beginning of software projects when software engineers have to understand the problem domain of the organisation for which the software is developed:

> We know that the first step towards the intellectual mastery of the world in which we live is the discovery of general principles, rules and laws that which bring order into chaos. By such mental operations we simplify the world of phenomena but we cannot avoid falsifying it in doing so, specially when we are dealing with processes of development and change.

More often than not, software developers are not experts in the problem domain and the *falsifying phenomenon* described by Freud is likely to occur. Therefore, software engineers—or more specifically *requirements engineers*—should develop the skills of understanding the problem domain as it actually is, being open–minded, avoiding excessive simplification, and focusing all their attention on eliciting users needs. Furthermore, in our experience in SE education, we have observed that most of our students are very interested in computer–related technology and programming but are hardly motivated to perform not so technical tasks like conceptual modelling, let alone meeting with C&Us for eliciting or validating requirements—partially due to an introverted personality very common in software engineers [6]. They should be provided with techniques to make them feel more comfortable in those stages of the software process when psycho–social skills are, at least, as relevant as the technical ones.

The aim of this paper is to present an experimental study in the University of Seville in order to analyse whether the practice of techniques like mindfulness improve the SE process, specifically the RE phase. Concretely, we have carried out a quasi–experiment in which we have evaluated whether the practice of mindfulness in SE students improves: (i) the quality of conceptual models they develop; and (ii) their conceptual modelling performance. The confirmation of the hypotheses, by means a family of experiments, would lead us to consider the possibility of introducing this discipline as a cross–training for future software engineers and, in the long term, to develop mindfulness–based personal growth programs for SE organizations.

The rest of the paper is organized as follows. In section 2, mindfulness is briefly described and the benefits of its practice in various contexts are discussed. In Section 3, the goals of the experiment, subjects, material, and experimental design are formally described; also, variables are identified and the experimental hypotheses are enunciated. In Section 4, the experiment execution is presented. In Section 5, the descriptive statistics are presented together with data preparation and hypotheses test. In Section 6, the experimental results, their implications in the field, and the threats to experimental validity are analysed. The generalization of the obtained results is also discussed in this section. In the final section, we discuss the conclusions and the future work.

# 2. THE PRACTICE OF MINDFULNESS

*Mindfulness* is an English word used to translate the Pali word *sati*, which means having awareness, attention, and remembering. Mindfulness can be simply defined as "the universal and basic human ability to be aware of the contents of the mind moment–by–moment" [29]. The term *mindfulness* is usually used with two complementary meanings. On the one hand, the practice of *formal* mindfulness, in which a person—or a group of people—draw away to a quiet place for a while (ten minutes at least) to meditate on breath. During the meditation session, the person's intention is to keep the mind calmed using breathing as support. Breathing is the usual meditation support because of its unavoidability. Other common options are focusing on an object using the sense of sight, or in surrounding sounds using the sense of hearing. A formal mindfulness session recipe, based on [22, 29], is summarized in Table 1. On the other hand, *informal* mindfulness consists of transferring this state of consciousness to everyday activities.

| Step | Description |
|---|---|
| 1 | Sit with your back straight in a relaxed way. |
| 2 | If there is limit in time, set an alarm. |
| 3 | Review your body trying to distend tensions. |
| 4 | Perform three slow, deep breaths. |
| 5 | Begin to breathe normally, unpretentious, focusing on breathing. |
| 6 | If thought comes, you experience it and gently let it go. |
| 7 | Enjoy, if it comes, a few seconds of spaciousness and openness in mind. |
| 8 | Back slowly to contact with external reality. |

**Table 1: Recipe to formal mindfulness practice (simplified)**

The practice of *formal* mindfulness leads to a high degree of awareness in daily life, i.e. a greater ability to stay in the present moment rather than rehashing the past or imagining the future. This is due to the ability to observe and abandon thoughts without identifying ourselves with them. Mindfulness helps us see the reality around us clearly and solve problems more efficiently, since it reduces mental wandering while performing tasks.

At a neurological level, these effects of mindfulness are explained by some changes in brain activity. The prefrontal cortex of the human brain have evolved to develop the ability to solve problems. Formerly, those problems used to have a physical nature, like making tools or weapons for haunting. Today, the nature of problems is very heterogeneous, including not only abstract but also social problems like helping other people to reach a consensus [26]. However, a hyperactivity of the prefrontal cortex has the undesired effects of rumination and wandering that, paradoxically, prevent us from having a clear vision of reality and solving problems properly. Notice that the increase of this hyperactivity is one of the consequences of our current relationship with technology, i.e. the ubiquity of Internet–connected devices and the continuous interruptions they generate. Neuroscientists have demonstrated that a continued practice of mindfulness reduce prefrontal cortex hyperactivity and increases the activity of other areas of the brain which are active when concrete tasks are performed [29].

## 2.1 Evidences of mindfulness benefits

There are many psychological studies reporting the benefits of mindfulness. For example, in [25] a qualitative study examined the influence of mindfulness in a 15–week course with graduate students. Participants reported an increase of their mental clarity, organization, awareness, and acceptance of emotions and personal issues. In [19], a controlled experiment based on Graduate Record Examinations (GRE) and assessing verbal, quantitative and analytical skills to measure reading comprehension, concentration, level of mind wandering, and working memory is described. The outcomes showed a great improvement in the group of people that

attended 15 mindfulness sessions. In [23], neuro–psychologists studied the effects of mindfulness in 136 heterogeneous patients showing that, after two months of daily 20–minute practice, a significant percentage experienced better personal well–being in terms of mental clarity, equanimity, wisdom and self–compassion based on standard health surveys (questionnaires). Mindfulness has also been successfully used as therapeutic treatment to individuals prone to anxiety, stressed people, etc. [9, 12, 28]. In [8], a careful literature review compiles the main benefits of mindfulness at a personal level: self–control, self–regard, equanimity, self–awareness, self–insight, intuition, regulation emotion, and well—being in general.

## 2.2 Mindfulness in the workplace

With respect to social relations, the main reported benefits of mindfulness are related to empathy, assertiveness, emotion regulation, decreased reactivity, increased response flexibility, counselling skills, and emotional intelligence in general [8]. Benefits of mindfulness in labour relations, especially in stressful working areas like health or teaching, have also been reported [21].

In the context of the software industry in Silicon Valley, the practice of mindfulness is fostered arguing improvements in Employee Relationship Management—reacting less emotionally, for example. It has also been observed that mindfulness improves memory and executive functions, and increases the ability to concentrate on fast–changing stimuli [27]. Particularly in Google, engineer Chade–Men Tan [35] is developing a training program based on mindfulness, using it to understand their co–workers' motivations, to enhance their creativity and productivity, and to cultivate emotional intelligence.

In the Management Science and Organisational Learning fields, there are some proposals of the so–called *organisational mindfulness*, which defines a set of principles like including deference to expertise, preoccupation with failure, reluctance to simplify interpretations, sensitivity to operations, and commitment to resilience [37]. Those principles have a positive influence in the capability to change direction, react in a flexible manner, improve reliability in control systems or increase innovation in information technology [24, 34]. When applied to RE, organisational mindfulness can help change wrong behaviour patterns. For example, in [24] a case study is presented in which RE tasks are defined, tasks inefficiencies are identified, and their causes are analysed using the aforementioned organisational mindfulness principles. Using the results of the analysis, the RE tasks are redefined and, as a result, inefficiencies decrease and become under control.

# 3. EXPERIMENT PLANNING

According to the SE experimental taxonomy presented in [30], the experiment described in this paper belongs to the *Requirements Analysis* and *Document Comprehension* categories since its main objective is to ascertain if the practice of mindfulness improves the capability of students in the performance of a concrete RE task: analysing an interview transcript and developing the corresponding conceptual model. Hereafter, the experimental process is described using the guidelines proposed by Jedlitschka *et al.* [13] with minor changes.

## 3.1 Goals

Based on the recommendations proposed in [38] about using the GQM template in experimentation in SE, the goal definition of the experiment can be summarized as:

**Analyse** the practice of mindfulness

**for the purpose of** evaluating its effects

**with respect to** the performance of students developing conceptual models, and the quality of the developed models

**from the point of view of** the experimenters

**in the context of** sophomore students of the degree on Software Engineering at the University of Seville.

## 3.2 Participants

Since we aim to study the effects of mindfulness on early stages of the software process, subjects are sophomores enrolled in the *Introduction to Software Engineering and Information Systems* (ISEIS) annual course. Specifically, 32 students (30 male, 2 female) participated in the experiment during the first semester of the 2013–2014 academic year. In order to increase the motivation, the students were offered extra points for taking part in the experiment.

The subjects were divided into two groups—the *mindfulness* group and the *public speaking* group—depending on the motivations and preferences expressed by the students in an interest questionnaire. Consequently, groups are unbalanced and the study is therefore a *quasi–experiment*, due to the absence of random assignment. Although this assignment mechanism has several disadvantages [11, 14], the experimenters considered individuals' motivation essential to avoid mortality and ensure that the treatment is actually applied, something crucial for the validity of the experiment.

## 3.3 Experimental material

The material used in the conduction of the experiment (available at `http://www.lsi.us.es/~beat/ExpMater/mindfulness`) comprises of the following items:

- A questionnaire with three sections: (i) questions regarding the importance of quality in software, and the relationship between the quality of the software product and that of the process leading to its creation; (ii) questions about the interest on participating in the research; and (iii) questions about the choice of mindfulness or public speaking.

- Two exercises of conceptual modelling. The statements of the exercises were presented as an interview transcript, i.e. questions and answers (Q&A) between a requirements engineer and a customer about the problem domain and the expected system behaviour.

  The first exercise is about an hypothetical information system for ERASMUS grants management, whereas the second one is about an hypothetical information system for *End of Degree Projects* (EODP) management. Both problem domains were chosen because of students' familiarity and interest, since all of them are potential candidates for an ERASMUS grant and have to develop an EODP.

## 3.4 Tasks

During the experiment, the subjects performed the following tasks (see Table 2): (i) filling out the interest questionnaire after a brief introduction to the ongoing research and mindfulness; (ii) doing the first conceptual modelling exercise (ERASMUS); (iii) attending the sessions on the cross–training topics of their interest, along with the ISEIS lessons, during four weeks, four days a week; (iv) doing the second conceptual modelling exercise (EODP).

## 3.5 Variables and parameters

The independent variables or factors that could have had an impact on the results are:

| Introductory presentation & Questionnaire of interest (30min) | | |
|---|---|---|
| ⇓ | | ⇓ |
| **Mindfulness group** | | **Public Speaking group** |
| Seminar about mindfulness (1h) | | Seminar about public speaking (1h) |
| ERASMUS conceptual modelling exercise (2h) | | |
| 4 times/week × 4 weeks × 10 minute mindfulness sessions | ISEIS scheduled lessons | 4 times/week × 4 weeks × 10 minute public speaking sessions |
| EODP conceptual modelling exercise (1.5h) | | |

**Table 2: Experiment schedule**

**Cross–Training** (training in complementary sessions): the main independent variable of the experiment is the specific topic taught during a set of complimentary sessions performed during four weeks. This factor has two levels, the treatment is *mindfulness* and the control condition is *public speaking*, i.e. the sessions on public speaking are used as an placebo, thus the experimenters implicitly assume they have no impact on the response variables. According to [13], this variable is considered as a resource, since it enhances the necessary skills for the performance of personal and professional roles of the students and it is in a nominal scale.

**Pre/Post–Exercise** (exercises done before and after cross–training sessions): this variable represents the conceptual modelling exercises that students have done. It has two levels: ERASMUS and EODP.

Since students are also learning SE during these 4 weeks, an improvement in both groups should be observed. However, we intend to analyse whether the improvement of students in the mindfulness group is significantly different from those in public speaking or not. According to [13], this is a material variable and it is measured in nominal scale.

In order to define dependent variables, an indirect measure called *semantic expressiveness* (SEMEX) was used. The semantic expressiveness measures to what extent—compared to a *canonical* conceptual model developed by the experimenters—the conceptual models developed by experimental subjects are able to represent all the problem domain scenarios described in the interview transcripts. Model elements are considered as *correctly identified* if there exist identical or semantically equivalent elements in the canonical model. The formula for computing this internal variable, measured in a ratio scale, is the following:

$$\text{SEMEX} = \text{CLASS}_{\text{OK}} - \frac{\text{CLASS}_{\text{KO}}}{2} + \text{ASSOC}_{\text{OK}} + \text{ATTR}_{\text{OK}}$$

where $\text{CLASS}_{\text{OK}}$ is the number of classes correctly identified; $\text{CLASS}_{\text{KO}}$ is the number of classes incorrectly identified; and $\text{ASSOC}_{\text{OK}}$ and $\text{ATTR}_{\text{OK}}$ the number of correctly identified associations and attributes respectively.

Taking SEMEX into account, the following dependent variables were considered:

**Effectiveness:** the percentage of semantic expressiveness achieved by a subject measured in ratio scale:

$$\text{EFFECTIVENESS} = \frac{\text{SEMEX}}{\text{CLASS}_C + \text{ASSOC}_C + \text{ATTR}_C}$$

where $\text{CLASS}_C$, $\text{ASSOC}_C$, $\text{ATTR}_C$ are the number of classes, associations, and attributes in the canonical solution previously agreed by the experimenters, i.e. their sum is the total semantic expressiveness.

**Efficiency:** this variable, following the recommendations in [16], is defined as the ratio between the achieved semantic expressiveness and the time spent by a subject in doing a conceptual modelling exercise:

$$\text{EFFICIENCY} = \frac{\text{SEMEX}}{\text{TIME}}$$

where TIME is the time spent in doing the conceptual modelling exercise measured in minutes.

Following [14], some parameters were identified: the lack of experience of the subjects in real projects, the similar size of the conceptual modelling exercises, the problem domain of the conceptual modelling exercises related to academic topics, and the number of sessions on each cross–training topic.

## 3.6 Hypotheses

Two groups of hypotheses, one for each dependent variable, were tested in the experiment.

**Effectiveness hypotheses**

$H_{0,1}$ **:** there is no difference in the effectiveness of subjects in the ERASMUS and the EODP exercises, i.e. neither the teaching of ISEIS nor the cross–training sessions have a significant effect on the effectiveness of subjects. // $H_{1,1} : \neg H_{0,1}$

$H_{0,2}$ **:** there is no difference in effectiveness between the subjects practising mindfulness and those subjects attending public speaking sessions—the experimenters assumed that the differences observed between the effectiveness for the ERASMUS and the EODP exercises are due to the teaching of ISEIS exclusively). // $H_{1,2} : \neg H_{0,2}$

**Efficiency hypotheses**

$H_{0,3}$ **:** there is no difference in the efficiency of subjects in the ERASMUS and the EODP exercises, i.e. neither the teaching of ISEIS nor the cross–training sessions have a significant effect on the efficiency of subjects. // $H_{1,3} : \neg H_{0,3}$

$H_{0,4}$ **:** there is no difference in efficiency between the subjects practising mindfulness and those subjects attending public speaking sessions—the experimenters assumed that the differences observed between the efficiency for the ERASMUS and the EODP exercises are due to the teaching of ISEIS exclusively). // $H_{1,4} : \neg H_{0,4}$

## 3.7 Design

In SE, experimentation is usually aimed at testing the goodness of methods, tools, or technologies, so the usual experimental designs are similar to those presented in [14]. When repeated measures are taken, the reason is usually either improving the precision by making comparisons within matched pairs of the experiment material [38], or maximizing data collection [2].

Nevertheless, for the experiment presented in this paper, which includes personal development aspects into empirical SE, it seems reasonable to use a design that allows the measure of the dependent variables before and after the practice of mindfulness. Considering the hypotheses, the experimental design should take into account two factors with two levels: the cross–training received

| | | Cross–Training | |
|---|---|---|---|
| | | **Mindfulness** | **Public Speaking** |
| **Pre/Post Exercise** | Pre–Exercise (ERASMUS) | $G_1$ | $G_2$ |
| | Post–Exercise (EODP) | $G_1$ | $G_2$ |

**Table 3: Experimental design — 2×2 mixed factorial**

| | | Interest | Pre | Attend. (mean) | Post | Sample |
|---|---|---|---|---|---|---|
| **Number of subjects** | $G_1$ | 38 | 18 | 23 | 17 | 16 |
| | $G_2$ | 37 | 24 | 10 | 18 | 16 |

**Table 4: Number of subjects during execution**

(a *between–subjects* factor) and the conceptual modelling exercises (a *within–subjects* factor). Since dedicating two ISEIS classes for the pre and post conceptual modelling exercises was affordable, a quasi–experiment with the 2×2 mixed factorial design from [5] was selected (see Table 3). In this design, commonly used in the fields of psychology and medicine when the evolution of patients under a given treatment (or therapy) after a certain amount of time needs to be studied [19, 21, 25], each subject is assigned to one single treatment (Cross–Training level) and two repeated measures on the response variables are taken.

# 4. EXECUTION

First of all, the ongoing research and a brief introduction to mindfulness were presented to 87 students during their ordinary class schedule. After that—during the same class—the interest questionnaire described in Section 3.3 was given to the students, who filled it out manually. Once the answers were analysed by the experimenters, the students were divided into two groups according to their preferences, except 12 students who showed no interest in participating and were therefore kept aside from the experiment.

Prior to the day of the experiment execution, each group was given an introductory seminar about the corresponding topic, i.e. mindfulness for $G_1$ and public speaking for $G_2$. The seminars took about one hour and were delivered out of ordinary class schedule. Students' attendance was controlled by a sign–in sheet.

Next, the two groups of students did the pre–exercise (ERASMUS) the same day during class schedule. The aim of the exercise was reading an interview transcript carefully and developing a conceptual model identifying classes, associations and attributes.

Then, during four weeks, each group of students attended complementary sessions of the corresponding topic four days a week. The sessions took place always in the same conditions, i.e. same classroom and same hour for each group.

Ultimately, the two groups of students did the post–exercise (EODP) the same day during class schedule. Each group did the exercise in different classrooms, always monitored by one experimenter. In the $G_1$ group, the initial 15 minutes were dedicated to a guided formal meditation in group, similar to those performed during the previous four weeks. After the 15–minute meditation, the students did the exercise individually. During the exercise, the two experimenters monitoring the exercise answered questions from the students that were related to the notation only (UML), trying to avoid any discussion about the interpretation of the interview transcript in order to avoid any influence on the outcome variables.

Once the students' participation was finished, both exercises were blindly checked and the obtained data was input into Excel and then imported into SPSS 22. For each subject, the following data were collected: answers to the interest questionnaire, number of sessions in the cross–training, and semantic expressiveness, effectiveness, time, and efficiency for both exercises.

Table 4 shows the number of students who participated in each phase of the experiment. The interest column shows the number of subjects who showed interest in mindfulness and public speaking respectively. The attendance column shows the average number of subjects who attended the cross–training sessions during four weeks. The pre and post columns show the number of subjects who did the pre and post–exercises. The sample column indicates the final number of subjects who where finally considered during analysis and evaluation of the experiment. The difference between the sample size and the number of participating subjects is motivated by the following deviations:

- In the $G_1$ group, 5 subjects attended the mindfulness sessions without having done the pre–exercise. Only 1 out of the 17 subjects who did the post–exercise did not do the pre–exercise. As a result, only the 16 subjects who did both exercises were included in the sample.
- Applying the same criteria in the $G_2$ group, only 16 subjects did both exercises and were included in the sample.

# 5. ANALYSIS AND EVALUATION

A summary of the descriptive statistics of the experiment in displayed in Table 5, showing the mean and standard deviation of the effectiveness and efficiency of the pre and post exercises. Figures 1 and 2 depict the dispersion of the same dependent variables.

| | | Mean ($\mu$) | | Std. Dev. ($\sigma$) | |
|---|---|---|---|---|---|
| | | **Pre–Ex.** | **Post–Ex.** | **Pre–Ex.** | **Post–Ex.** |
| **Effective–ness** | $G_1$ | 0.395 | 0.673 | 0.183 | 0.131 |
| | $G_2$ | 0.416 | 0.611 | 0.084 | 0.092 |
| **Efficiency** | $G_1$ | 0.172 | 0.502 | 0.075 | 0.177 |
| | $G_2$ | 0.188 | 0.365 | 0.042 | 0.108 |

**Table 5: Descriptive statistics of the sample with 32 subjects**

With respect to the potential outliers depicted in Figures 1 and 2, the low effectiveness of subject 8 was scrutinized finding that, although some textual notes were developed by the subject, they could not be considered as a valid conceptual model and therefore no valid conceptual element could be identified, i.e. SEMEX $= 0$ in the pre–exercise, so the subject was left out of the data analysis.

The very high efficiency of subjects 10 and 11 was also investigated, but no other cause than a high performance was found. As a consequence, they were considered as genuine values since they showed an outstanding performance before and after the cross–training sessions. Regarding subject 31, although he showed a high efficiency only in the pre–exercise, no reason was found to discard him. However, the statistical tests of the hypothesis shown in Sections 5.1 and 5.2 have been performed on both samples, i.e. including and excluding subjects 10 and 11, and the results regarding hypotheses acceptance or rejection were not affected when using $\alpha = 0.01$.

Then, due to the lack of random assignment to groups, the differences between $G_1$ and $G_2$ were examined before starting the cross–training sessions. In order to check the hypothesis of similarity of groups in terms of the values of the dependent variables [5], one–way analysis of variance (ANOVA) was conducted on each of the

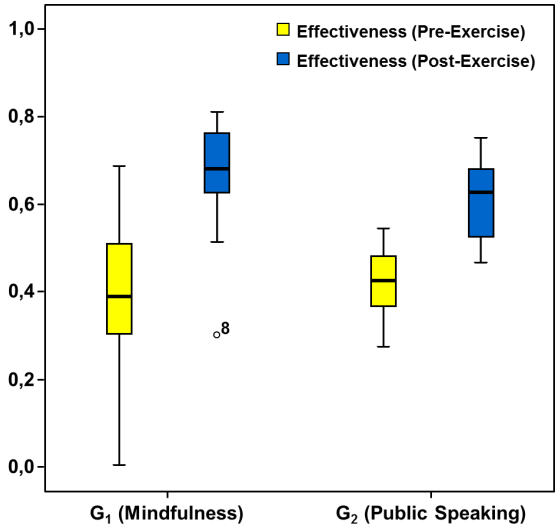


**Figure 1: Box plot of effectiveness**

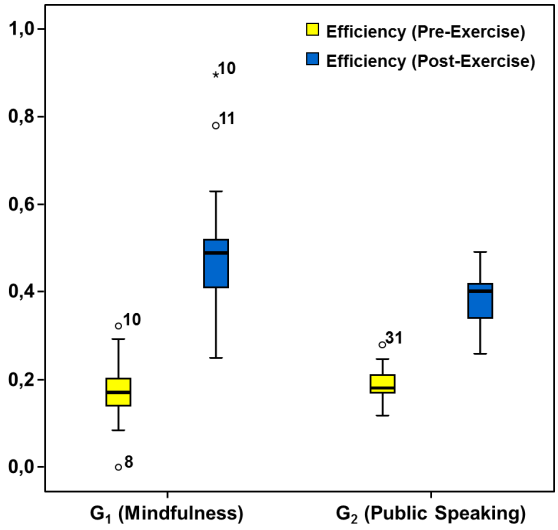


**Figure 2: Box plot of efficiency**

outcome measures. Since the null hypotheses were not rejected (effectiveness(p–value) = 0,896 and efficiency(p–value) = 0,816), there is no evidence of significant differences between groups, which mitigates to some extent the problems and risks caused by the lack of random assignment [5].

## 5.1 Testing effectiveness hypotheses

The purpose of these hypotheses is to determine how the independent variables have influenced on effectiveness, i.e.:

- Whether the ISEIS lessons have influenced on effectiveness.
- Whether the interaction between the cross–training and the ISEIS lessons has influenced on effectiveness.

The means of effectiveness obtained for each level of the independent variables are shown in figure 3. For both cross–training levels, the mean for effectiveness after treatment is higher than before. This is not surprising, since students' knowledge of SE has been improved due to the ISEIS lessons, having an impact on effectiveness. It is also noticeable that the line showing the improvement of the mindfulness group has a steeper slope than the line of the public speaking group.

To test the effectiveness hypotheses, it was first evaluated whether the data followed a normal distribution or not (see Shapiro–Wilk results in table 6).

Given that all the groups were normal at the level of 0.05, a mixed–model ANOVA was conducted, with Cross–Training as a *between–subjects* factor and Pre/Post–Exercise as a *within–subjects* factor. In table 7 the results obtained by means of ANOVA for effectiveness are shown, in which the last column represents the partial eta–squared ($\eta_p^2$) as an indicator of the effect size for groups. The first row of the table shows the data for the Pre/Post–Exercise independent variable of the experiment; the second row shows the data for the interaction between the aforementioned variable and Cross–Training; finally, the error is shown in the last row. The analysis summarised in table 7 revealed a significant effect ($\alpha = 0.01$) for Pre/Post–Exercise, i.e. both groups have a statistically better effectiveness after training than before training, measured by the percentage of semantic expressiveness achieved. However, the interaction between Pre/Post–Exercise and Cross–Training is not significant at the 0.01 level. This means that, although the mean of effectiveness varied significantly for both exercises, the effect of the Cross–Training is not necessarily linked to these differences. Therefore, the null hypothesis $H_{01}$ is rejected at the $\alpha = 0.01$ significance level, but there is not enough evidence as to reject $H_{02}$ at such significance level. However, this same analysis for effective-

| Dependent variable | Group | Shapiro–Wilk significance level |
|---|---|---|
| **Effectiveness** | $G_1$ (pre) | 0.830 |
| | $G_2$ (pre) | 0.709 |
| | $G_1$ (post) | 0.377 |
| | $G_2$ (post) | 0.405 |

**Table 6: Shapiro–Wilk normality test results to effectiveness**

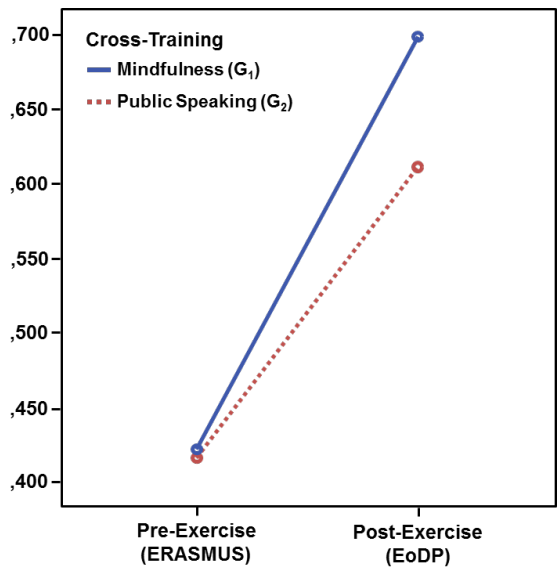


**Figure 3: Profile plot of mean of effectiveness**

| Source of variation | Type III Sum of Squares | Degrees of Freedom | Mean Square | F–ratio | Significance | $\eta_p^2$ |
|---|---|---|---|---|---|---|
| Pre/Post–Exercise | 0.860 | 1 | 0.860 | 90.253 | 0.000 | 0.757 |
| Pre/Post–Exercise * Cross–training | 0.026 | 1 | 0.026 | 2.713 | 0.110 | 0.086 |
| Error(Pre/Post–Exercise) | 0.276 | 29 | 0.010 | | | |

**Table 7: ANOVA of effectiveness**

ness was repeated removing outliers 10 and 11. The p–value for the interaction between independent variables in this case was 0.048, i.e. the effectiveness of students attending ISEIS and practising mindfulness is better that those that just attended ISEIS and public speaking sessions. According to this result, $H_{02}$ could be rejected at the $\alpha = 0.05$ significance level. We do not claim the existence of a significant difference on the effectiveness due to mindfulness, since the p–value is almost in the limit of statistical significance and the removal of outliers 10 and 11 could be objected. However, the results raise the suspicion about the existence of such a difference in practice. Furthermore, since the number of mindfulness sessions could be too low as to show a significant effect at $\alpha = 0.01$, we consider that it is worthy to replicate the experiment in order to discern if a cause–and–effect relationship actually exists between the practice of mindfulness and the improvement of effectiveness.

## 5.2 Testing efficiency hypotheses

The purpose of these hypotheses is to determine how the independent variables have influenced on efficiency, i.e.:

- Whether the ISEIS lessons have influenced on efficiency.
- Whether the interaction between the cross–training and the ISEIS lessons have influenced on efficiency.

The means of efficiency obtained for each level of the independent variables are shown in figure 4. As for effectiveness, the mean for efficiency after treatment is higher than before for both cross–training levels. In order to test the efficiency hypotheses, it was firstly evaluated whether the data followed a normal distribution or not (see Shapiro–Wilk results in table 8).

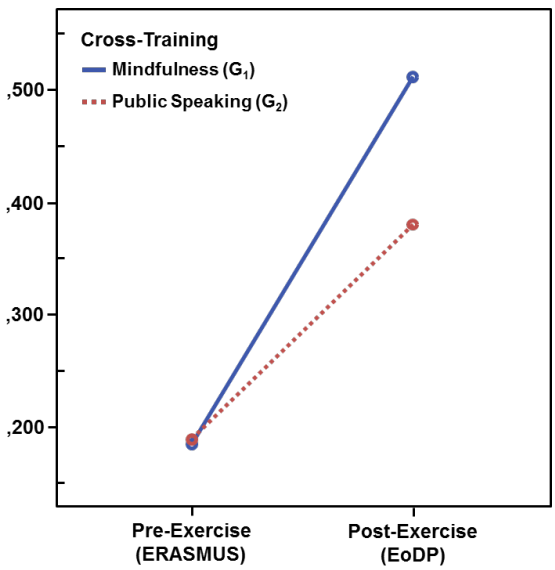


**Figure 4: Profile plot of mean of efficiency**

| Dependent variable | Group | Shapiro–Wilk significance level |
|---|---|---|
| **Efficiency** | $G_1$ (pre) | 0.275 |
| | $G_2$ (pre) | 0.560 |
| | $G_1$ (post) | 0.039 |
| | $G_2$ (post) | 0.243 |

**Table 8: Shapiro–Wilk normality test results to efficiency**

Although the data of $G_1$ (post) was not normal at the level of 0.05, it was decided to carry out a mixed–model ANOVA in a similar way as with the effectiveness hypotheses tests, since most of the data was normal and the statistical test is robust [10]. The results obtained by means of ANOVA for effectiveness are shown in Table 9. This analysis revealed a significant effect ($\alpha = 0.01$) for Pre/Post–Exercise, i.e. both groups had a statistically better efficiency after training than before training, as measured by the semantic expressiveness achieved per unit of time. At this significance level ($\alpha = 0.01$), the interaction between Pre/Post–Exercise and Cross–Training is also significant. This means that, although the mean efficiency varied significantly for both exercises in both groups, the practice of mindfulness is linked to a higher improvement in the efficiency. Therefore, the null hypotheses $H_{03}$ and $H_{04}$ are rejected at the $\alpha = 0.01$ significance level. In a very similar way to effectiveness, we repeated the analysis of the efficiency without taking into account outliers 10 and 11. In this case, the results were essentially the same, i.e. the p–value for the interaction of independent variables was 0.001, thus $H_{04}$ was rejected at the $\alpha = 0.01$ significance level.

# 6. DISCUSSION

The main finding of the experiment described in this paper is that the practice of mindfulness make SE students capable of achieving similar results in less time—they become more *efficient*—in the conceptual modelling of a problem domain described by an interview transcript. Compared to this, the observed mindfulness effect on effectiveness has been relatively small and not enough evidences exist to reject the null hypothesis. This could be because (i) either there is no casual relationship between mindfulness and students' effectiveness in conceptual modelling; or (ii) because the number of mindfulness sessions was insufficient to produce the expected effect. We think the latter is more plausible since most reports indicate that is after two months of mindfulness practice as a therapy when their effects begin to be perceived. Nevertheless, the main experiment that have influenced our work [19], used only 15 sessions of mindfulness as a way of enhancing some cognitive skills but not as a therapy. For this reason, we think that is necessary to adjust the number of mindfulness sessions in future replications of the experiment. Regarding the type of task chosen for the experiment, we think that conceptual modelling was a better choice than reading comprehension [30]. The reason is that mindfulness is supposed to increase abstraction as a result of a better mental clarity, and this is an essential skill for conceptual modelling.

| Source of variation | Type III Sum of Squares | Degrees of Freedom | Mean Square | F–ratio | Significance | $\eta_p^2$ |
|---|---|---|---|---|---|---|
| Pre/Post–Exercise | 1.047 | 1 | 1.047 | 247.515 | 0.000 | 0.895 |
| Pre/Post–Exercise * Cross–Training | 0.072 | 1 | 0.072 | 17.001 | 0.000 | 0.370 |
| Error(Pre/Post–Exercise) | 0.123 | 29 | 0.004 | | | |

**Table 9: ANOVA of efficiency**

With respect to the possibilities of replicating this experiment in other universities, the main problem is finding a person committed to guide the mindfulness sessions. Other authors like [25] use CDs with recorded meditation guides for individual use, but we think that young students are not likely to keep on practising mindfulness individually without the feedback of a group. Other possibilities, like practising mindfulness in group but distributively via videoconferencing, have not been explored yet.

## 6.1 Threats to validity

In [38], a thorough compilation of threats to the validity of the results of empirical studies is presented. In this section, the threats related to the *conclusion*, *internal*, *construct*, and *external* validities are analysed and the actions performed to reduce them are described.

**Conclusion validity:** this validity is concerned with the statistical relationship between the treatment and the outcome. In the experiment described in this paper, the main threat to conclusion validity was the small size of the sample. Nevertheless, although small, the sample size was acceptable for the statistical tests applied [14], and all the assumptions of each test were verified before their application. In addition, those subjects who missed one of the exercises were excluded from the sample, in order to apply the tests properly.

**Internal validity:** this validity is concerned with the treatment—and no other uncontrolled or unmeasured factor—being the cause of the outcome. The following threats described in [38] were identified:

**History:** since both groups performed the experiment tasks simultaneously, without any significant incident, this threat was neutralised.

**Maturation:** both groups maturated simultaneously with respect to their knowledge in SE, since they all attended the same number of ISEIS sessions between the pre and post exercises. Therefore, this threat was also neutralised.

**Mortality:** in order to reduce this threat, the students were offered extra points for participating in the experiment. Furthermore, no subjects abandoned the mindfulness sessions due to tedium or lack of interest.

**Selection & assignment bias:** the main threats to the internal validity of the experiment were *selection* and *assignment bias*, since it was a quasi–experiment. However, there are no evidences of significant differences in the groups prior to treatment, as verified by means of a one–way ANOVA test on the measures of the pre–exercise dependent variables (see Section 5 for details).

On the other hand, motivation was also similar in both groups because the subjects participated in the experiment on their own choice, they freely chose the cross–training topic, and the pre and post exercises were relevant for their marks in ISEIS.

However, given the shown implication of students in this experiment, we plan to perform random assignment in future replications.

**Construct validity:** this validity is concerned with the relation between theory and observation. Since the experiment includes two cross–training topics as treatments, and two exercises, the *mono–operation bias* was reduced because of the cause construct being completely represented. The *mono–method bias* was also reduced by the analysis of the effectiveness and efficiency dependent variables.

**External validity:** the greater the external validity, the more the results of an empirical study can be generalized to current SE practice. Two threats were identified which can limit such generalization:

**Materials used:** the size of the interview transcripts might not be representative of industrial problems, but it was appropriate for the available time for the pre and post exercises. However, we think that the intellectual processes applied during conceptual modelling—and potentially improved by mindfulness—are basically the same regardless of the size of the problem at hand.

**Subjects:** since the tasks performed during the pre and post exercises did not require high levels of industrial experience, using students as subjects instead of SE professionals could be considered as appropriate [20]. Moreover, students are the next generation of professionals, so they are close to the population under study [16].

# 7. CONCLUSIONS AND FUTURE WORK

In this paper, the results of an experiment to analyse the effects of mindfulness on the development of conceptual models are presented. Previously, the main psycho–social factors that influence the SE process were reviewed, showing the connections with the alleged benefits of mindfulness.

In our opinion, conceptual modelling is a difficult task for our students. We think that the proven benefits of mindfulness, i.e. mental clarity, reading comprehension, concentration, etc., could be a relevant support for this task. The experiment results revealed that students who practised mindfulness obtained similar results in less time than the others, i.e. a significant improvement in efficiency was observed. However, the observed improvement in effectiveness, although important, was not statistically significant.

If the results of this experiment were confirmed by replication, the improvement in efficiency would be quite interesting for SE organizations. A key factor for the adoption of mindfulness in industry is persuasion, i.e. to convince employees of the goodness of its practice. Once convinced, it is necessary to facilitate continuity, an appropriate atmosphere in the organisation, software guides for meditation (a mobile *app* for example), providing follow–up meetings, etc.

A future work in the short term is replicating the experiment increasing the number of cross–training sessions and using random assignment. Another option is conducting a distributed quasi–experiment, in which an experiment is run across multiple sites, organised as a single study rather than as a set of replications [3]. Then, changing the experimental object, it would be possible to study the effects of mindfulness on the quality and performance of the subjects during other SE activities, such as elicitation and negotiation of requirements, or resolution of technological problems. Furthermore, we would like to examine whether mindfulness improves team–work among introverted software engineers.

Once the strengths of the practice of mindfulness have been identified, we would like to perform case studies in real companies in order to develop a mindfulness–based personal growth program for its further adoption.

## 8. ACKNOWLEDGMENTS

The authors would like to thank the ISEIS students who participated in the experiment for their excellent attitude. We would also like to thank Dr. M. Genero, from the University of Castilla–La Mancha, for being always willing to answer our experimental *uncertainties*. We also thank the staff of the Escuela Técnica Superior de Ingeniería Informática of the University of Seville for arranging anything we needed in the classrooms for the mindfulness and public speaking sessions. Finally, we thank E. Pérez Infantes and J. Murga for guiding some of the mindfulness sessions.

This work was partially supported by the EU Commission (FEDER), the Spanish and Andalusian R&D&I programme grants TAPAS (TIN2012–32273), COPAS (P12–TIC–1867), and THEOS (TIC–5906).

## 9. REFERENCES

[1] S. T. Acuña, M. Gómez, and N. Juristo. How do personality, team processes and task characteristics relate to job satisfaction and software quality? *Information and Software Technology*, 51(3):627–639, 2009.

[2] L. Briand, Y. Labiche, and R. Madrazo-Rivera. An experimental evaluation of the impact of system sequence diagrams and system operation contracts on the quality of the domain model. In *Empirical Software Engineering and Measurement (ESEM), 2011 International Symposium on*, pages 157–166. IEEE, 2011.

[3] D. Budgen, B. Kitchenham, S. Charters, S. Gibbs, A. Pohthong, J. Keung, and P. Brereton. Lessons from conducting a distributed quasi-experiment. In *Empirical Software Engineering and Measurement, 2013 ACM/IEEE International Symposium on*, pages 143–152. IEEE, 2013.

[4] B. S. Butler and P. H. Gray. Reliability, mindfulness, and information systems. *MIS Quarterly*, 30(2):211–224, 2006.

[5] D. T. Campbell and S. Julian. *Experimental and Quasi–Experimental Designs for Research*. Wadsworth, United States, 1963.

[6] L. F. Capretz. Personality types in software engineering. *International Journal of Human–Computer Studies*, 58(2):207–214, 2003.

[7] A. Davis. *201 principles of software development*. McGraw–Hil, New York, 1995.

[8] D. M. Davis and J. A. Hayes. What are the benefits of mindfulness? a practice review of psychotherapy-related research. *Psychotherapy*, 48(2):198, 2011.

[9] C. K. Germer, R. D. Siegel, and P. R. Funton. *Mindfulness and Psychotherapy*. The Guilford Press, New York, 2013.

[10] G. Glass, P. D. Peckham, and J. R. Sanders. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, 42(1):237–288, 1972.

[11] J. A. Gliner, G. A. Morgan, and N. L. Leech. *Research methods in applied settings: an integrated approach to design and analysis (second edition)*. Tylor and Francis Group, 2009.

[12] P. Grossman, L. Niemann, S. Schmidt, and H. Walach. Mindfulness–based stress reduction and health benefits: A meta–analysis. *Journal of Psychosomatic Research*, 57(1):35–43, 2004.

[13] A. Jedlitschka, M. Ciolkowski, and P. D. *Guide to Advanced Empirical Software Engineering*, chapter Reporting experiments in Software Engineering, pages 201–228. Springer Verlag, 2008. Eds. Forrest Shull and Janice Singer and Dag I.K. Sjøberg.

[14] N. Juristo and A. M. Moreno. *Basics of Software Engineering Experimentation*. Kluwer Academic Publishers, 2001.

[15] T. Kilamo, I. Hammouda, and M. Chatti. Teaching collaborative software development: A case study. In *34th International Conference on Software Engineering (ICSE) Proceedings*, pages 1165–1174. ACM, June 2012.

[16] B. A. Kitchenham, S. L. Pfleeger, D. Hoaglin., K. E. Emam, and J. Rosenberg. Preliminary Guidelines for Empirical Research in Software Engineering. *IEEE Transactions on Software Engineering*, 28(8):721–734, August 2002.

[17] E. J. Langer. *Mindfulness*. Addison-Wesley, Massachusetts, 1989.

[18] S. Matook and K. Kautz. Mindfulness and agile software development. In *19th Australasian Conference on Information Systems (ACIS) Proceedings*, pages 638–647. Association for Information Systems/University of Canterbury, December 2008.

[19] M. D. Mrazek, M. S. Franklin, D. T. Phillips, B. Baird, and J. W. Schooler. Mindfulness training improves working memory capacity and gre performance while reducing mind wandering. *Psychological Science*, 24(5):776–781, May 2013.

[20] A. A. Porter, L. G. Votta, and V. R. Basili. Building Knowledge through Families of Experiments. *IEEE Transactions on Software Engineering*, 25(4):456–473, July 1999.

[21] P. A. Poulin, C. S. Mackenzie, G. Soloway, and E. Karayolas. Mindfulness training as an evidenced-based approach to reducing stress and promoting well-being among human services professionals. *International Journal of Health Promotion and Education*, 46(2):72–80, 2008.

[22] A. Puddicombe. *Get some headspace*. Hodder and Stoughton, USA, 2011.

[23] D. Riebel, J. Greeson, G. Brainard, and S. Rosenzweig. Mindfulness–based stress reduction and health–related quality of life in a heterogeneous patient population. *Jefferson Myrna Brind Center of Integration Medicine Faculty*, 2(1):1–20, July 2001.

[24] D. Sammon, T. Nagle, and J. McAvoy. Analysing ISD performance using narrative networks, routines and mindfulness. *Information and Software Technology*, 2014.

[25] M. B. Schure, J. Christopher, and S. Christopher. Mind–body medicine and the art of self-care: Teaching mindfulness to counseling students through yoga, meditation, and qigong.

*Journal of Counseling and Development*, 86(1):47–56, Winter 2008.

[26] M. Seligman. *Flourish: A Visionary New Understanding of Happiness and Well–being*. Free Press, New York, 2012.

[27] N. Shachtman. *Enlightenment engineers: meditation and mindfulness in Silicon Valley*. WIRED Magazine, 2013.

[28] S. L. Shapiro, J. A. Astin, S. R. Bishop, and M. Cordova. Mindfulness-based stress reduction for health care professionals: results from a randomized trial. *International Journal of Stress Management*, 12(2):164–176, 2005.

[29] V. Simón. *Aware and Awake*. Descleé De Brouwer, Spain, 2013.

[30] D. I. Sjøberg, J. E. Hannay, O. Hansen, V. B. Kampenes, A. Karahasanovic, N.-K. Liborg, and A. C. Rekdal. A survey of controlled experiments in software engineering. *Software Engineering, IEEE Transactions on*, 31(9):733–753, 2005.

[31] The Standish Group. *The CHAOS Report*, 2008.

[32] The Standish Group. *CHAOS Manifesto 2013: Think Big, Act Small*, 2013.

[33] The Standish Group. *Project Emotional Maturity Test Kit*, 2013. Available at `http://blog.standishgroup.com/project-management/36-pm/381-emtk`.

[34] E. B. Swanson and N. C. Ramiller. Innovating mindfully with information technology. *MIS quarterly*, pages 553–583, 2004.

[35] C.-M. Tan. *Search inside yourself*. Addison-Wesley Publishing Company, Massachusetts, 2012.

[36] R. Vidgen and X. Wang. Coevolving systems and the organization of agile software development. *Information Systems Research*, 20(3):355–376, September 2009.

[37] K. E. Weick, K. M. Sutcliffe, and D. Obstfeld. Organizing for high reliability: Processes of collective mindfulness. *Crisis management*, 3:81–123, 2008.

[38] C. Wohlin, P. Runeson, M. Hst, M. C. Ohlsson, B. Regnell, and A. Wessln. *Experimentation in software engineering: an introduction*. Springer Publishing Company, Incorporated, 2012.