

Seminars on Software Maintenance and Evolution: An Empirical Study of the Background Factors Affecting Student Success

Jussi Koskinen*

Department of Computer Science and Information Systems, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland

Abstract: Software maintenance and evolution (SME) is an important but challenging topic area for university-level computer science education. Seminars can be used to provide students with versatile and up-to-date knowledge on scientifically relevant issues. We organized three systematic university-level seminars on SME. In these seminars 127 groups have each been assigned the task of analyzing one scientific SME article. The main results concern background factors relating to the students, groups and articles as these affect student success in the seminars. This paper presents a strict statistical analysis and a discussion of these factors. Fourteen hypotheses were set and tested regarding the relation of various background factors and a student's success in the seminar. The results indicate a clear relation between some of the factors and success. Most of the student- and group-related factors clearly affected student success, whereas most of the article-related factors did not. The study also revealed many important ancillary results. The results support organizing, studying, and improving feasible seminars in this area.

Keywords: Software maintenance, software evolution, software engineering, teaching, seminars, curricula.

1. INTRODUCTION

Software maintenance and evolution (SME) is a large and economically important sub-field of software engineering [1]. Software maintenance most obviously includes making changes to a software system, but it also includes many processes which aim at supporting and managing the making of changes. These processes include, for example, reverse engineering, program comprehension, and maintenance cost estimation.

SME-activities have traditionally accounted to 50-75% of total software life-cycle costs in the case of successful systems with a long lifetime [2, 3]. According to some of the more recent studies, including [4, 5] this percentage may even be increasing. Therefore, SME is an important area of research.

The empirically derived so-called Lehman's first law [6] states that systems implementing computer applications in the real world must continually be adapted, otherwise they become progressively less satisfactory. The need for adaptations and enhancements is due to the new user requirements and technical requirements. Successful systems with a long lifetime are hard to replace but also hard to maintain. Thus, their maintenance is elaborate but often can not be avoided.

Because of the importance and problematic nature of SME there is a constant need in computer science education to follow scientific progress, and transfer the results achieved

to future professionals [7]. There is also a need to improve the present forms of teaching and representing issues related to SME. Seminars help in increasing both interactivity and the scope of the subject areas which can be covered in academic teaching. Despite these obvious needs and possibilities very few scientific studies have been published on software engineering seminars. Moreover, no earlier scientific studies appear(s) to have been done on SME-seminars.

Hence, a clear need exists to study the feasibility, potential and limitations of seminars in the context of SME-related knowledge. We discussed the organization of SME-related seminars and confirmed the feasibility of the here-applied seminar-based approach on a general level in an earlier study [7]. This paper is based on that earlier study, but has been greatly extended, has a different focus, and includes a much deepened analysis.

There has been a specific need to reveal and analyze background factors underlying student success in these kinds of seminars. Therefore, we analyze in detail the relationship between background factors and students' success in this paper. The analysis covers three software maintenance seminars involving 127 seminar groups comprising nearly 300 students. Tests of various hypotheses on these issues are reported; the results of these tests form the main contribution of this paper.

The rest of the paper is organized as follows. We first review the related works as an introduction to the general state, problems and options of teaching SME (Section 2). Next, we describe the research setting and the research methodology used, provide descriptive statistics on the organized seminars and list the set hypotheses (Section 3). The main body of the paper (Section 4) consists of statistical analyses and their

*Address correspondence to this author at the Department of Computer Science and Information Systems, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland; E-mail: koskinen@jyu.fi

results. The results are discussed in detail in a separate section (Section 5). Finally, the paper is summarized and conclusions are presented (Section 6).

2. RELATED WORKS

The background and context of our research on SME-seminars is that we have focused on teaching and investigating SME at the University of Jyväskylä (UJ) since 2000. Teaching of SME in the Faculty of Information Technology (FIT) of UJ has been organized by the author of this paper. FIT has systematically aimed at following on general level the recommendations of IEEE and ACM [9, 10] since 2002, concerning the composition of the courses it provides, e.g. regarding the coverage and emphasis of the issues deemed as essential [7], and currently, there are four advanced-level non-mandatory SME-courses provided for students.

We have earlier [7] identified the following three main practical problems in teaching SME: 1) weak coverage of SME in university-level teaching, 2) limitations of the existing educational materials, and 3) hardness of reproducing realistic maintenance situations in teaching SME. It should be emphasized that we do not aim at solving these very hard fundamental problems as such in this paper. Instead they are the background and reason for the approach we have developed earlier. We will here use the taxonomy of problems to organize the related works section.

2.1. Coverage of SME in Teaching

The IEEE's project SWEBOK has suggested an informed, reasonable, and extensive characterization of the general software engineering body of knowledge [8]. That knowledge has later been used in the joint ventures of IEEE and ACM to produce software engineering curriculum guidelines [9, 10]. General software engineering curricula development has also been discussed at general level e.g. in [11, 12].

The first general problem of teaching SME is that it generally has received very little attention from those involved in software engineering education as compared to its importance. SME has unfortunately traditionally not been among the most fashionable subject areas among students, companies, or academic staff, for that matter. The main reason to this has been that developing, and studying new kinds of software systems and techniques have been considered more interesting. This phenomenon has been recognized in many studies, including [13-15]. Due to this background it is especially important to offer the needed knowledge to the students in an interesting way in order to promote SME and to preserve existing interest.

The incorporation of SME into the software engineering curricula related to the CCSE-initiative [9] has been discussed, for example, in [16]. SME has rarely been taught extensively at university-level before the third year of studies [17, 18]. This is related to the fact that other issues have already reserved a relatively large portion of the available curricula [11].

Possible differences in this and other regards in the university-level education in Finland and USA have been inves-

tigated more closely in a separate sub-study in 2006 [18]. That study revealed that the amount of teaching hours dedicated to the subject area of SME in almost all of the Finnish universities and many of the top US universities was still clearly below the latest recommendations of IEEE and ACM [9, 10]. The study also revealed that interactive SME-seminars are rare or non-existent. The study covered such generally well-performed US universities as Carnegie-Mellon, MIT, Stanford, Berkeley, and all 14 Finnish science universities. There are naturally numerous courses which somehow deal with SME in various universities, but only few of them have been systematically reported in scientific series.

2.2. Educational Materials

The second general problem of teaching SME is that the existing educational materials have their limitations [19]. There exists many books on the area, but the basic limitation of text-books is that they generally do not stimulate interaction. There exists classic SME-books, but they are mainly out-dated from the technical point-of-view [2, 20]. Newer, management-oriented books include [21-24], but they do not cover all the important technical aspects.

There are some good books, including [5, 25-26] describing some of the technical aspects such as reengineering. There are also many practice-oriented books on closely related areas, especially for language-dependent debugging, e.g. [28] and tool-specific configuration management, e.g. [29]. However, many of these kinds of books are heavily dependent on specific techniques. There also exists good books which, however, strongly focus on more general, relatively loosely related, aspects, such as software reuse [27].

Additionally, SWEBOK recommends references for software maintenance including some of these books and a list of refereed articles [8, Chapter 6], which are potentially suitable also as a basis for SME-seminars.

2.3. Reproducing Realistic Maintenance Situations

The third general problem of teaching SME is that realistic maintenance situations appear only in realistic environments. This problem has been attacked in many studies with various approaches, including teaching emphasizing maintenance projects [30, 31], software evolution [32], practices of software industry [33], a project-based course on teaching SME focusing on code comprehension methods [34], and implementing change-requests on open-source systems [35].

Although practical project-based approach can be very valuable and desirable, dealing with real-world highly connected systems within the currently typical software engineering courses is often difficult in practice [7]. There are many reasons to this: students' limited knowledge and experience, limitations on the available instructor resources, technical challenges, limitations of the size of the applicative works, problems in program comprehension of large-scale systems etc.

We have earlier [7] concluded that this problem can be effectively attacked only by organizing the applicative tasks related to larger dedicated work-projects, which are per-

Table 1. Software Maintenance Seminars [7] © CISSE 2008 (reproduced with permission)

Aspect	SMS-I	SMS-II	SMS-III	Total
Involved groups	28	45	54	127
Active groups	24	37	46	107
Students	76	111	109	296
Seminar sessions	10 h.	13 h.	15 h.	38 h.
Organization	75 h.	105 h.	125 h.	305 h.
Total used teaching resources	85 h.	118 h.	140 h.	343 h.
Reports produced	173 p.	251 p.	259 p.	683 p.

formed in an ideal case in cooperation with industrial partners [30] or related to open-source systems [35]. This approach has been followed also e.g. in FIT in form of student projects.

2.4. Seminar-Based Approaches

Based on information retrieved from IEEE Xplore and ACM Digital Library there are no scientific studies on SME-seminars. We are not aware of any studies in this area, except our earlier work [7], in any scientific publication. The earlier research on seminars is also very scarce in the more general area of software engineering. There exists only some remotely related studies on areas such as knowledge work and tools [36] and electronic commerce [37].

As represented above, there are three main SME problems and two different kinds of issues to be taught; theoretical and applicative. Students clearly need large applicative team-based SME-projects. However, before they can realistically be expected to be mature enough to successfully commit themselves to those projects they need theoretical knowledge on SME. Ideally, they would be provided with a versatile and motivating introduction consisting of solid, theoretical and up-to-date science-based knowledge on SME.

Seminars can help in overcoming some of the problems related to the provision of that knowledge. Seminars enable much content within the typically strict limitations in the software engineering curricula and their interactive nature is especially suitable since preserving student interest is challenging in this area. Since there have been no earlier reported scientific studies on SME-seminars, we have organized seminars, systematically gathered novel empirical data concerning their characteristics, and performed statistical analyses.

2.5. Conclusion Based on the Related Works

This section has provided an overview of the related works which have some meaningful connections to our study. The review has revealed the state of the research in the area of SME. The most important observation is that there are no internationally reported scientific studies on SME-seminars, besides our earlier paper [7]. There are neither systematic scientific studies on general software engi-

neering seminars which would focus on studying the general factors which affect student success.

Most of the other works have focused on studying other forms of teaching, which are not our focus. We are interested here in the background factors which affect seminar success regardless of the peculiarities of the selected specific teaching forms or other situation-dependent design choices. That sort of basic knowledge concerning the background factors is useful for understanding seminars as an effective way to deal with university level students, in order to be able to profile students and to predict their success.

3. METHODOLOGY

We have organized three software maintenance seminars. The main characteristics of the seminars as such have been described earlier in [7]. We will repeat in this section such general information which is helpful for the reader to understand the context of the performed empirical study. The general context of the seminars has been identical. They have been organized as part of Software Engineering -course in FIT.

The instructor of all these seminars was the author of this paper. The general substance area of SME has been introduced in the course to the students by the seminar instructor, and the actual seminar has followed. Seminar work tasks have been assigned to students to be performed as group works. The group works have been mandatory. Seminar groups organized themselves autonomously. The main task of each seminar group was to analyze and report one scientific article.

3.1. General Descriptive Statistics of the Seminars

There have been a total of 127 seminar groups, and 1-5 students in each group. The basic statistics of the seminars (SMS-I, SMS-II and SMS-III) are shown in Table 1. Organization of the seminars has required 343 h. of teaching resources. The involved students have produced 207 credit units, corresponding to expected 8280 h. of studying. Additionally, about 540 h. have been used by the author of this paper to gathering data, performing the statistical analyses and preparing the related two scientific articles. Thereby, the total amount of resources allocated to the organization, stu-

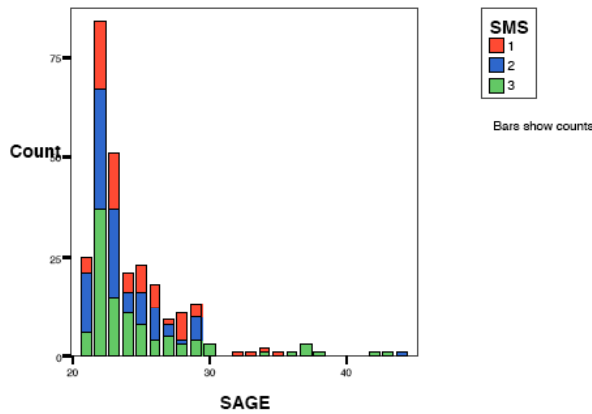


Fig. (1). Students' age distribution (SAGE) (SMS indicates the seminars).

dent involvement, and reporting of the seminars by now has been about 9100 h.

This is an exploratory empirical study. Guidelines for conducting empirical research in software engineering are provided e.g. by Kitchenham [38]. The statistical analyses of this paper have been received by using SPSS (v. 12.0.1 and v. 15.0.1) for Microsoft Windows (XP Professional v. 5.1.2600).

The main characteristics of the studied variables are presented in Appendix 1 for student-related variables (Table 6), group-related variables (Table 7), and article-related variables (Table 8).

3.1.1. Students

The notation which will be used consistently in all the oncoming figures is as follows: SMS-I = red, SMS-II = blue, SMS-III = green, circles = women, and triangles = men. There were 282 students who proceeded beyond the reservation of the seminar articles. 198 of these were men and 84 women. Average age = 24.3 years, median = 23.

Fig. (1) shows the age (SAGE) distribution of the students as stacked based on the individual seminars. Average year of studies at the university = 2.6 (median = 2). FIT grants degrees on information systems science (ISS), and information technology and computer science (IT). 93% of the students within the study were aiming at a degree within FIT. 82.3 % of the students were focusing on information systems science [7].

The total accumulated amount of credit units for each student (SCUT) was determined, mean = 150, minimum = 19, maximum = 335 credit units, standard deviation = 47, N = 269. Fig. (2) shows the distributions of SCUT and SAGE as a combined scatter plot (the linear regression between the variables is: $SCUT = 129.7 + 0.817 * SAGE$). Similarly, the total accumulated amount of computing credit units for each student (SCUC) was determined, mean = 76, minimum = 5, maximum = 159 credit units, standard deviation = 26, N = 269. Computing includes here all the mentioned computing disciplines of FIT. Correlation between SCUT and SCUC is high, Spearman's ρ (p) = 0.600, p = 0.000 (***)).

We use in this paper the following conventional notation regarding the risk level (p) of the results: $0.01 \leq p < 0.05$ are statistically almost significant results (*), $0.001 \leq p < 0.01$ are significant (**), and $p < 0.001$ are highly significant (***)). Most of the correlations in this paper are determined based on Spearman's ρ (p) since it does not require normality of the distribution of the tested variables.

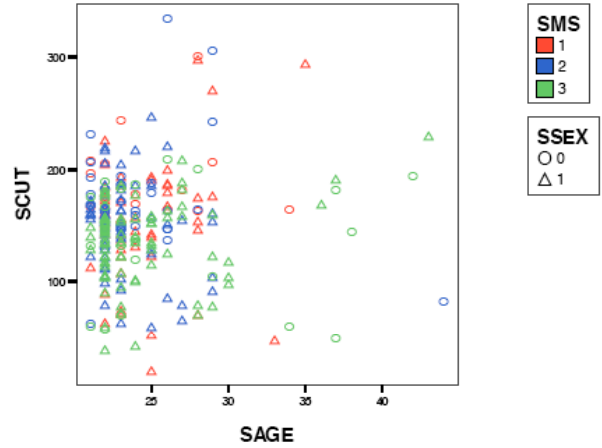


Fig. (2). Gathered student credit units: Total (SCUT) and students' age (SAGE) (SMS indicates the seminars, SSEX: 0 = women, 1 = men).

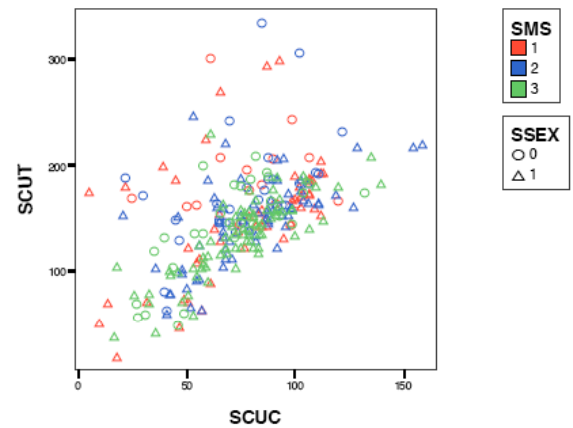


Fig. (3). Gathered student credit units: Total (SCUT) and computing (SCUC) (SMS indicates the seminars, SSEX: 0 = women, 1 = men).

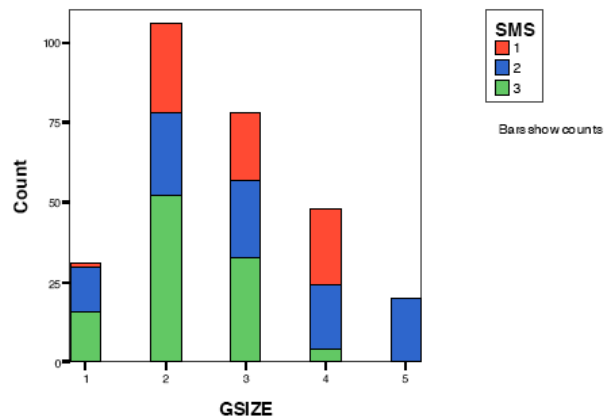


Fig. (4). Seminar group size (GSIZE) (SMS indicates the seminars).

Table 2. The Studied Articles and their Classification ¹

Theme: Sub-theme	N	Citations
Software maintenance techniques:	(50)	
Reverse engineering (TE:RVEN)	33	[43,45,49,55,58,63,64,68,70,73,74,75,77,79,80,81,84,86,89,90,105,106,110,122,123,128,147,150,154,155,162,163,166]
Reengineering (TE:REEN)	7	[66,87,88,119,146,151,153]
Program slicing (TE:SLIC)	5	[41,52,92,100,156]
Program visualization (TE:PVIS)	5	[54,59,67,135,165]
Software maintenance tasks:	(48)	
Maintenance processes (TA:MTPR)	12	[61,78,95,104,112,118,125,136,142,144,149,164]
Program comprehension (TA:PCMP)	23	[42,44,50,69,71,82,93,97,109,111,113,114,115,116,117,120,130,137,138,148,152,160,161]
Maintainability improvements (TA:MNTB)	8	[52,83,91,107,126,127,141,159]
Debugging (TA:DEBG)	5	[62,94,99,157,158]
Economic estimation of maintenance:	(29)	
Maintenance cost estimation (EE:COST)	5	[53,60,102,139,145]
Maintenance effort estimation (EE:EFFE)	12	[40,46,48,56,57,65,72,85,96,121,124,140]
Metrics and measurement (EE:METR)	12	[47,76,98,101,103,108,129,131,132,133,134,143]
Total:	127	

¹ N = number of the articles

Fig. (3) shows the distributions of SCUT and SCUC as a combined scatter-plot (the linear regression between them is: $SCUT = 72.4 + 1.013 * SCUC$). As noted earlier students organized themselves into seminar groups. The group size distribution is given in Fig. (4). Median of the group size was 3.

3.1.2. Articles

The seminars have covered most of the important issues in the SME-field. The set of articles offered to the students has been selected by the seminar instructor based on his experience on the area. The Ph.D. thesis of the instructor [39] included a large literature review on SME with over 400 references. That review served as a starting point for the literature gatherings and selections for this study. All the selected articles had to fulfill the following criteria: 1) English as a language, 2) publication on refereed international scientific series or established books, 3) publication year later than 1975, and 4) SME as a focus area. Complete bibliographic information of the articles offered to the students is provided in the reference list [40-166].

Each article is also characterized as an annotation to its reference in terms of: the seminar in which it was studied, its main theme, sub-theme, main research method, length, and formality. That information is provided in the reference list in order to support transparency, replication of the seminars, and reliable permanent storage of the information. The studied articles have been classified based on their themes and

sub-themes in Table 2. There were three primary themes and eleven sub-themes.

The articles have appeared in 33 different publication series. The distribution of the studied articles related to the series is shown in Appendix 2 (Table 9). Most of the studied papers have been published in good academic journals or conference proceedings and others in established widely-used software engineering text-books.

The counts of the students analyzing articles on the specific main themes are presented in Fig. (5). Similarly, the counts of the students analyzing articles on the specific sub-themes are presented in Fig. (6).

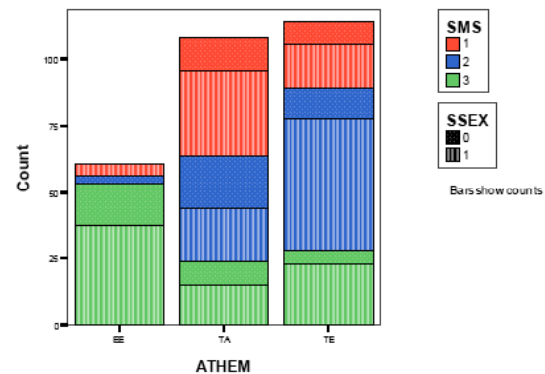


Fig. (5). Article themes (ATHEM: EE = economic estimation, TA = tasks, TE = techniques) (SMS indicates the seminars, SSEX: 0 = women, 1 = men).

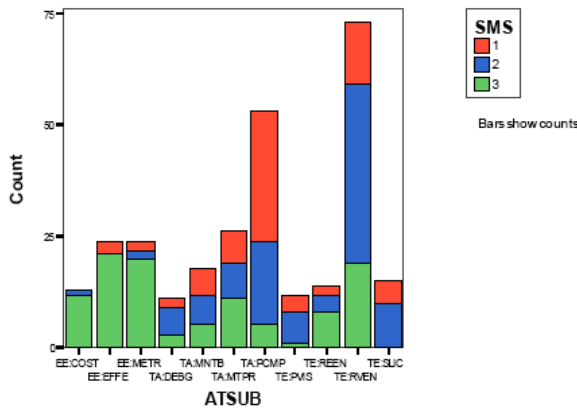


Fig. (6). Article sub-themes (ATSUB) (as explained in Table 2) (SMS indicates the seminars).

Each individual student has been given an opportunity to freely select; from the predefined set, the paper which best matches his or her background, talents, and interests. The selection procedure is detailed in [7]. The primary research methods applied in the analyzed papers were as follows: Theoretical analyses (T) (38.2%), empirical (E) (32.5%), constructive (C) (16.3%), and surveying (S) (13.0%).

3.2. Tasks, Objectives and Evaluation

The task of each seminar group was to:

- 1) Analyze one scientific article.
- 2) Present its analysis at the seminar.
- 3) Write a carefully finished summary report of the article.

- 4) Serve as opponent to one other group.
- 5) Listen the presentations of other groups.

The general nature of educational objectives can be classified e.g. based on so-called Bloom categories [8, p. 1-7]. These concern increasing depth on: Knowledge, comprehension, application, analysis, synthesis, and evaluation. In our case, the students' general starting level has been such that they have had no earlier experience on software engineering related seminars or other SME-courses. For most of them this was also the first time they conducted academic group works and analyzed scientific articles written in non-native tongue. Therefore, the main general learning objectives have been to increase knowledge on SME, and capabilities on analyzing science-based articles of the field [7].

The presentations have required the students to be able to summarize the main points of the analyzed method or technique. Since the analyzed articles have been scientific and the seminars did not intend to include applicative tasks, there were no applicative objectives. Since the tasks were performed mainly as group works communication and cooperation skills were implicitly needed during the process [7].

Students have been evaluated based on a standardized form. The use of that form has aimed at paying attention to all the central relevant aspects in these kinds of seminars during the whole lifecycle of the seminar works. The students have been given feedback by the instructor multiple times and by the assigned opponent group at the seminar.

During the first phase review of the seminar reports the form was used as a basis for giving the groups further improvement suggestions. Final reports have gone through two-phase examination by the instructor. The seminar work qual-

Table 3. Evaluation Form [7] © CISSE 2008 (reproduced with permission)

Number	Criteria	Details
1	Layout of the report	There were strict instructions for the coherent format of the reports
2	Extent of the report	The size of the report within the allowed range for the article
3	Relevance of the report's focus	The report had to focus on the issues inside the article
4	Quality of language	Use of complete sentences, readability, style etc.
5	Correctness of the details	Accuracy of the presented assertions, numerical data, references
6	Definition of the central concepts	Their presence and correspondence to the original ones
7	Description of the research problem	The problem as represented in the article
8	Description of the hypotheses	For empirical works
9	Description of the applied research methods	Procedure of the study
10	Description of the results	Main results and their implications
11	Group's own observations and conclusions	Students' evaluation of the article
12	Clarity of the presentation	Presence of the main points, oral and audio-visual appearance
13	Interaction between the group and audience	Answers to the opponents and others
14	Conformance to the predefined schedules	Meeting the mile-stones for reviews etc.
15	Action of the group in the role of opponents	Constructive comments

ity has been systematically monitored by the seminar instructor based on an evaluation form which contains 15 criteria related to the report, presentation, and other factors [7].

The used criteria and the contents of the corresponding standardized evaluation form are presented in Table 3. Also the students have been aware of these criteria already prior to starting their seminar works. Therefore, these have been the operational learning objectives. Criteria 1-10 relate to the seminar report and criteria 11-15 to its oral presentation.

The instructor performed the evaluation related to the oral presentation right after the corresponding seminar session. Listening the presentations of other groups in turn was controlled only by requiring participation to the seminar sessions. The instructor made the final evaluation of the report-related factors after the group returned its final version of the seminar report. The instructor evaluated the seminar works group-wisely.

3.3. Seminar Points and Course Grade

All of the above-listed criteria have had a uniform weight to the final result of each group. The criteria have been considered carefully, in detail, and aiming at impartiality for each group by the seminar instructor. The above-described evaluation form was used to support objectivity in the evaluations. The final result (SPS = Student's Points of Seminar) has been received based on these factors. SPS has been handled as a continuous variable. This variable measures the success of individual students on the seminars. The distribution of SPS as categorized (rounded downwards to integers in to give the reader a compact and illustrative view) into six classes (SPS_CATEG) is shown in Fig. (7), which also shows the distribution for each seminar separately.

The applied scaling roughly corresponds to international grading systems: A ... F, where: A = 5 = excellent, B = 4 = very good, C = 3 = good, D = 2 = satisfactory, E = 1 = sufficient, F = 0 = fail. The distribution of the grades, however, differs e.g. from ECTS grading system [167], since ECTS requires specific percentiles for the categories (A: 10%, B: 25%, C: 30%, D: 25%, E: 10%). The distribution of SPS is not normal as tested by (two-tailed, one sample) Kolmogorov-Smirnov test, $p < 0.00$, since the average of the grades is relatively high due to the general success of the seminars. This must be noted while interpreting the grades based on the standard keywords. However, because we are here mainly interested in the actual relative differences between the students' performance within the sample as measured numerically, instead of drawing conclusions based on the keywords, this is not a methodological problem for the study.

Other SPS characteristics were, mean = 3.55, minimum = 0, maximum = 5.30, standard deviation = 1.33, $N = 282$. 107 groups completed their works. 20 groups could not follow the schedule. The seminars produced 683 pages of thoroughly inspected seminar reports which were made available to all students of FIT.

The distribution of the students' pure examination points in the main course, i.e. the Software Engineering course (SPE_EXE_CATEG) is shown in Fig. (8). The examination consisted of four questions each requiring a short; about one

page essay on the subject as an answer. As an incentive for the students the received seminar points were finally added to the examination points to receive the final course grade (SPE).

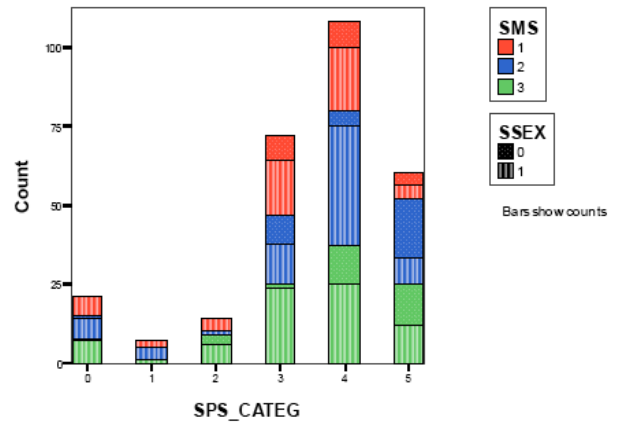


Fig. (7). Students' categorized seminar points (SPS_CATEG) (SMS indicates the seminars, SSEX: 0 = women, 1 = men).

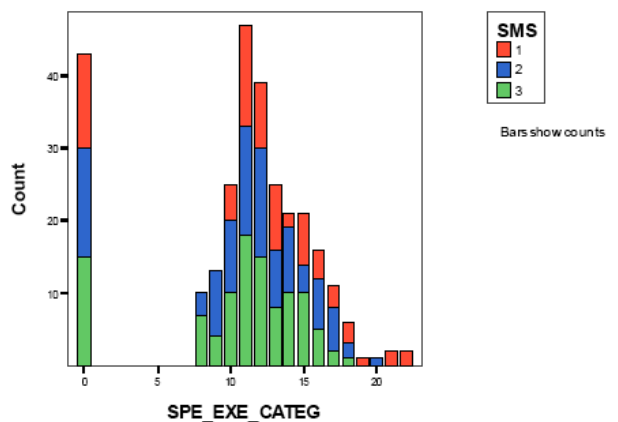


Fig. (8). Students' Software Engineering course examination points (SPE_EXE_CATEG) (SMS indicates the seminars).

The relative effect of SPS to SPE was 17.9 ... 29.4%. SPE characteristics were, mean = 14.52, minimum = 0, maximum = 28, limit for pass = 12, and for excellent grade = 23 points, standard deviation = 6.59, $N = 281$.

3.4. Hypotheses

The set hypotheses were as follows (S refers here to students, G to groups, and A to articles):

H_{S1}) General studying efficiency and especially studying efficiency on computing positively affects the student's seminar success. Computing efficiency especially is relevant for the topic area of the seminars.

H_{S2}) Student sex (i.e. gender) does not have significant influence on the student's seminar success, since men and women both have their own strong sides which are relevant in these seminars. It is likely that men are better in mathematics and technical issues, whereas women in tasks requiring communication and social skills.

- H_{S3}) Student age does not have significant influence on the student's seminar success, since age has both positive and negative effects. The positive effects include e.g. increased experience and the negative ones slowly weakening physical fitness due to aging having effect on studying efficiency.
- H_{S4}) Student focus area does have influence on the student's seminar success, such that those majoring computer science (CS) and similar subjects reach better seminar success. This is due to the better and more extensive relevant prior knowledge on the general topic area of the seminars.
- H_{G1}) Group size does not have significant influence on the student's seminar success, since larger groups are assigned proportionally larger tasks.
- H_{G2}) General average studying efficiency of the group positively affects the student's seminar success. This is due to the probably effective utilization of the capabilities of the gifted members of the group inside the group. The group-related hypotheses H_{G2}-H_{G4} were formed after the statistical analyses of the student-related factors and the group size were completed. Those results suggested that it would be worthwhile to study the group-related factors more closely. Therefore, the knowledge gained from those earlier parts of the study has already been available for and affected the assumptions in these cases.
- H_{G3}) The composition of the group in terms of the sex of the individual members does have influence on the student's seminar success in case that the sexes differ regarding the seminar success on the student-level (H_{S2}). The reason to this is the same as in case of the hypothesis H_{G2}; the group is assumed to be capable of utilizing its gifted members.
- H_{G4}) The composition of the group in terms of age of the individual members does affect the student's seminar success in case that the students having different age differ regarding the seminar success on the student-level (H_{S3}).
- H_{A1}) The type of the article's publication series has effect on student's seminar success since the nature and characteristics of the articles differ in different kinds of series.
- H_{A2}) Students who analyze conference articles have better seminar success than those analyzing journal articles, since conference articles are less detailed and therefore easier to understand by the scientifically rather inexperienced average students.
- H_{A3}) The article length does not have significant influence on student success since larger articles are targeted to larger groups, and thereby the relative required effort is the same.
- H_{A4}) The general article theme and sub-theme do not have significant influence on student's seminar success, since each student selects his or her article.
- H_{A5}) Formality of the article does not have effect on student success, since many informal articles are among the

possible choices and each student can select his or her article freely in this sense.

- H_{A6}) The primary research method applied in the analyzed articles is not assumed to have any particular effect.

4. RESULTS

As noted earlier, this section describes the performed statistical analyses and the received results. We are especially interested in the general background factors which contribute to student's seminar success. Therefore, the main relevant dependent variable here is the above-described Student's Points of Seminar (SPS). There are two main categories of potential causes of differences in performance: Student background, including group formation, and characteristics of the articles. We have aimed at keeping all other characteristics constant during the seminars, as far as it has been possible, in order to focus on these ever-present background factors.

4.1. Student Background Effects

4.1.1. General Studying Efficiency (STEFF)

SCUT indicates student credit units as defined earlier. Another variable called SYCS indicates the years a student has studied in the university before that time. General long-term studying efficiency is approximated by a variable called STEFF = SCUT / SYCS. STEFF characteristics were, mean = 30, minimum = 3, maximum = 91 credit units, standard deviation = 12, $N = 269$. Fig. (9) shows as a scatter-plot the relation between total studying efficiency and seminar success (the linear regression is: $SPS = 2.585 + 0.031 * STEFF$). As expected there is a clear, statistically highly significant correlation between these two variables. Spearman's $\rho = 0.253$, $p = 0.000$ (***)). SPS is here the dependent variable. Similarly, there is also a statistically highly significant correlation between SPS and the received points in the related course examination (SPE_EXE), Spearman's $\rho = 0.253$, $p = 0.000$ (***)). Spearman's correlation coefficient for the pair (STEFF, SPE_EXE) = 0.221, $p = 0.000$ (***)).

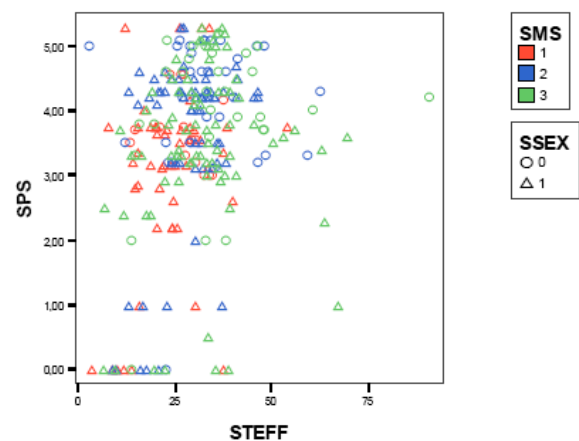


Fig. (9). Total studying efficiency (STEFF) and seminar success (SPS) relation (SMS indicates the seminars, SSEX: 0 = women, 1 = men).

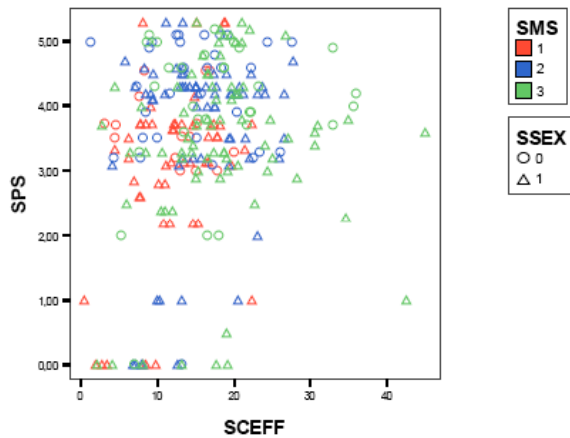


Fig. (10). Computer science studying efficiency (SCEFF) and seminar success (SPS) relation (SMS indicates the seminars, SSEX: 0 = women, 1 = men).

Other related significant correlations are presented in their main context, in the oncoming subsections: computer science studies (SCEFF), and group efficiency effects (GEFF).

4.1.2. Computer Science Studying Efficiency (SCEFF)

SCUC indicates student credit units on computer science and information technology as defined earlier. Long-term student efficiency in this regard is approximated by a variable called $SCEFF = SCUC / SYCS$. SCEFF characteristics were, mean = 15.6, minimum = 0, maximum = 45 credit units, standard deviation = 7, $N = 269$. Fig. (10) shows as a scatter-plot the relation between computing studying efficiency and seminar success, likewise as in the previous figure: $SPS = 2.880 + 0.042 * SCEFF$. As can be expected there is a clear, statistically highly significant correlation also between these two variables. However, the difference is somewhat less significant than in case of STEFF. Spearman's $\rho = 0.193$, $p = 0.002$ (**). As Fig. (9) and Fig. (10) suggest, STEFF and SCEFF are very strongly related, Spearman's $\rho = 0.782$, $p = 0.000$ (***). Likewise as in case of STEFF, there is a statistically highly significant correlation for the pair (SCEFF, SPE_EXE), Spearman's $\rho = 0.243$, $p = 0.000$ (***).

4.1.3. Sex (SSEX)

Women were clearly more successful than men. Women: $SPS = 4.02$, $N = 84$, standard deviation = 1.02. Men: $SPS = 3.34$, $N = 198$, standard deviation = 1.39. Fig. (11) shows the relation between sex and the student's seminar success for each seminar as a box-plot. Box-plot is used since sex has only two valid values: 0 indicates women, and 1 men. Box-plots show the median, quartiles, and outliers. Medians within each category are shown by horizontal center lines. Circles and asterisks show the outliers. The difference of the means of groups can be tested by t-test. All the t-tests in this paper are performed as Student's independent-samples 2-tailed tests. The tests include Levene's test for equality of variances among the groups and potentially needed remedy of the inequality. The difference of the means of the groups for the whole data-set is here statistically highly significant, $t = 4.563$, $df = 211.4$, $p = 0.000$ (***). Possible differences

between these groups regarding other relevant factors potentially affecting seminar success need also to be checked. These groups did not differ statistically significantly regarding age (SAGE) or seminar group size (GSIZE).

On the other hand women were more successful than men also in terms of the Software Engineering course examination (SPE_EXE). Women: $SPE = 11.9$, $N = 84$. Men: $SPE_EXE = 10.9$, $N = 198$. Fig. (12) shows similarly the relation between sex and SPE_EXE for each seminar as a box-plot. 0 indicates women, and 1 men. The difference, however, is not statistically significant. Therefore, women performed overall relatively better in the seminar than in the examination.

Due to the general success of women in the course, the background variables were studied even more thoroughly to find out whether women were better also in general in their studies. It was found out that women were better also in terms of long-term studying efficiency (STEFF). Women: $STEFF = 33.2$, $N = 75$. Men: $STEFF = 28.8$, $N = 194$. The difference is statistically significant: t-test, $t = 2.810$, $df = 267$, $p = 0.005$ (**). On the other hand, men were somewhat more successful in terms of long-term computer science studying efficiency (SCEFF). That difference, however, is not statistically significant. Other related significant correlations are presented in their main context, in the oncoming subsections: group sex effects (GSEX) and article-related effects (AMETH).

4.1.4. Age (SAGE)

The relation between SPS and student age (SAGE) was studied by dividing the sample into two groups. The two groups were as follows: SAGE more than average (i.e. > 24.3 years) ('older students', $SPS = 3.22$, $N = 89$), and SAGE less than average ('younger students', $SPS = 3.69$, $N = 181$).

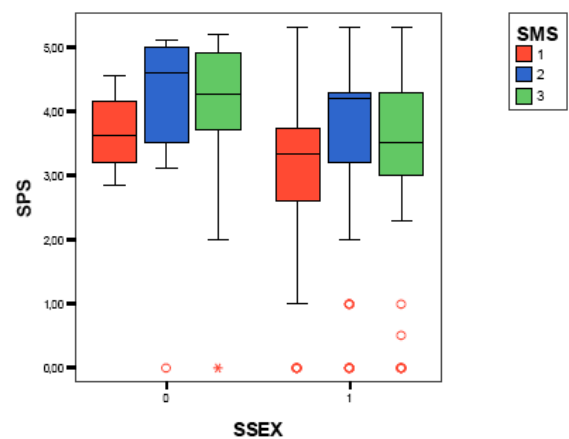


Fig. (11). Relation between student sex (SSEX: 0 = women, 1 = men) and seminar success (SPS) (SMS indicates the seminars).

The difference is statistically significant: t-test, $t = 2.765$, $df = 268$, $p = 0.006$ (**). Spearman's $\rho = -0.223$, $p = 0.000$ (***). Fig. (13) shows the relation between SAGE and SPS as a scatter-plot (the linear regression is: $SPS = 4.42 - 0.037 * SAGE$).

This same pattern was identified also while comparing student success in terms of Software Engineering course

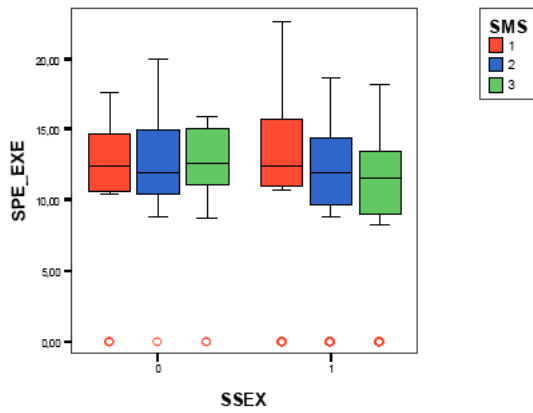


Fig. (12). Relation between student sex (SSEX) (SSEX: 0 = women, 1 = men) and Software Engineering course examination success (SPE_EXE) (SMS indicates the seminars).

examination (SPE_EXE). 'Older students', SPE_EXE = 10.3, $N = 90$. 'Younger students', SPE_EXE = 11.8, $N = 181$. Also this difference is statistically significant, although less significant than in case of SPS and SAGE: t -test, $t = 2.100$, $df = 269$, $p = 0.037$ (*). Therefore, older students performed relatively worse in the seminar than in the examination. The groups did not differ statistically significantly regarding sex or studying efficiency. Other related significant correlations are presented in their main context: group age effects (GAGE) and article length (ALEN).

4.1.5. Student Focus Areas

Most of the students were aiming at a degree within FIT, but some were from other universities or faculties. Those students who were from other universities and supplementing their degree in FIT (SM) performed worse (SPS = 3.22, $N = 48$) than others (SPS = 3.60, $N = 220$). Also those students who were from other faculties (SIT) performed slightly worse (SPS = 3.32, $N = 20$) than others (SPS = 3.55, $N = 248$). Student's major subject (SMAJ) was also investigated. There were 15 different major subjects. The students were divided into two groups based on whether their major subject was technically-oriented or not (STECH). The students with information systems science as major subject were more economics-oriented whereas the students with information technology were more technically-oriented. Also the students of software engineering were technically-oriented. Students coming from outside the faculty were not technically-oriented. Students of information systems science or other even less technical disciplines (SPS = 3.60, $N = 216$) did succeed, perhaps a little bit surprisingly, slightly better than other more technically-oriented students (SPS = 3.26, $N = 52$). None of these differences, however, is statistically significant.

Focus areas of studies were also looked more closely. During 1998-2004 FIT provided 15 different lines (SLINE), which are effectively alternative acceptable combinations of courses focusing on different aspects of information technology. These lines cover the area of ICT well. These lines were: System development, software engineering, software production, electronic commerce, software business, digital media, group technologies, multimedia, user-friendly com-

puting, teaching, embedded systems, data communication, mobile computing, optimization and simulation. Students studying user friendly computing and digital media were a little bit more successful than others, but the found differences were not statistically significant. Therefore, there were no statistically significant differences in student performance related to any of their focus areas.

4.2. Group Effects

4.2.1. Group Size (GSIZE)

There were 20 students who did not pass the seminar. 10 of these were in 1-person groups and eight in a 2-person group. Therefore, 90% of the seminar failures were related to small groups. All 5-person groups (20 students) were successful (SPS = 3.83).

Fig. (14) shows the relation between group size and seminar success for each seminar as a box-plot.

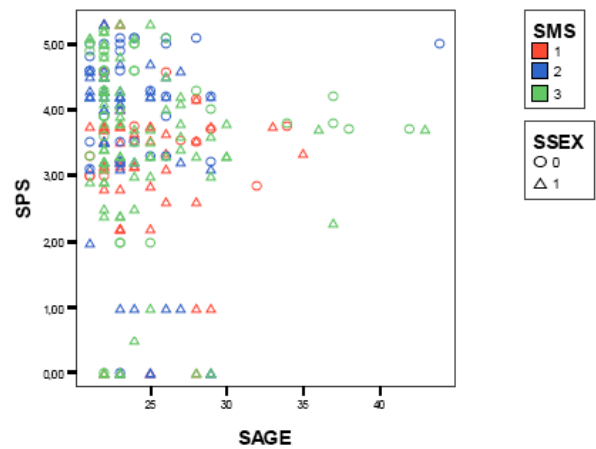


Fig. (13). Relation between student age (SAGE) and seminar success (SPS) (SMS indicates the seminars, SSEX: 0 = women, 1 = men).

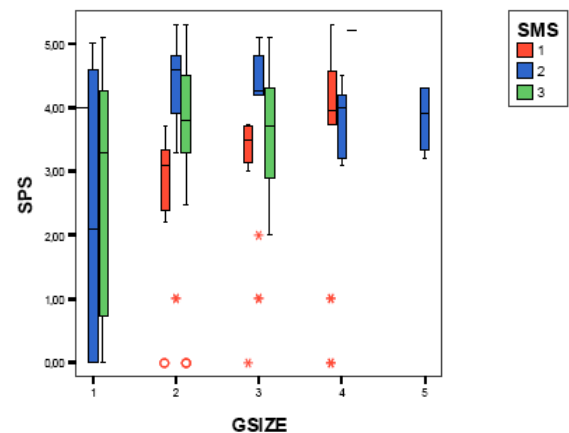


Fig. (14). Relation between seminar group size (GSIZE) and seminar success (SPS) (SMS indicates the seminars).

Larger groups were in average statistically almost significantly more successful than smaller ones as tested by correlation analysis for the whole data-set; Spearman's $\rho = 0.132$, $p = 0.027$ (*).

The distribution of men and women in the different size groups is shown in Fig. (15). The different size groups or analysis groups did not differ statistically significantly regarding sex (SSEX, GSEX) or studying efficiency (STEFF), but they differed regarding student's age and group's age. The relation of student age and group size is depicted as a box-plot for each seminar in Fig. (16). Spearman's ρ for the pair (GFSIZE, SAGE) = -0.209, $p = 0.001$ (**). Spearman's ρ for the pair (GFSIZE, GAGE) = -0.158, $p = 0.008$ (**). Other related significant correlations are presented in their main context, in the oncoming subsections: article series (AJOUR) and article length (ALEN).

4.2.2. Group's Composition in Terms of the General Studying Efficiency (GEFF)

We calculated a variable GEFF for the groups as an average of the general studying efficiency (STEFF) of the group members. Characteristics were, mean = 30, minimum = 3, maximum = 67.5 credit units, standard deviation = 11, $N = 126$. Fig. (17) shows as a scatter-plot the relation between GEFF and seminar success (linear regression: $SPS = 2.257 + 0.038 * GEFF$).

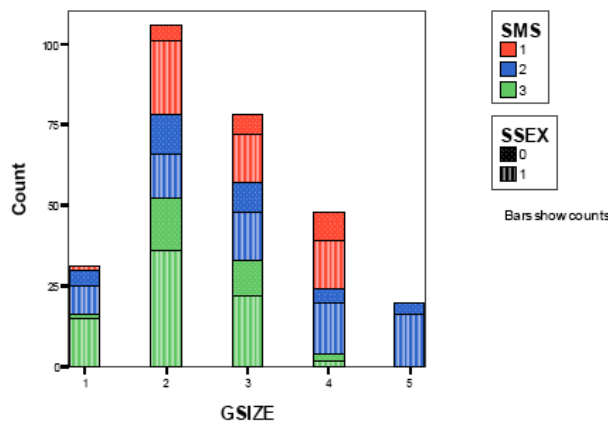


Fig. (15). Group composition in terms of student sex (SSEX: 0 = women, 1 = men) and group size (GFSIZE) (SMS indicates the seminars).

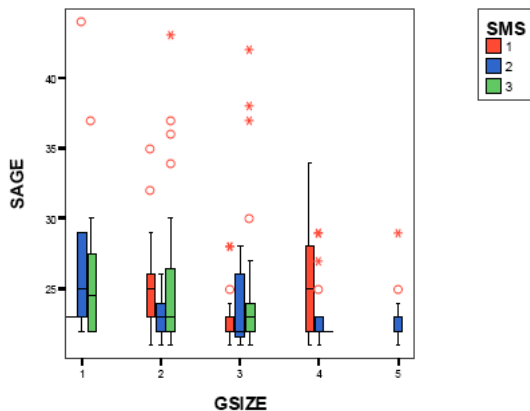


Fig. (16). Group composition in terms of student age (SAGE) and group size (GFSIZE) (SMS indicates the seminars).

As expected there is a clear, statistically highly significant correlation between these two variables. Spearman's $\rho = 0.336$, $p = 0.000$ (***). GEFF is by definition related to

STEFF, Spearman's $\rho = 0.792$, $p = 0.000$ (***). GEFF is also related to SCEFF, Spearman's $\rho = 0.565$, $p = 0.000$ (***). Additionally, there is a statistically highly significant correlation between GEFF and GSEX; Spearman's $\rho = -0.213$, $p = 0.000$ (***), because women had better STEFF-levels as described earlier. Group's efficiency also correlated statistically significantly with the oncoming individual course examination grade (SPE_EXE); Spearman's ρ for the pair (GEFF, SPE_EXE) = 0.180, $p = 0.002$ (**).

We also studied the relation of the individual student and the rest of the group (i.e. the complement). We determined the average studying efficiency of the rest of the group (GCEFF). GCEFF approximates here; in a way, the quality of the potential relevant support from the other group members. E.g. for the singleton groups the value of this variable = 0. There is a statistically highly significant correlation between STEFF and GCEFF; Spearman's $\rho = 0.341$, $p = 0.000$ (***). The situation is depicted in Fig. (18) as a scatter-plot.

4.2.3. Group's Composition in Terms of Sex (GSEX)

We calculated a variable GSEX for the groups as an average of the sexes (SSEX) of the members of the group with characteristics: mean = 0.72, minimum = 0, maximum = 1, standard deviation = 0.40, $N = 127$.

Fig. (19) shows as a scatter-plot the relation between GSEX and seminar success (the linear regression is: $SPS = 4.207 - 1.166 * GSEX$). As expected there is a clear, statistically highly significant correlation between these two variables. Spearman's $\rho = -0.250$, $p = 0.000$ (***). GSEX is by definition related to SSEX; t-test, $t = -23.752$, $df = 116.9$, $p = 0.000$ (***).

The distribution of the groups was such that there were only 19 mixed groups (14.9%). The other 108 groups consisted solely of one sex.

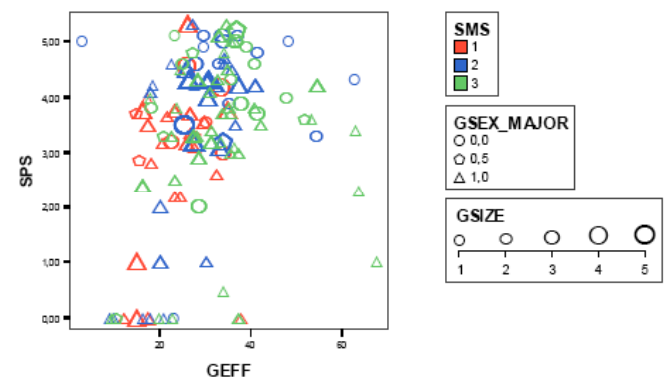


Fig. (17). Group's studying efficiency (GEFF) and seminar success (SPS) relation (SMS indicates the seminars, GFSIZE is the group size, and GSEX_MAJOR the majority sex of the group: 0 = women, 1 = men).

We studied the relation of the individual student's sex and the composition of the rest of the group (i.e. the complement) in this regard (GCSEX). There is a statistically highly significant relation between SSEX and GCSEX; t-test, $t = -10.925$, $df = 114.8$, $p = 0.000$ (***). Therefore, the groups

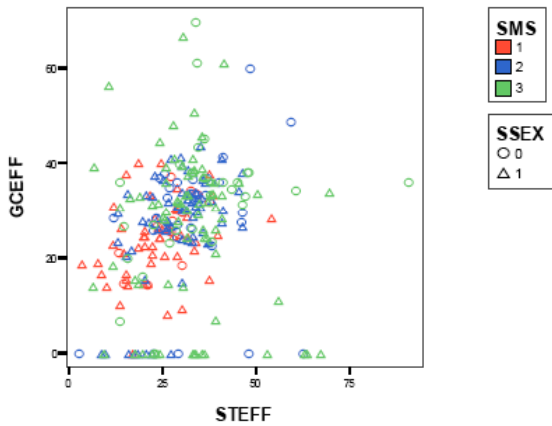


Fig. (18). Relation between student's studying efficiency (STEFF) and the average studying efficiency of the rest of the group (GCEFF) (SMS indicates the seminars, SSEX: 0 = women, 1 = men).

are highly polarized in this sense. Additionally, there exists the relation to GEFF as described above.

4.2.4. Group's Composition in Terms of Age (GAGE)

We calculated a variable called GAGE for the groups as an average of the ages (SAGE) of the members of the group. GAGE characteristics were, mean = 24.7, minimum = 21, maximum = 44, standard deviation = 3.57, $N = 126$. Fig. (20) shows as a scatter-plot the relation between GAGE and seminar success (the linear regression is: $SPS = 4.808 - 0.051 * GAGE$) There is a statistically highly significant correlation between these two variables. Spearman's $\rho = -0.264$, $p = 0.000$ (***). GAGE is by definition related to SAGE, Spearman's $\rho = 0.783$, $p = 0.000$ (***)

Additionally, there consequently exists the relation between age and group size as described earlier. Another related significant correlation is presented in its main context: article length (ALEN).

We also studied the relation of the individual student and the rest of the group (i.e. the complement), in this sense. We determined the average age of the rest of the group (GCAGE).

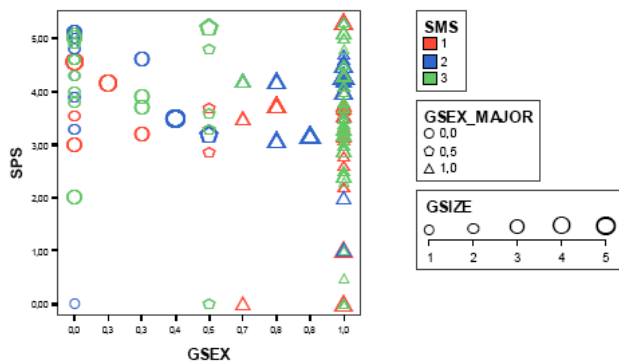


Fig. (19). Group's composition in terms of the majority sex of the group (GSEX_MAJOR: 0 = all women, 1 = all men) and seminar success (SPS) relation (SMS indicates the seminars, and GSIZE is the group size).

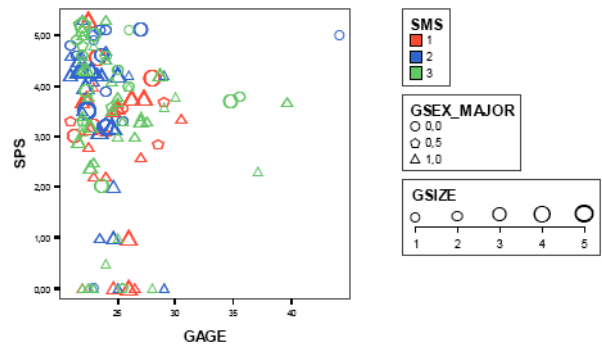


Fig. (20). Group's composition in terms of the members' age (GAGE) and seminar success (SPS) relation (SMS indicates the seminars, GSIZE is the group size, and GSEX_MAJOR the majority sex of the group: 0 = all women, 1 = all men).

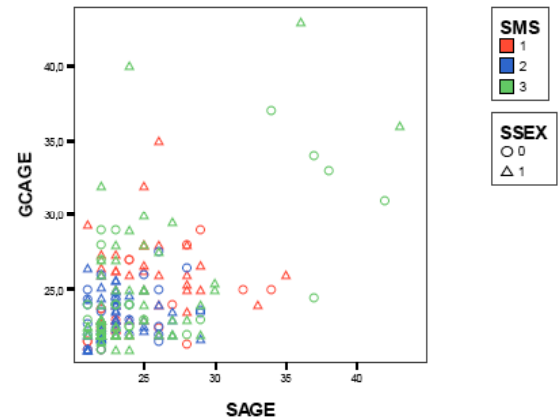


Fig. (21). Relation between student's age (SAGE) and the average age of the rest of the group (GCAGE) (SMS indicates the seminars, SSEX: 0 = women, 1 = men).

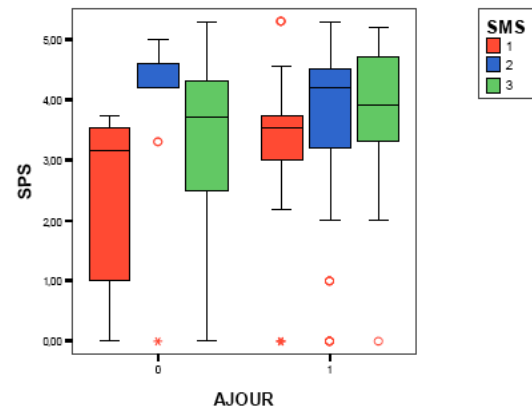


Fig. (22). Relation between journal article selection (AJOUR) and seminar success (SPS) (SMS indicates the seminars).

There is a statistically highly significant correlation between SAGE and GCAGE; Spearman's $\rho = 0.412$, $p = 0.000$ (***). The situation is depicted in Fig. (21) as a scatter-plot.

4.3. Article Effects

4.3.1. Article Series (AJOUR)

Students who analyzed journal articles were more successful ($SPS = 3.67$, $N = 199$) than those who analyzed con-

ference articles (SPS = 3.25, $N = 83$). The relation is shown in Fig. (22) as a box-plot for each seminar (1 denotes the journal article case). The difference between the cases (AJOUR) is statistically almost significant; t-test, $t = -2.218$, $df = 126.7$, $p = 0.028$ (*). The groups did not differ statistically significantly regarding the relevant background variables including: STEFF, SSEX, SAGE, GEF, and GSEX.

However, they differed clearly most importantly regarding group size (students having selected journal articles, $G\text{SIZE} = 3.0$, students having selected conference articles, $G\text{SIZE} = 2.1$). The difference is statistically highly significant: t-test, $t = -7.854$, $df = 218.5$, $p = 0.000$ (***). The reason to this variance is quite obvious since journal articles are longer (average within the sample = 27.9 pages) than conference articles (average = 14.2) and larger groups selected longer articles than smaller groups. There is a statistically highly significant difference in article length (ALEN) regarding AJOUR; t-test, $t = -11.054$, $df = 280.3$, $p = 0.000$ (***). Therefore, also the effects of the article length needed to be studied. Students who had much credit units (SCUT) selected more keenly journal articles than others; t-test, $t = -3.363$, $df = 267$, $p = 0.001$ (**). However, they were not more efficient than others in their studies (STEFF).

We also tested the potential effects of the article's publication year (AY) to SPS, since more recent articles might be more interesting to the students and thereby causing better success. However, there was no significant correlation between AY and SPS ($p > 0.05$).

4.3.2. Article Length (ALEN)

Length of the articles (ALEN) varied, mean = 23.9, minimum = 4, maximum = 66, standard deviation = 14.4. Students selecting articles, whose length was above the average were more successful (SPS = 3.79, $N = 116$) than others (SPS = 3.38, $N = 166$). The difference is statistically almost significant: Spearman's $\rho = 0.145$, $p = 0.015$ (*). The relation between SPS and ALEN is shown graphically in Fig. (23). The figure shows, e.g., that there were no failures related to the longer articles.

The cases did not differ statistically significantly regarding most of the other relevant background variables including: STEFF, SSEX, GEF, GSEX, but they differed regarding student's (SAGE) and group's age (GAGE). Spearman's ρ for the pair (ALEN, SAGE) = -0.232, $p = 0.000$ (***), and Spearman's ρ for the pair (ALEN, GAGE) = -0.207, $p = 0.000$ (***). This situation can be explained by the fact that the longer articles were clearly selected by the larger groups (G\text{SIZE}), which in turn, as explained earlier, were clearly more successful than others and there were also significantly less older students in the larger groups. It is very unlikely that the length of the article would instead be the real explaining factor of success since length conversely can be expected to cause additional mental work. On the other hand, age as such is also a very unlikely real cause since there were no significant correlations between SPS and STEFF regarding ALEN.

However, length was compensated by larger average group sizes. Spearman's ρ for the pair (ALEN, G\text{SIZE}) = 0.859. The correlation is very strong and statistically highly

significant, $p = 0.000$ (***). The group size effects can be neutralized in this sense by computing a new variable: $\text{ALENPP} = \text{ALEN} / \text{G\text{SIZE}}$ (mean = 8.7, minimum = 3.5, maximum = 26.0, standard deviation = 3.00). This variable approximates the mental effort of each student in terms of the length of the original text pages to analyze. Fig. (24) illustrates as a scatter-plot the relation between SPS and ALENPP. The student population was divided into two groups based on the value of this variable. There was no statistically significant difference remaining regarding SPS between these groups. Therefore, the effect on SPS is due to some other factor(s); most probably mainly due to the group size (G\text{SIZE}). ALENPP neither correlated statistically significantly with G\text{SIZE} ($p > 0.05$). Therefore, the articles assigned by the groups did not cause bias in terms of their length. Another related significant correlation is explained in its main context, in the oncoming subsection: article formality (AMATH).

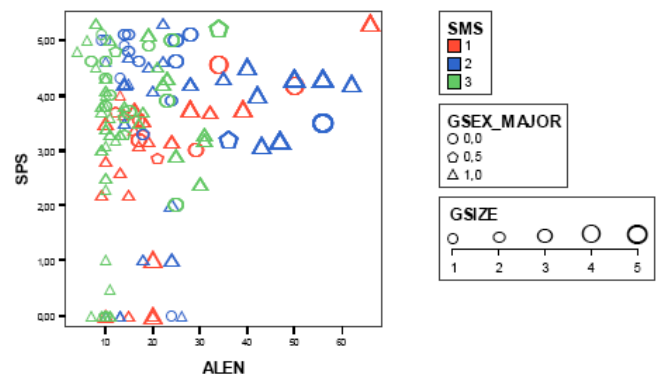


Fig. (23). Relation between article length (ALEN) and seminar success (SPS) (SMS indicates the seminars, G\text{SIZE} is the group size, and GSEX_MAJOR the majority sex of the group: 0 = all women, 1 = all men).

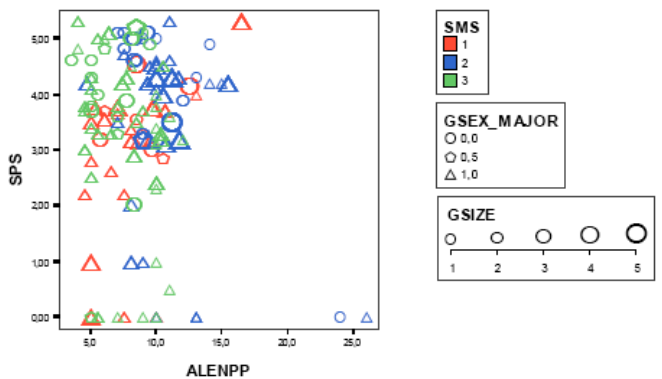


Fig. (24). Relation between article length per person (ALENPP) and seminar success (SPS) (SMS indicates the seminars, G\text{SIZE} is the group size, and GSEX_MAJOR the majority sex of the group: 0 = all women, 1 = all men).

Additionally, related to the article's length, we determined the number of references in the articles (AREFS) and tested the potential effect to SPS. That measure may approximate the effort of analyzing an article. However, there was no significant correlation between AREFS and SPS ($p > 0.05$).

4.3.3. Article Theme (ATHEM)

As noted earlier, there were three general themes: Software maintenance techniques (SPS = 3.49, $N = 117$), software maintenance tasks (SPS = 3.49, $N = 105$), and economic estimation of software maintenance (SPS = 3.76, $N = 60$). There were no statistically significant differences in student success in these regards. When looking the sub-themes more closely some significant differences appear to emerge. Students with debugging articles have clearly been the least successful (SPS = 2.04, $N = 11$): t-test, $t = 3.187$, $df = 10.5$, $p = 0.009$ (**). There were no statistically significant differences in this regard in the relevant background variables (SSEX, STEFF, SAGE, GSIZE, GEFF, GAGE, GSEX). On the other hand, students with cost estimation articles have been the most successful (SPS = 4.39, $N = 13$). The difference to others is statistically almost significant: t-test, $t = -2.372$, $df = 280$, $p = 0.018$ (*). However, the reason to this is that many of those students were women and they had also much higher general studying efficiency (STEFF) than others.

There were 117 students (41.7%) who selected technically oriented articles. Since sex was a good explaining factor of success, the article selection patterns of men and women were studied more closely. Men selected in average more technically oriented articles (ATTECH) than women. Instead, women selected more process-oriented articles. The distribution based on article themes categorized by sex for each seminar is shown in Fig. (25). EE indicates 'Economic estimation of software maintenance', TA indicates 'Software maintenance tasks', TE indicates 'Software maintenance techniques', 0 indicates women, and 1 men. While looking more closely the selection habits of technical articles between men and women it becomes clear that the difference is statistically significant: t-test, $t = -3.031$, $df = 274.0$, $p = 0.003$ (**). The difference regarding the composition of the group in terms of sex in this regard is even clearer and statistically highly significant: t-test, $t = -3.550$, $df = 271.3$, $p = 0.000$ (***)).

Additionally, we tested the potential effects of dealing with different kinds of articles in terms of their technical content and technical freshness. The revealed ancillary result was that there were no statistically significant correlation with SPS ($p > 0.05$). The tests included whether the article focused on object-orientation (AOO), C-language (AC), or COBOL-language (ACOBOL).

4.3.4. Article Formality (AMATH)

Formality of the articles is here described based on a variable AMATH defined on scale: 1 ... 5, based on the amount and complexity of mathematical expressions in each article. For testing the effects of the formal and mathematical nature of the articles the student population was divided into two groups based on the value of AMATH. 'Math-seekers': SPS = 3.49, $N = 140$. 'Math-avoiders': SPS = 3.60, $N = 142$.

The difference between the groups is not statistically significant, which means that the students were well capable of judging the sufficiency of their own mathematical capabilities while selecting the articles.

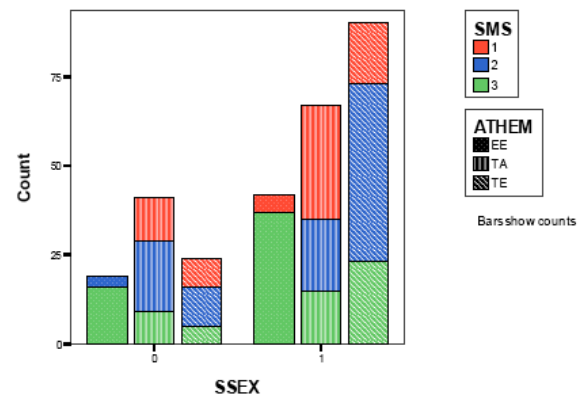


Fig. (25). Relation between student sex (SSEX: 0 = women, 1 = men) and themes of the selected articles (ATHEM: EE = economic estimation, TA = tasks, TE = techniques) (SMS indicates the seminars).

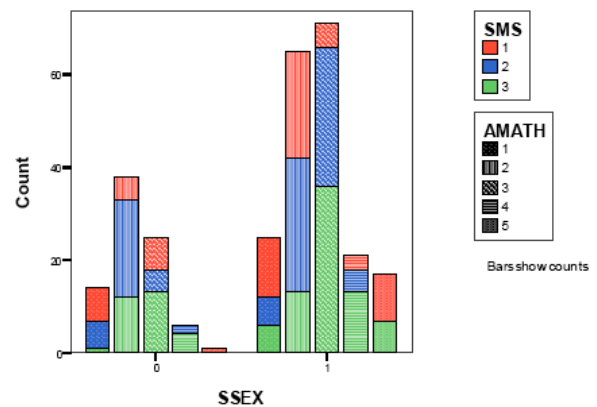


Fig. (26). Relation between student sex (SSEX: 0 = women, 1 = men) and formality of the selected articles (AMATH) (SMS indicates the seminars).

Men selected in average clearly more formal articles (AMATH = 2.70, $N = 199$) than women (AMATH = 2.31, $N = 84$). The distribution based on the formal and mathematical nature of the selected articles categorized by sex is shown in Fig. (26).

The difference between the sexes (SSEX) is here statistically significant: t-test, $t = 2.894$, $df = 281$, $p = 0.004$ (**). The difference is even clearer when considering the composition of the groups in terms of sex against the made article selections: Spearman's $\rho = 0.233$, $p = 0.000$ (***). There is also a statistically significant relation between AMATH and technically-oriented articles (ATTECH); t-test, $t = 2.672$, $df = 276.3$, $p = 0.008$ (**).

Formal articles were statistically highly significantly shorter than other articles: Spearman's ρ for the pair (AMATH, ALEN) = -0.241, $p = 0.000$ (***). Despite the obvious differences in the selections, there are no statistically significant correlations between AMATH and seminar success within the two groups: men and women, as taken separately.

4.3.5. Research Method (AMETH)

There were four general research methods applied in the analyzed articles: Empirical approach (SPS = 3.69, $N = 92$),

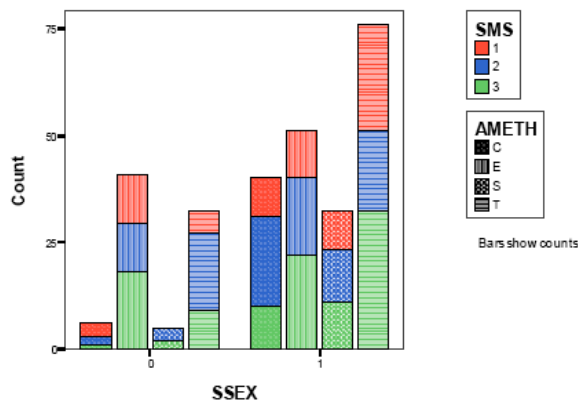


Fig. (27). Article selection habits of men and women (SSEX: 0 = women, 1 = men) regarding the research methods (AMETH: C = constructive, E = empirical, S = survey, T = theoretical) used in the articles (SMS indicates the seminars).

theory formation (SPS = 3.50, $N = 108$), constructive approach (SPS = 3.19, $N = 46$), and surveys (SPS = 3.78, $N = 36$). There were no statistically significant differences in SPS between the different methods in this regard. There were three important ancillary results regarding the article selection habits of men and women. The distribution of the students in this regard is presented in Fig. (27).

It was revealed that women selected more keenly empirical articles. There is a statistically highly significant difference regarding AMEMP and SSEX in this regard; t-test, $t = 3.678$, $df = 138.6$, $p = 0.000$ (***)). Similarly, there is a statistically highly significant difference regarding AMEMP and GSEX; t-test, $t = 4.331$, $df = 155.1$, $p = 0.000$ (***)). Another revealed ancillary result was that men selected more keenly constructive articles (AMCON); t-test, $t = -3.380$, $df = 82.4$, $p = 0.001$ (**). Similarly, there is a statistically highly significant difference regarding AMCON and GSEX; t-test, $t = -4.767$, $df = 117.3$, $p = 0.000$ (***)).

The third ancillary result was that men also selected somewhat more keenly survey-articles (AMSUR); t-test, $t = -2.892$, $df = 57.8$, $p = 0.005$ (**). The difference is statistically significant. Similarly, there is a statistically highly significant difference regarding AMSUR and GSEX; t-test, $t = -3.719$, $df = 65.0$, $p = 0.000$ (***)). There were no statistically significant differences in selecting theoretical articles between the sexes, although that they were selected clearly more often by men than articles of other categories. Survey articles were selected more often by larger groups (AMSUR, GSIZE); t-test, $t = -4.594$, $df = 42.0$, $p = 0.000$ (***)).

While looking at the article selection habits of different kinds of groups in terms of their average age (GAGE) it was noticed that there were no significant relations except in case of survey articles (AMSUR); younger students selected statistically highly significantly more keenly survey articles; t-test, $t = 6.591$, $df = 182.9$, $p = 0.000$ (***)).

5. DISCUSSION

This section first explains the focus, goals, limitations, and external relations of the study. Then the results are summarized, and suggestions for utilizing them are provided.

Their more detailed analysis follows. Finally, two kinds of sensitivity analyses are reported and possible further research directions are outlined.

5.1. Focus, Limitations, and External Relations

This paper has a very clear focus but since it also has many relevant external connections, we repeat and describe in this subsection first what the focus has been and then explain what the limitations, and main related aspects are.

5.1.1. Focus, Goals, and Main Characteristics

In this paper we have focused solely on analyzing the objective background factors related to the students, groups and articles affecting the actual measured student success in the seminars. That information is especially relevant to profiling students. We performed a strict and detailed statistical analysis of the success factors in this sense with a relatively large number of subjects. The seminars were characterized by the scientific nature of the analyzed articles, extensive coverage of the central aspects of SME, and interactivity. The general nature of the main learning objectives was to increase both knowledge about SME and skills in analyzing scientific articles. The students' task consisted of five sub-tasks. The detailed operational learning objectives were precisely determined by the predefined criteria used in evaluating the students' success. The level of success identified was analyzed based on the background factors.

5.1.2. The Seminar Based Approach as such

The approach which has been applied in this paper has been developed and reported earlier in [7]. That paper discussed the feasibility of the approach, the relationship between the general problems identified in the teaching of SME and the seminars, the teaching of SME-specific issues, the details of organizing seminars, and relation of seminars and other studies and research.

The general experiences on the seminars have been positive as reported in that paper, but this paper does not aim at repeating those findings. The seminars have in part been organized in relation to our earlier research projects, namely: Hyper-Soft (Hypertextual Software Maintenance), which is described e.g. in [168, 169], and ELTIS (Extending the Life-Time of Information Systems) [170, 171]. The archived and published seminar reports contribute to the long-term improvement of the science-based teaching of SME within FIT.

5.1.3. Different Forms of Teaching SME

The conventional "without seminar approach" has been considered by the author before the seminars were organized. The then acknowledged problems and challenges in the SME-teaching as they were described in Section 2 served as a motivation for developing the here-applied seminar-based approach. Unfortunately, we have not gathered any strictly comparative quantitative data about the pre-seminar situation. This is due to the novelty and exploratory nature of organizing the seminars and the fact that conventional lecturing and the seminars have different kinds of goals.

Basically, an SME-seminar is a group-intensive, interactive supplementary activity to the traditional course lectures. The seminar-based approach as such has been characterized in this paper based on the theoretical possibilities in Section 2 and partly based on the qualitative experiences which have been gained during the development and organization of the seminars.

We have obviously focused on studying SME-seminars. Based on our experiences on organizing the SME-seminars we can recommend their use in higher education. Especially, we recommend organizing such seminars which take into account the relevant background factors and the lessons related to their organization as presented in this paper. Since this paper has focused solely on the role of the background factors to the seminar success, it obviously does not provide additional empirical data on the methods of teaching SME. As noted above, the approach as such and our experiences have been reported earlier in [7].

For example, this study deliberately did not include the technical application of maintenance skills, since activities of that kind do not lend themselves to seminar work as described in Section 2. Those activities have been reported in other literature, whereas no earlier studies on SME seminars have been published. We have not aimed at comparing dif-

ferent forms of teaching since that would be too laborious and would complicate the research setting. Instead we have kept the teaching approach applied constant over all the experiments to be specifically able to study the background factors which invariably affect student success.

5.2. Results: Summary, Utilization, Details

The main results of this paper concern the background factors affecting student success in the SME seminars. There were many statistically highly significant relations identified between the success and the background variables, and also many unexpected results. The results have the following relations to the set hypotheses. Table 4 summarizes the valid results concerning the set hypotheses. Column 1 gives the acronyms used for the hypotheses. Column 2 shows the main general-level factor groups, i.e. student, group and article, and column 3 shows the factors. Column 4 shows whether an effect was assumed and what kinds of attribute values were assumed to lead to better success. Column 5 shows the p values in the statistical tests and column 6 the results of testing the hypotheses. Finally, column 7 shows actual effects identified and what kinds of attribute values resulted better success. Table 5 shows the ancillary results related to the hypotheses.

Table 4. Main Results Concerning the Set Hypotheses

Hypothesis	Factor Group	Factor	Assumed Effect	p (Risk Level)	Result	Actual Effect
H _{S1}	Student	Studying efficiency	Yes: More	0.000 (***)	Confirmed	Yes: More ₁
H _{S2}	Student	Sex	No	0.000 (***)	Rejected	Yes: Women
H _{S3}	Student	Age	No	0.000 (***)	Rejected	Yes: Younger
H _{S4}	Student	Focus area	Yes: CS	> 0.05	Rejected	No
H _{G1}	Group	Size	No	0.027 (*)	Rejected	Yes: Larger ²
H _{G2}	Group	Studying efficiency	Yes: More	0.000 (***)	Confirmed	Yes: More ³
H _{G3}	Group	Sex	Yes: Women	0.000 (***)	Confirmed	Yes: Women
H _{G4}	Group	Age	Yes: Younger	0.000 (*)	Confirmed	Yes: Younger
H _{A1}	Article	Series	Yes	0.028 (*)	Confirmed	Yes: Journal ₄
H _{A2}	Article	Conference	Yes: Conf.	0.028 (*)	Rejected	Yes: Journal ₄
H _{A3}	Article	Length	No	> 0.05	Confirmed	No ₅
H _{A4}	Article	Theme	No	> 0.05	Confirmed	No ₅
H _{A5}	Article	Formality	No	> 0.05	Confirmed	No ₅
H _{A6}	Article	Research method	No	> 0.05	Confirmed	No ₅

¹ The same result was received in multiple different ways of measuring the efficiency; thereby this result is very reliable. CS = computer science.

² This result is mainly due to the failed students (i.e. the students who have failed to pass the seminar altogether) who in turn were mostly in single-person groups. Thereby, this result is relatively weak.

³ This is the most reliable one of the results, the correlation is statistically highly significant even if the effects of the failed students are filtered out.

⁴ These results are relatively weak; mainly due merely to the underlying more fundamental issues of larger groups requiring journal articles and also performing better than smaller ones, and to the effects of the students who have failed to pass the seminar. Thereby these results should be interpreted with great caution.

⁵ These important negative results concern the effects of the attributes of the studied articles and confirm the expectations.

Table 5. Ancillary Results as Related to the Hypotheses

Hypo-thesis	Factor Group	Ancillary Result	<i>p</i> (Risk Level)
H _{S1}	Student	Seminar success predicted the student success in the later software engineering course examination.	0.000 (***)
H _{G1}	Group	90% of the total failures were related to the small (<3 student) groups.	N/A
H _{G2}	Group	Efficient students tended to group together.	0.000 (***)
H _{G3}	Group	Students of the same sex tended to group together (<15% were mixed groups).	0.000 (***)
H _{G4}	Group	Students with similar age tended to group together.	0.000 (***)
H _{A1}	Article	The publication year of the article did not have effect on the student success. Students with much credit units selected more keenly journal articles.	> 0.05 0.001 (**)
H _{A3}	Article	Large articles can be conveniently handled in the seminars. The number of references in the article did not have effect on the student success.	0.015 (**) > 0.05
H _{A4}	Article	Men selected more technically-oriented articles. The technical freshness of the articles did not have effect on the student success.	0.003 (**) > 0.05
H _{A5}	Article	Men selected more formal articles.	0.004 (**)
H _{A6}	Article	Women selected more empirical articles. Men selected more constructive articles. Men selected more survey-articles. Younger students selected more survey articles.	0.000 (***) 0.001 (**) 0.005 (**) 0.000 (***)

N/A= not applicable.

Appendix 3 provides additional detailed information about the factors actually explaining student success in general (Table 10), about the important real explaining factors both based on the whole data-set and separately for the organized three seminars (Table 11), and about the relations between miscellaneous factors which could have potentially explained student success (Table 12).

All the student background factors, except the focus area of the student's studies, clearly affected success. On the other hand, of the article-related factors, only the type of the publication series was relevant to success. Additionally, most group-related factors clearly affected success. All the hypotheses concerning the effect of the article-related factors, except the role of conference articles, were confirmed. On the other hand, of the student background factors only the hypothesis concerning the effect of studying efficiency was confirmed. Others, as well as the one concerning the effect of the group size were rejected. Therefore, the results related to the effect of the student-related factors were largely unexpected. The results for the group-level factors supported the results for the student-level factors.

It is useful to know which factors are relevant and which are not in order to reduce the risks related to the successful and adequate planning and organizing of similar seminars. Having the relevant information reduces the risks of unexpected effects. Most of the student- and group-related factors were highly relevant and should therefore be taken into account regardless of possible variations in seminar type within the SME field. On the other hand, most of the article-related factors are not likely obstacles to organizing seminars. Be-

low, we present a more detailed summary of the main results of the statistical analyses.

Special attention has been paid to revealing possible significant inter-correlations which could bias the results. All the relevant variables have been investigated and tested where deemed to be potential alternative explaining factors. All the significant correlations obtained are reported and all the remaining real causes of the variance of student success are discussed.

5.2.1. Student Background Effects

H_{S1} was confirmed. Both the prior general studying efficiency and studying efficiency in computing clearly positively correlated with the student's seminar success. This was a very clear and expected result. It was a little bit surprising though that the general studying efficiency was a better explaining factor than computing studying efficiency. Therefore, studying efficiency, even in very general terms, is a very relevant predictor of success.

As an ancillary result it was revealed that the seminar success was a very good indicator of the student's success in the examination of the related Software Engineering course which was taken after the seminar. Therefore, seminar success was nicely in line with the general studying success both before and after the seminar. This supports the idea that, generally speaking, similar kinds of capabilities are needed in these kinds of seminars and other typical university-level studies of computer science students. This idea is supported also by the fact that most of the other studies of these students consisted mainly of course examinations.

It may be generally necessary to limit the number of students allowed to participate to seminars because of the typical resource limitations in organizing university-level teaching. Since the student's prior general studying efficiency is both easily accessible for the instructors via student registers and a very relevant predictor of the student success, it can be used as a reliable student selection criteria in this sense. Moreover, if there is a recent software engineering course examination result available prior to the seminars, it may be even better and easier way to determine the start level of the students.

The students' studying efficiency, and other similar measures can also be used to reliably and straightforwardly support the group formation process. The groups should have relevant sufficient collective capabilities to pass the seminars. This is important, for example, because seminars are resource intensive forms of teaching and failures of individual students may risk the success of all the members in a group. Since one of the purposes of the group works is that the students can learn from each other and their strong sides will be available in synergetic sense for the group work, it is better to have in average-terms rather uniform groups regarding this background factor, than to have both exceptionally good and very poor groups. Thereby, it is beneficial that there is at least one good or very good student in each group. This is possible because there are more students than groups. That student may take the obviously needed leading role inside the group.

H_{S2} was rejected. Student sex did have significant influence on the student's seminar success; women performed clearly better. Explanation to this may be that women have better social skills which are especially important while performing group works. On the other hand, women also tended to intentionally select themselves more other kinds of studies than computer science studies. These facts suggest that the seminars generally suited very well to women. The implications for the group formation are similar as in the case of the studying efficiency; the groups should be in average-terms sufficiently uniform, so that exceptionally poor or one-sided groups can be avoided.

H_{S3} was rejected. Student age unexpectedly had significant influence on the student's seminar success; younger students performed clearly better than older ones. Experience possibly gained by age was neutralized by some other factors. One of these factors clearly was that the older students were in average in smaller groups than younger students. Smaller groups in turn performed worse. Older students appeared to be socially less connected to other students. Other potential explaining background factors include the following: older students may have less time for studies due to other responsibilities related to work and family and they may have lower motivation to studies due to better financial circumstances or poor experiences on prolonged studies. However, age had not lowered the general prior studying efficiency significantly. Therefore, the seminars did not suit especially well to older students. The groups should be formed such that their members' capabilities supplement each other. It would be beneficial, from the view-point of the general success, that the group's average ages would not vary

much, and that older students would not be isolated into singleton groups.

H_{S4} was clearly rejected. Student focus area did not have any significant influence on the student's seminar success. First of all, a very plausible reason to this is that the effect of some of the other factors is so clear that it well compensates the possible weaknesses of the individual students in this regard. Another explanation to this may be that those not majoring computer science and similar subjects, and therefore being volunteers, had specifically good reasons and interest for having selected the related Software Engineering course at the first place. This result can also be partly explained based on the fact that technically-oriented students were a clear minority within the sample, which may have had some hampering influence on group formation for them. The main lesson learnt from this issue is that, at least on the individual level, the focus area of the studies is not a matter to be worried about during the group formation phase. On the other hand, isolation of the students who have very different kind of background than the majority will likely affect success negatively; at least in some degree, and should therefore be avoided if convenient.

5.2.2. Group Effects

H_{G1} was rejected. Group size did have influence on the student's seminar success. There was synergy between most of the group members. About 90% of the failures were related to small groups. Therefore, selection of the articles and group formation are very important issues. Communication is much needed in these kinds of seminars and isolation of the students in this sense should be avoided by all possible means. It can be speculated that those who do not succeed well in the group formation phase tend to postpone the initiation of doing the seminar work and thereby increase the probability of their failure. Based on the performed statistical analysis we can seriously recommend that in the group formation the following issues are specifically considered: sufficiency of the size of the group and uniform composition of the groups in terms of the student background factors which have been shown to have effect on the students' success. This suggestion is made in order to avoid likely failures related to groups which do not have sufficient collective tendencies and capabilities to perform the seminars. Additionally, special attention should be paid on generally not allowing, or at the very least not favoring, singleton groups and by monitoring the time-table of the group formation process.

H_{G2} was confirmed; group's studying efficiency did have significant effect on the success. This means that the group-level characteristics are about as important as the individual level characteristics in this sense. On the other hand, this means that the groups were generally able to utilize their most gifted members. Otherwise the most gifted members in larger groups would not have succeeded relatively as well as they now did. Similar kind of logic in these two regards applies also to group's sex. The analysis of the group complements in general revealed that effects of the background of the rest of the group are sometimes more important than the background of the individual student. Naturally the group's size affects this so that in the larger groups the relative effect of the individual student's characteristics is smaller than in

the smaller groups. As an important ancillary result it was revealed that efficient students tend to group together and thereby the rest of the students group separately among themselves. This phenomenon has a positive effect to the high quality groups, but negative effect to the low quality groups.

H_{G3} was confirmed; group's composition in terms of its members' sex did have significant effect on the success. Since women as individuals were more successful in the seminars than men, this means that this favorable effect remained also on the group-level. So, there were no negative effects in case of groups whose clear majority or all of the members were women. As an important ancillary result it was revealed that students of the same sex strongly tend to group together. This unfortunately leads to isolation of the characteristic capabilities that men and women have.

H_{G4} was confirmed; group's composition in terms of its members' age did have significant effect on the success. This result supports the idea presented above. As an ancillary result it was revealed that students with similar age tend to group together. Therefore, all the group-related factors suggest that students with similar kinds of backgrounds tend to group together. This is a rather undesired tendency due to the problems with the low-quality groups and calls for guidance from the instructor to optimize the groups' composition.

5.2.3. Article Effects

H_{A1} was confirmed; the publication series of the articles initially seemed to have significant effect on the success. Therefore, the effects of the different kinds of articles needed to be studied and considered more thoroughly as described below.

H_{A2} was rejected. Students who analyzed conference articles did not have better seminar success than others, instead they had worse seminar success. This was an unexpected result, but can be well explained based on the performed analysis by the fact that larger groups, which were more successful than others, also selected longer articles which in turn typically were journal articles. Nevertheless, because larger groups; which were generally more successful than the others, require longer articles, we can hereby conclude that journal articles are in general rather suitable for these kinds of seminars. Since they generally also have higher quality than other kinds of articles they probably are a better choice than conference articles. Thereby, we can recommend journal articles as a basis for similar seminars. The journal articles studied in these seminars probably are good choices to be used also in other seminars in the field of SME since these seminars were successful in general.

H_{A3} was confirmed. The article length did not really have significant influence on student success although revealing this fact required taking into account the effects of group size, since longer articles typically were selected by larger groups, which also succeeded significantly better than smaller ones. As an ancillary result the study showed that it is quite possible to successfully deal with long scientific articles in these kinds of seminars as group works. There were no failures related to the articles which were above 30 pages. Thereby, the main lesson here is that the group-based semi-

nar works enable successful dealing with large materials, and that the division of work is not a significant problem.

H_{A4} was confirmed. The general article theme did not have significant influence on student's seminar success. The article sub-themes neither had any really significant relations to student's seminar success. These negative results were clear and the variances in the other relevant student background factors did not compromise the results since their correlations to this factor were not statistically significant. However, while potentially drawing further conclusions based on this issue it should be remembered that the students selected their articles relatively freely and to note that there were differences among the sexes in this regard.

As an important ancillary result it was revealed that men selected significantly more technically-oriented articles. Therefore, such articles should be offered to them. This is especially important since men succeeded generally significantly worse than women. Therefore, the main conclusion here is that it is generally important to allow the free selection of the articles by the students themselves. Otherwise, the average success might decrease significantly. However, this conclusion is based merely on the observed subjective article selection habits of men, not on the actual verified beneficiality of such articles for them in a sense of thereby reaching better seminar success since there were no such correlations.

H_{A5} was confirmed. Formality of the article did not have effect on student success on general level. The reason to this was that the students were fully capable of selecting proper articles in this regard reflecting their skills. Thereby, in these kinds of seminars, inclusion of the formal, mathematically-oriented articles into the set of the initially offered ones should not be a problem. As an ancillary result it was revealed that implicit expectations that men select significantly more formal articles were valid. The logic here concerning the suitability and selection of these articles by or for men is identical to the observation presented above concerning technically-oriented articles. Thereby, it is important to allow free selection of the articles by the students also due to this aspect.

H_{A6} was confirmed. The primary research method applied in the analyzed articles did not have any particular effect on student success. This issue shows that by allowing the free selection of the articles in this sense different articles can appear as seminar materials without compromising the general success of the seminars and students. As an important ancillary result it was revealed that men and women had significantly different kinds of article selection habits in this regard.

5.3. Sensitivity Analyses

5.3.1. Differences between the Organized Seminars

The general aspects of the organized seminars were identical. All the organized seminars were similar regarding their context, the main course, the instructor, the instructions, the procedure, the incentives, the evaluation, and the general field of study. Due to practical reasons, each group had to study a different article. Since the seminars were identical regarding the above-listed general aspects, our view is that it

is not essential in which specific seminar the articles were studied. All the studied articles nevertheless clearly represented topics within the SME-field which was the intended focus area. We purposefully aimed at covering well the whole area of the SME during the set of three seminars in order to be able to generalize the results to the area of SME. Therefore, there naturally were different kinds of articles in the sample.

The effects of the differences of the attributes of the articles were studied in our statistical analyses. Those analyses revealed that the article-related factors mainly did not have any significance to the seminar success (e.g. the article theme did not have significance). The only exception was the type of publication (i.e. conference vs journal) which in turn was of general nature and not specifically bound to any particular article. Even for that factor most of the effect was actually due to other underlying reasons.

Due to the reasons presented above, we strongly feel that there have been no significant research methodological problems in merging the data sets of the studied three seminars as has been done. The merge has been conducted purposefully in order to process more information about those factors for which the initial seminar did not show clear results.

However, in order to facilitate transparency of the performed analysis, additional information has been provided. Firstly, all the figures have presented the data in such a way that both the combined data and the differences between the three seminars can be observed visually on general level. Secondly, as a background information the main statistical descriptives of the studied scale variables have been presented also for each seminar as taken separately in Appendix 1 (Tables 6, 7, and 8). Thirdly, an analysis of the differences between the organized seminars regarding the received correlations between the student success and the determined real explaining factors is provided in Appendix 3 (Table 11).

5.3.2. Student's Attitude and Motivation

We have not measured students' attitudes. It is obvious that at least the individual student's general attitude towards the seminar (i.e. his/her motivation or seriousness) may have an implicit effect to the student's success. Students who had poorer attitude probably succeeded worse than the rest.

Our view is that the student's motivation is basically an intermediary variable between the other measured, objective, and more stable background factors and the student's actual success in the seminar. The main problem with motivation as such is that its reliable measurement would be hard to achieve in practice.

Firstly, motivation is a subjective measure. Its use would probably not increase the accuracy of the results and it might even decrease it, because the students would less likely report low levels of motivation in order to seem good candidates, or would not be able to estimate their level of motivation accurately at the first place due to their lack of earlier experience.

Secondly, it should be noted that all the data which was actually gathered here was objective. Therefore, students' attitudes did not have a potential biasing effect to the results

in this sense. Instead of taking the risk of filtering out students from the data set based on their assumed level of motivation, we have decided to honestly present the results based on the whole data set.

Thirdly, the motivation level might change during the seminars, in which case the initial measurement would not accurately describe the actual attitude. Fourthly, it could be argued that in this case a proper measure of the motivation is actually the student's final success in the seminar, which however was in these analyses a dependent, not an independent variable.

Since the worst possible effect of the low level of motivation would be a failure to pass the seminar, we have additionally performed an analysis of the effects of the failed students to the general main results. This "sensitivity" analysis has been achieved by performing the main correlation analyses both based on the whole sample and based solely on the students who have passed the seminar (Appendix 3, Table 10).

The amount of the failed students has been 20 representing 6.8% of the whole data set. It is obvious that despite the relative low proportion of the failed students, they have a noticeable effect to the average success since they represent one of the far ends of the distribution. The analysis nevertheless revealed that for most of the factors (Geff, Gceff, Gcage, Gage, Steff, Gsex, Sage, Ssex, Gcsex, and Ajour) most of the effect was due to the passed students.

For some of the factors (SPE_EXE, SCEFF, GSIZE, and ALEN) most of the effect was due to the failed students. Of the factors in the latter category SPE_EXE is not used to predict success, ALEN is not significant in the final analysis regardless of the inclusion/exclusion of the failed students, and SCEFF can be conveniently substituted with STEFF, which is not particularly problematic in this sense.

Therefore, the group size (GSIZE) is clearly the most affected one of the factors, which really affect student's seminar success by the inclusion of the failed students into the data set. For that variable only 19% of the variation of the correlation was due to the passed students. In practice this means that most of the failed groups were 1-person groups (Figure 14), and success of most of the passed 1-person groups was inferior to the average success of the students in the whole sample.

This means that the low value of group size (i.e. the case of 1-person groups) has an especially strong effect to the success, whereas the variation above that group size is much lower. To know whether the small group size was the fundamental reason to the low level of success of the failed students or the low level of motivation to the small group size of the failed students would require separate new studies and gathering more information on the group formation process

5.4. Potential Further Research Directions

5.4.1. Independent Replication of the Seminars

Since these seminars were successful, we can recommend their independent replication in the SME-context by others. The detailed descriptions of the seminars and the ex-

periments in this paper support this option. Although the results received so far regarding most of the studied issues are clear, studying new seminars could further consolidate and extend the present empirical basis of teaching SME successfully.

Individual seminars can be arranged by single research groups by themselves, but in practice comparative studies between different kinds of seminars would require cooperation between multiple research groups. At least, in our case it has been quite impossible to perform ambitious additional comparative studies due to their elaborate nature and the typical obvious teaching resource limitations. However, we feel that our current results may provide some support also to that kinds of studies since we have focused on the general background factors which probably need to be taken into account in the studies which have more explaining variables to be controlled in cases of predicting student success regardless of the exact focus of those experiments.

5.4.2. General Software Engineering Seminars

Studies on the possible similarities and differences of these success factors and those of seminars in general software engineering context could be useful for identifying and validating whether the general nature of the here revealed success factors means that they are valid also in the software engineering field in general.

5.4.3. Applicative SME-Capabilities

Studying applicative capabilities of SME-students is a branch of research separate from seminars since the seminars had theoretical emphasis. Such studies could however provide interesting supplementing information for the general organization of the teaching of SME. Comparative studies would be useful but they would require relatively large number of subjects. Especially reliable comparison of the effectiveness of different forms of teaching; such as seminars and applicative exercises, would be very challenging since they have distinct educational purposes.

6. CONCLUSIONS

Software maintenance and evolution is a wide, economically important, and nowadays also rapidly advancing sub-field of software engineering. In the optimal situation, university-level computer science students could be provided with versatile and up-to-date scientific SME knowledge sufficiently early during their studies in an interactive and motivating way.

Attainment of this ideal on the basis of the currently available educational materials, within the typical curricula constraints, and by lecture- or text-book-based approaches is an almost impossible task. On the other hand, seminar-based approaches are basically well-suited to the purpose. This situation has emphasized the need to study SME seminars empirically.

The general feasibility of the here applied seminar-based approach as such to the teaching of SME has been confirmed already earlier [7]. Basically, SME-seminars enable supplementing the traditional course lectures with a group-based

interaction focusing on scientifically important or current issues. Related to that earlier study, we identified both a need and an opportunity to explore in detail the background factors which invariably affect student success in these kinds of seminars.

This paper is an example of empirical exploratory basic research and has focused on studying the background factors affecting student success in the three organized SME-seminars. To best of our knowledge, our studies are the only scientific studies on SME seminars. Moreover, this particular study is also the first strict statistical analysis of these issues in the larger context of software engineering. We have illustrated our main findings such that both the whole combined data set and the differences between the seminars can be seen.

The main results concerned the background factors which affected each individual student's actual measured success in the seminars. Fourteen main hypotheses concerning student, group, and article-related factors were set and tested. Most of the results were very clear and statistically significant. Most of the results on the effect of the student background factors were unexpected, whereas most of the results on the effect of the article-related factors were as expected. Most of the student and group-related factors had a significant influence on success whereas most of the article-related factors did not.

These results mean that it is very important to pay attention to the group formation process, and that most scientific articles on SME are likely to be adequate seminar material. Consequently, it would be beneficial to be able to classify students prior to the group formation process according to their capabilities in terms of predicting success. On the other hand, there is not much need to analyze the convenience of articles, despite numerous potential differences between them.

The most important background factors actually affecting student success were: 1) the group's prior general studying efficiency, 2) the group's average age, 3) the student's prior general studying efficiency, 4) the sex composition of the group, 5) student's age, 6) student's sex, and 7) small group size. The factors affecting success were thus related especially to studying efficiency, age, sex, and small group size.

The main factors related to the background of individual students were as follows. A student's general prior studying efficiency correlated highly positively with seminar success. Seminar success and success in the related course examination also correlated strongly, which underlines the usefulness of profiling students on the basis of these generally highly-available measures. These results also support the idea that similar capabilities are needed in these kinds of seminars and in other typical university-level courses in this area.

Women were clearly more successful than men. They were more successful also in their other studies, but especially in the seminars. Thereby, the seminars were especially well suited to women. Younger students were also more successful than older ones.

Group composition clearly affected student success. The group contributed as a whole to the success of each of its members. Likewise, as in case of the individual student's

background, average prior studying efficiency and age of the group were good predictors of success. These factors, also as determined for the rest of the group, were even somewhat better predictors than the student factors on the individual level. This underlines the importance of group characteristics. The sex composition of the group also clearly affected success.

Most of the failures were related to small groups. Therefore it is particularly important to avoid the formation of too small groups and beneficial to form or instruct the groups to be formed such that they are not biased in terms of the characteristics and capabilities of their members which are relevant to success in the seminar. As an important ancillary result it was found out that students with a similar background tend to group together by themselves. This emphasizes the need to guide the group formation process in order to avoid too widely varying groups.

Most of the article-related factors did not have much relevance to success. The most important ancillary result concerning the article-related issues was that men were very keen to select formal and technical articles. This and other similar observations underline the importance of allowing students to freely select their articles.

Generally speaking, we suggest that guidance and control should be focused on the group formation process, whereas article selection choices can and should mainly be delegated to the students in these kinds of seminars. The students were very capable of selecting convenient articles but less capable of forming and maintaining good groups.

The general background information profiling students can be used in avoiding the formation of groups that are likely to be unsuccessful and to enable the selection of promising students to be given priority to participate in seminars at the first place.

The results of this paper support, in particular, paying attention to issues which can be expected to be generally relevant to student success in these kinds of seminars. Knowing those issues is important when organizing and managing those seminars.

Due to the general nature of the studied factors, the results may also have some value as indispensable background information for other studies. For example, comparisons of different kinds of seminars and courses may benefit from the present investigation since it focused on analyzing the generally relevant background factors which affect success and which therefore invariably need to be taken into account when conducting empirical experiments related to software engineering seminars.

ACKNOWLEDGMENTS

The participation of the nearly 300 students to these seminars has made this paper possible. We also wish to thank the anonymous reviewers for their constructive comments, which have helped to enhance the quality of the paper. Mr. Jukka Kesonen and Mr. Pekka Rahkonen were instrumental by providing expert guidance regarding the use of statistical methods and the features of SPSS. Useful general comments were received also from Mr. Veikko Halttunen.

APPENDIX 1: CHARACTERISTICS OF THE STUDIED MAIN VARIABLES

Table 6. Characteristics of the Student-Related Variables^{1,2}

Factor	<i>N</i>	min.	max.	mean	std. dev.	Explanation
SPS	282	0	5.3	3.55	1.33	Seminar points
(SMS-I)	73	0	5.3	3.19	1.24	
(SMS-II)	104	0	5.3	3.77	1.34	
(SMS-III)	105	0	5.3	3.57	1.33	
SSEX	283	0	1	0.70	0.46	Sex
(SMS-I)	74	0	1	0.73	0.45	
(SMS-II)	104	0	1	0.67	0.47	
(SMS-III)	105	0	1	0.71	0.45	
SPE_EXE	283	0	22.7	11.2	5.36	Received points in the related course examination without the seminar bonus
(SMS-I)	74	0	22.7	11.7	6.13	
(SMS-II)	104	0	20.0	11.2	5.20	
(SMS-III)	105	0	18.2	10.8	4.93	
SCUT	269	19	335	150	47.4	Credit units; total
(SMS-I)	67	19	301	161	55.5	
(SMS-II)	98	60	335	157	45.8	
(SMS-III)	104	38	230	136	39.7	

Table 6. contd...

Factor	N	min.	max.	mean	std. dev.	Explanation
SCUC	269	5	159	76	26.1	Credit units; computing
(SMS-I)	67	5	120	77	29.0	
(SMS-II)	98	21	159	79	25.7	
(SMS-III)	104	17	140	73	24.5	
SM	269	0	1	0.18	0.38	Study area; whether the student was from some other university
SIT	269	0	1	0.92	0.27	Study area; whether the student was from some other faculty
SMAJ	283	N/A	N/A	N/A	N/A	Major subject (15 different disciplines, listed in text)
STECH	269	0	1	0.19	0.40	Study area; some technical discipline
(SMS-I)	67	0	1	0.13	0.34	
(SMS-II)	98	0	1	0.24	0.43	
(SMS-III)	104	0	1	0.18	0.39	
SLINE	283	N/A	N/A	N/A	N/A	Study focus area (15 different lines, listed in text)
STEFF	269	3	91	30.0	11.8	General prior studying efficiency
(SMS-I)	67	4	54	24.4	8.9	
(SMS-II)	98	3	63	30.3	10.1	
(SMS-III)	104	7	91	33.3	13.6	
SCEFF	269	0	45	15.6	7.10	General prior computing studying efficiency
(SMS-I)	67	0	22	12.0	5.35	
(SMS-II)	98	1	28	15.4	5.87	
(SMS-III)	104	2	45	18.2	8.07	
SAGE	271	21	44	24.3	3.74	Age
(SMS-I)	68	21	35	24.6	3.20	
(SMS-II)	99	21	44	23.6	3.02	
(SMS-III)	104	21	43	24.8	4.52	
SYCS	269	1	25	2.6	2.36	Year course
(SMS-I)	67	1	15	2.8	2.30	
(SMS-II)	98	0	25	2.6	2.68	
(SMS-III)	104	0	19	2.5	2.06	
AINDEX	283	N/A	N/A	N/A	N/A	The article selected by a specific student

¹ The subrows show the information for each variable as calculated in the following manner: 1) for the whole sample, 2) for the SMS-I, 3) for the SMS-II, and 4) for the SMS-III.

² N = number of the valid cases, N/A = not applicable (e.g. all statistical descriptives are not meaningful for nominal variables).

Table 7. Characteristics of the Group-Related Variables ^{1,2}

Factor	N	min.	max.	mean	std. dev.	Explanation
GSIZE	283	1	5	2.7	1.09	Amount of the group members
(SMS-I)	74	1	4	2.9	0.87	
(SMS-II)	104	1	5	3.1	1.33	
(SMS-III)	105	1	4	2.2	0.75	

Table 7. contd...

Factor	N	min.	max.	mean	std. dev.	Explanation
GEFF	282	3	68	30.0	9.56	Average general prior studying efficiency of the members
(SMS-I)	73	12	38	24.5	6.62	
(SMS-II)	104	3	63	30.6	8.34	
(SMS-III)	105	10	68	33.4	10.70	
GSEX	283	0.0	1.0	0.7	0.39	Composition in terms of the sex of the members
(SMS-I)	74	0.0	1.0	0.73	0.36	
(SMS-II)	104	0.0	1.0	0.67	0.41	
(SMS-III)	105	0.0	1.0	0.71	0.40	
GAGE	282	21	44.0	24.3	3.16	Average age of the members
(SMS-I)	74	21.0	31.0	24.7	2.39	
(SMS-II)	103	21.0	44.0	23.6	2.62	
(SMS-III)	105	22.0	40.0	24.8	3.93	
GCEFF	269	0	70	26.7	13.1	Group complement for the studying efficiency
(SMS-I)	67	0	40	24.2	8.0	
(SMS-II)	98	0	60	27.2	12.4	
(SMS-III)	104	0	70	27.9	16.0	
GCSEX	252	0.0	1.0	0.69	0.42	Group complement for the sex composition
(SMS-I)	73	0.0	1.0	0.73	0.40	
(SMS-II)	90	0.0	1.0	0.68	0.40	
(SMS-III)	89	0.0	1.0	0.67	0.44	
GCAGE	241	21	43.0	24.1	3.17	Group complement for the age
(SMS-I)	67	21.0	35.0	24.7	2.72	
(SMS-II)	86	21.0	27.5	23.1	1.51	
(SMS-III)	88	21.0	43.0	24.7	4.29	

¹ The subrows show the information for each variable as calculated in the following manner: 1) for the whole sample, 2) for the SMS-I, 3) for the SMS-II, and 4) for the SMS-III. ² N = number of the valid cases.

Table 8. Characteristics of the Article-Related Variables ^{1,2}

Factor	N	min.	max.	mean	std. dev.	Explanation
AY	283	1985	2002	1996	4.43	Publication year (shown in the references)
(SMS-I)	74	1986	1999	1994	3.19	
(SMS-II)	104	1985	2002	1994	4.50	
(SMS-III)	105	1987	2002	1999	3.54	
AFOR	283	N/A	N/A	N/A	N/A	Publication forum (33 different forums, see Appendix 2, Table 9)
AJOUR	283	0	1	0.71	0.46	Publication series (journal or other) (can be determined based on the reference information).
(SMS-I)	74	0	1	0.78	0.41	
(SMS-II)	104	0	1	0.88	0.33	
(SMS-III)	105	0	1	0.49	0.50	

Table 8. contd....

Factor	N	min.	max.	mean	std. dev.	Explanation
ALEN	283	4	66	23.9	14.4	Length in pages (specified in the reference list)
(SMS-I)	74	9	66	24.7	14.3	
(SMS-II)	104	9	62	30.9	15.6	
(SMS-III)	105	4	34	16.2	8.0	
ALENPP	283	4	26	8.7	3.00	Length per group size
(SMS-I)	74	5	17	8.3	2.94	
(SMS-II)	104	5	26	10.3	2.97	
(SMS-III)	105	4	12	7.2	2.15	
AREFS	278	6	102	34	21.8	Number of references
(SMS-I)	74	9	81	31	17.5	
(SMS-II)	103	11	102	42	27.5	
(SMS-III)	101	6	60	27	14.0	
ATHEM	283	N/A	N/A	N/A	N/A	Theme (3 main themes: ATTECH, ATTASK, ATECON) (specified in the reference list)
ATSUB	283	N/A	N/A	N/A	N/A	Sub-theme (11 different) (specified in the reference list)
AMETH	283	N/A	N/A	N/A	N/A	Applied main reseach method (4 different: AMEMP, AMCON, AMTHE, AM-SUR) (specified in the reference list)
AMATH	283	1	5	2.6	1.05	Formality (specified in the reference list)
(SMS-I)	74	1	5	2.4	1.34	
(SMS-II)	104	1	4	2.4	0.78	
(SMS-III)	105	1	5	2.9	0.97	
AOO	283	0	1	0.18	0.38	Dealing with object-orientation
(SMS-I)	74	0	1	0.19	0.39	
(SMS-II)	104	0	1	0.18	0.38	
(SMS-III)	105	0	1	0.16	0.37	
AC	283	0	1	0.21	0.41	Dealing with C-language
(SMS-I)	74	0	1	0.38	0.49	
(SMS-II)	104	0	1	0.21	0.41	
(SMS-III)	105	0	1	0.09	0.28	
ACOBOL	283	0	1	0.14	0.35	Dealing with COBOL-language
(SMS-I)	74	0	1	0.14	0.34	
(SMS-II)	104	0	1	0.14	0.35	
(SMS-III)	105	0	1	0.13	0.34	

¹ The subrows show the information for each variable as calculated in the following manner: 1) for the whole sample, 2) for the SMS-I, 3) for the SMS-II, and 4) for the SMS-III.

² N = number of the valid cases, N/A = not applicable (e.g. all statistical descriptives are not meaningful for nominal variables).

APPENDIX 2: PUBLICATION SERIES

The analyzed articles have appeared in the different publication series as listed in Table 9. Columns 1-2 show the name of the series and its used acronym. Column 4 shows the numbers of articles (N_A), and column 5 the number of students (N_S) who have selected the articles. Columns 6-8 show the number of students who have selected the articles related to the specific seminars (SMS-I, SMS-II, and SMS-III). The studied articles as such are listed in the reference list.

Table 9. Publication Series in Descending Order of the Amount of the Articles in the Seminars ¹

Publication Series	Acronym	Publisher	N_A	N_S	$N_{S-SMS-I}$	$N_{S-SMS-II}$	$N_{S-SMS-III}$
<i>Software Maintenance / Journal of Software Maintenance (and Evolution: Research and Practice)</i>	SM / JSM (JSME)	Wiley	21	50	9	13	28
<i>Proceedings of the International Conference on Software Maintenance</i>	ICSM	IEEE	20	32	3	1	28
<i>IEEE Transactions on Software Engineering</i>	TSE	IEEE	10	19	6	8	5
<i>Proceedings of the International Conference on Software Engineering</i>	ICSE	IEEE, ACM	9	18	6	2	10
<i>International Journal of Human-Computing Studies / International Journal of Man-Machine Studies</i>	IJHCS / IJMMS	Elsevier	7	16	6	10	0
<i>Communications of the ACM</i>	CACM	ACM	6	11	0	5	6
<i>Advances in Computers</i>	AIC	Elsevier	5	22	12	10	0
<i>Software – Practice and Experience</i>	SPE	Wiley	5	15	3	12	0
<i>Empirical Studies of Programmers</i>	ESP	Ablex	5	11	4	7	0
<i>Journal of Systems and Software</i>	JSS	Elsevier	5	8	5	1	2
<i>Working Conference on Reverse Engineering</i>	WCRE	IEEE	4	4	0	1	3
<i>Software Engineering: A Practitioner’s Approach</i>	SEPA	McGraw-Hill	3	8	0	0	8
<i>IEEE Software</i>	-	IEEE	3	5	0	1	4
<i>IEEE Computer</i>	-	IEEE	3	5	4	0	1
<i>ACM Transactions on Software Engineering and Methodology</i>	TOSEM	ACM	2	8	4	4	0
<i>International Journal of Software Engineering and Knowledge Engineering</i>	IJSEKE	World Scientific	2	6	0	6	0
<i>Journal of Visual Languages and Computing</i>	JVLC	Elsevier	1	5	0	5	0
<i>Cognitive Psychology</i>	-	Elsevier	1	5	0	5	0
<i>Interacting with Computers</i>	-	Elsevier	1	4	4	0	0
<i>Computing Systems</i>	-	USENIX	1	4	0	4	0
<i>Management Science</i>	-	INFORMS	1	3	3	0	0
<i>Information Sciences</i>	-	Elsevier	1	3	0	3	0
<i>Annals of Software Engineering</i>	-	Springer	1	3	0	0	3
<i>Global and Multiple Criteria Optimization</i>	-	Åbo Akademi	1	3	0	0	3
<i>Colloquium on Combining Paradigms for Software Development</i>	-	Springer	1	3	3	0	0
<i>Information Processing & Management</i>	-	Elsevier	1	2	0	0	2
<i>ACM Transactions on Programming Languages and Systems</i>	TOPLAS	ACM	1	2	0	2	0
<i>The Computer Journal</i>	-	Oxford Journals	1	2	2	0	0
<i>International Conference on Information Systems</i>	ICIS	AIS	1	2	0	0	2
<i>Journal of Object-Oriented Programming</i>	JOOP	SIGS Public.	1	1	0	1	0
<i>Software – Concepts and Tools</i>	-	Springer	1	1	0	1	0
<i>International Workshop on Program Comprehension</i>	IWPC	IEEE	1	1	0	1	0
<i>IBM Center for Advanced Studies Conference</i>	CASCON	IBM	1	1	0	1	0
Total:			127	283	74	104	105

¹ N_A = number of the articles, N_S = number of the students. IEEE = IEEE Computer Society, ACM = ACM Press

APPENDIX 3: MAIN RELEVANT CORRELATIONS

Table 10. Main Significant Correlations between Student Success (SPS) and the Explaining Factors for the whole Data-Set and for the Passed Students in Descending General Order of the Significance¹

Factor	N	Spearman's rho or Student's t^2	p (Risk Level)	Effect of the Passed Students / Total Effect	Notes
GEFF (passed)	281 262	$\rho = 0.336$, $\rho = 0.257$	0.000 (***) 0.000 (***)	76.5 %	Real explaining factor.
GCEFF (passed)	268 249	$\rho = 0.277$, $\rho = 0.198$	0.000 (***) 0.002 (**)	71.4 %	Group complement effects; are included in GEFF.
GCAGE (passed)	240 230	$\rho = -0.266$, $\rho = -0.259$	0.000 (***) 0.000 (***)	97.4 %	Group complement effects; are included in GAGE.
GAGE (passed)	281 262	$\rho = -0.264$, $\rho = -0.245$	0.000 (***) 0.000 (***)	92.8 %	Real explaining factor.
STEFF (passed)	268 249	$\rho = 0.253$, $\rho = 0.153$	0.000 (***) 0.015 (*)	60.5 %	Real explaining factor. STEFF is very similar to SCEFF but easier to determine and has better significance for the passed students.
SPE_EXE (passed)	282 262	$\rho = 0.253$, $\rho = 0.126$	0.000 (***) 0.042 (*)	49.8 %	SPE_EXE does not (and is not intended to) predict SPS since SPS precedes SPE_EXE in time.
GSEX (passed)	282 262	$\rho = -0.250$, $\rho = -0.222$	0.000 (***) 0.000 (***)	88.8 %	Real explaining factor.
SAGE (passed)	270 251	$\rho = -0.223$, $\rho = -0.215$	0.000 (***) 0.001 (**)	96.4 %	Real explaining factor.
SSEX (passed)	282 262	$t = 4.563$, $t = 3.890$	0.000 (***) 0.000 (***)	85.3 %	Real explaining factor.
GCSEX (passed)	251 241	$\rho = -0.203$, $\rho = -0.190$	0.001 (**) 0.003 (**)	93.6 %	Group complement effects; are included in GSEX.
SCEFF (passed)	268 249	$\rho = 0.193$, $\rho = 0.071$	0.002 (**) 0.262	36.8 %	Real explaining factor, if failed students are included.
ALEN (passed)	282 262	$\rho = 0.145$, $\rho = 0.059$	0.015 (*) 0.340	40.7 %	Does not explain SPS since the effect is mainly due to GSIZE, and the potential remaining effect is mainly caused only by the failed students.
GSIZE (passed)	282 262	$\rho = 0.132$, $\rho = 0.025$	0.027 (*) 0.691	19.0 %	Real explaining factor, but only if failed students are included.
AJOUR (passed)	282 262	$t = -2.218$, $t = -1.253$	0.028 (*) 0.213	56.5 %	Some significance if failed students are included, but does not explain SPS well since e.g. the effect is partly due to GSIZE.

¹ The subrows show the information for each variable as calculated in the following manner: 1) for the whole sample, and 2) for the students who have passed the seminar (i.e. to those whose SPS > 0).

² Spearman's ρ has been calculated for the cases where the factor is a scale variable, and Student's t ; e.g. for the cases where the factor is a nominal variable.

Table 11. Main Significant Correlations between the Student Success (SPS) and the Determined Real Explaining Factors Separately for the whole Data-Set and the Organized three Seminars in Descending General Order of the Significance¹

Factor	N	Spearman's rho or Student's t^2	p (Risk Level)	Notes
GEFF (SMS-I) (SMS-II) (SMS-III)	281 73 103 105	$\rho = 0.336$ $\rho = 0.283$ $\rho = 0.266$ $\rho = 0.268$	0.000 (***) 0.015 (*) 0.007 (**) 0.006 (**)	All the component p -values (i.e. for the SMS-I, SMS-II, and SMS-III) have significance, but the combined data for the whole sample provides even clearer results.
GAGE (SMS-I) (SMS-II) (SMS-III)	281 73 103 105	$\rho = -0.264$ $\rho = 0.019$ $\rho = -0.283$ $\rho = -0.245$	0.000 (***) 0.872 0.004 (**) 0.012 (*)	SMS-I initially did not show significance, but SMS-II and SMS-III did. The combined data provides even clearer results.
STEFF (SMS-I) (SMS-II) (SMS-III)	268 66 98 104	$\rho = 0.253$ $\rho = 0.166$ $\rho = 0.162$ $\rho = 0.193$	0.000 (***) 0.183 0.111 0.049 (*)	The component p -values as taken separately are low, but the combined data nevertheless succeeds to show the significance very well in this case.
GSEX (SMS-I) (SMS-II) (SMS-III)	282 73 104 105	$\rho = -0.250$ $\rho = -0.301$ $\rho = -0.242$ $\rho = -0.304$	0.000 (***) 0.010 (***) 0.013 (*) 0.002 (**)	All the component p -values have significance, and the combined data provides even clearer results.
SAGE (SMS-I) (SMS-II) (SMS-III)	270 67 99 104	$\rho = -0.223$ $\rho = -0.052$ $\rho = -0.200$ $\rho = -0.203$	0.000 (***) 0.674 0.047 (*) 0.038 (*)	SMS-I initially did not show any significance. Also the p -values for the SMS-II and SMS-III are relatively low. However, the combined data provides very clear results.
SSEX (SMS-I) (SMS-II) (SMS-III)	282 73 104 105	$t = 4.563$ $t = 3.059$ $t = 2.494$ $t = 2.140$	0.000 (***) 0.003 (**) 0.014 (*) 0.035 (*)	All the component p -values have significance, but the combined data provides even much clearer results.
SCEFF (SMS-I) (SMS-II) (SMS-III)	268 66 98 104	$\rho = 0.193$ $\rho = 0.241$ $\rho = 0.050$ $\rho = 0.157$	0.002 (**) 0.051 0.627 0.111	None of the component p -values were significant. The combined data manages nevertheless to provide a decent level of significance.
ALEN (SMS-I) (SMS-II) (SMS-III)	282 73 104 105	$\rho = 0.145$ $\rho = 0.573$ $\rho = -0.154$ $\rho = 0.007$	0.015 (*) 0.000 (***) 0.119 0.940	SMS-I initially did show a highly significant p -value. However, that result could not be replicated in SMS-II or SMS-III. Even the combined data shows only a relatively low level of significance.
GSIZE (SMS-I) (SMS-II) (SMS-III)	282 73 104 105	$\rho = 0.132$ $\rho = 0.538$ $\rho = -0.133$ $\rho = 0.162$	0.027 (*) 0.000 (***) 0.177 0.099	Only the SMS-I did show a significant p -value. The combined data shows some significance.
AJOUR (SMS-I) (SMS-II) (SMS-III)	282 73 104 105	$t = -2.218$ $t = -2.231$ $t = 0.737$ $t = -2.289$	0.028 (*) 0.037 (*) 0.463 0.024 (*)	SMS-I and SMS-III showed some significance, but SMS-II did not. Even the combined data shows only a relatively low level of significance.

¹ The subrows show the information for each variable as calculated in the following manner: 1) for the whole sample, 2) for the SMS-I, 3) for the SMS-II, and 4) for the SMS-III.² Spearman's ρ has been calculated for the cases where the factor is a scale variable, and Student's t ; e.g. for the cases where the factor is a nominal variable.

Table 12. Main Significant Relevant Correlations between the Factors Potentially Explaining Student Success (SPS) in Descending General Order of the Significance

Factor 1	Factor 2	Spearman's ρ or Student's t ¹	p (Risk Level)	Notes on the Underlying Reasons
ALEN	GSIZE	$\rho = 0.859$	0.000 (***)	Larger groups selected longer articles as instructed.
GEFF	STEFF	$\rho = 0.792$	0.000 (***)	Relation by definition.
SAGE	GAGE	$\rho = 0.783$	0.000 (***)	Relation by definition.
STEFF	SCEFF	$\rho = 0.782$	0.000 (***)	Relation by definition.
SAGE	GCAGE	$\rho = 0.412$	0.000 (***)	Students with similar age grouped together.
STEFF	GCEFF	$\rho = 0.341$	0.000 (***)	Students with similar studying efficiency grouped together.
AMATH	ALEN	$\rho = -0.241$	0.000 (***)	Formal articles were shorter.
AMATH	GSEX	$\rho = 0.233$	0.000 (***)	Groups with more men selected more formal articles.
ALEN	SAGE	$\rho = -0.232$	0.000 (***)	Longer articles were selected by younger students.
STEFF	SPE_EXE	$\rho = 0.221$	0.000 (***)	Students with better studying efficiency succeeded better in the oncoming course examination.
GEFF	GSEX	$\rho = -0.213$	0.000 (***)	Groups with more women had better studying efficiency.
ALEN	GAGE	$\rho = -0.207$	0.000 (***)	Groups with older students selected shorter articles.
SSEX	GSEX	$t = 23.752$	0.000 (***)	Relation by definition.
AJOUR	ALEN	$t = -11.054$	0.000 (***)	Journal articles were longer.
SSEX	GCSEX	$t = -10.925$	0.000 (***)	Students grouped mainly according to their sex.
AJOUR	GSIZE	$t = -7.854$	0.000 (***)	Journal articles were selected by larger groups.
GAGE	AMSUR	$t = 6.591$	0.000 (***)	Younger students selected more survey articles.
AMSUR	ALEN	$t = -5.020$	0.000 (***)	Survey articles were longer.
AMCON	GSEX	$t = 4.767$	0.000 (***)	Groups with more men selected more constructive articles.
AMSUR	GSIZE	$t = -4.594$	0.000 (***)	Survey articles were selected by larger groups.
AMEMP	GSEX	$t = 4.331$	0.000 (***)	Groups with more women selected more empirical articles.
AMSUR	GSEX	$t = -3.719$	0.000 (***)	Groups with more men selected more survey articles.
AMEMP	SSEX	$t = 3.678$	0.000 (***)	Women selected more empirical articles.
ATTECH	GSEX	$t = -3.550$	0.000 (***)	Groups with more men selected more technical articles.
GSIZE	SAGE	$\rho = -0.209$	0.001 (**)	Larger groups contained less older students.
AJOUR	SCUT	$t = -0.363$	0.001 (**)	Students with much credit units selected journal articles.
AMCON	SSEX	$t = 3.380$	0.001 (**)	Men selected more constructive articles.

¹ Spearman's ρ has been calculated for the cases where both factors are scale variables, and Student's t for the cases where one or both of the factors are nominal variables.

REFERENCES^{1,2}

- [1] J. Koskinen, "Software maintenance fundamentals", in *Encyclopedia of Software Engineering*, P. Laplante, Ed., Taylor & Francis Group, 2009 (accepted, in press).
- [2] B. Lientz, *Software Maintenance Management: A Study of the Maintenance of Computer Application Software in 487 Data Processing Organizations*, Addison-Wesley, 1980.
- [3] S. Huff, "Information systems maintenance", *The Business Quarterly*, vol. 55, pp. 30-32, 1990.
- [4] L. Erlikh, "Leveraging legacy system dollars for E-business", *(IEEE) IT Pro*, vol. 2(3), pp. 17-23, 2000.
- [5] R. Seacord, D. Plakosh, and G. Lewis, *Modernizing Legacy Systems: Software Technologies, Engineering Processes, and Business Practices*, Addison-Wesley, 2003.
- [6] M. Lehman, D. Perry, and J. Ramil, "Implications of evolution metrics on software maintenance", in *ICSM*, pp. 208-217, 1998.

¹ Annotations for the seminar articles:

{Seminar#,Theme:Subtheme,Research method,Length,Formality}

² Publication forum acronyms for the seminar articles:

These have been specified in Appendix 2, Table 9.

- [7] J. Koskinen, "Experiences from software maintenance seminars: Organizing three seminars with 127 groups", in *International Conference on Engineering Education, Instructional Technology, Assessment, and E-learning (EIAE 2008)* (as part of the *CISSE 2008*), 2009 (presented, in press).
- [8] A. Abran, J.W. Moore, P. Bourque, and R. Dupuis, Eds. *SWEBOK: Guide to the Software Engineering Body of Knowledge 2004 Version*, IEEE Computer Society, 2004. [E-book] Available: http://www.swebok.org/ironman/pdf/SWEBOK_Guide.2004.pdf [Accessed Aug. 14, 2007].
- [9] *Computing Curricula 2001 Computer Science: Final Report (15 December 2001)*, The Joint Task Force on Computing Curricula: IEEE Computer Society, Association for Computing Machinery, 2001. [E-Book] Available: http://www.computer.org/portal/cms_docs_ieeeecs/ieeecs/education/cc2001/cc2001.pdf [Accessed Nov. 28, 2006].
- [10] *Software Engineering 2004: Curriculum Guidelines for Undergraduate Degree Programs in Software Engineering: A Volume of the Computing Curricula Series (23 August 2004)*, The Joint Task Force on Computing Curricula: IEEE Computer Society, Association for Computing Machinery, 2004. [E-book] Available: <http://sites.computer.org/ccse/SE2004Volume.pdf> [Accessed Nov. 28, 2006].
- [11] J.P. Myers Jr., "Software engineering throughout a traditional computer science curriculum", *Journal of Computing Sciences in Colleges*, vol. 16(2), pp. 31-41, 2001.
- [12] B. Bracken, "Progressing from student to professional: The importance and challenges of teaching software engineering", *Journal of Computing Sciences in Colleges*, vol. 19(2), pp. 358-368, 2003.
- [13] B.J. Corneliuss, M. Munro, and D.J. Robson, "An approach to software maintenance education", *Software Engineering Journal*, vol. 4(4), pp. 233-236, 1989.
- [14] M. Newby, "Legacy systems, software maintenance and computing curricula", in *Proceedings of the Software Education Conference*, 1994, pp. 96-102.
- [15] J. Slimick, "An undergraduate course in software maintenance and enhancement", in *CSEE&T*, pp. 61-73, 1997.
- [16] A. van Deursen, and T.C. Lethbridge, "How should software evolution and maintenance be taught?", in *ICSM 2002*, pp. 248-250.
- [17] *Ohjelmistotekniikan Korkeakoulutuspäivä* (in Finnish). Tampere, Finland: University of Tampere, 2002 (National seminar on university-level software engineering education).
- [18] M. Kalliovaara, *Ohjelmistojen Ylläpidon ja Evoluution Kurssitarjonta – Vertailu Suomen ja Yhdysvaltojen välillä* (in Finnish). Jyväskylä, Finland, University of Jyväskylä, 2006 (Comparison of the SME-courses in Finland and in the USA).
- [19] M. El-Ramly, "Experience in teaching a software reengineering course", in *ICSE 2006*, pp. 699-702.
- [20] E.B. Swanson, and C. Beath, Eds., *Maintaining Information Systems in Organizations*, John-Wiley & Sons, 1989.
- [21] T.M. Pigoski, *Practical Software Maintenance: Best Practices for Managing Your Software Investment*, Wiley, 1996.
- [22] C.L. McClure, *Managing Software Development and Maintenance*, Van Nostrand Reinhold, 1997.
- [23] M. Polo, M. Piattini, and F. Ruiz, *Advances in Software Maintenance Management: Technologies and Solutions*, IGI Global, 2002.
- [24] P. Grubb P, and A. Takang, *Software Maintenance: Concepts and Practice* (2nd ed.), World Scientific, 2003.
- [25] R. Arnold, *Software Reengineering*, IEEE Computer Society, 1993.
- [26] M. Fowler, K. Beck, J. Brant, W. Opdyke, and D. Roberts, *Refactoring: Improving the Design of Existing Code*, Addison-Wesley, 1999.
- [27] S. Jarzabek, *Effective Maintenance and Evolution: A Reuse-Based Approach*, Auerbach, 2007.
- [28] C. Pappas, and W. Murray, *Debugging C++: Troubleshooting for Programmers*, McGraw-Hill Osborne Media, 2000.
- [29] M. Haug, E. Olsen, and G. Cuevas, Eds. *Managing the Change: Software Configuration & Change Management*, Springer Verlag, 2001.
- [30] J.H. Andrews, and H.L. Lutfiyya, "Experiences with a software maintenance project course", *IEEE Transactions on Education*, vol. 43(4), pp. 383-388, 2000.
- [31] K.R. Pierce, "Teaching software engineering principles using maintenance-based projects", in *CSEE&T 1997*, pp. 53-60.
- [32] M. Postema, J. Miller, and M. Dick, "Including practical software evolution in software engineering education", in *CSEE&T 2001*, pp. 127-135.
- [33] M. Kajko-Mattsson, S. Forssander, and U. Olsson, "Corrective maintenance maturity model (CM³): Maintainer's education and training", in *ICSE 2001*, pp. 610-619.
- [34] M.A. Austin III, and M.H. Samadzadeh, "Software comprehension/maintenance: An introductory course", in *Proceedings of the 18th International Conference on Systems Engineering (ICSEng 2005)*, pp. 414-419.
- [35] J. Buchta, M. Petrenko, D. Poshyvanik, and V. Rajlich, "Teaching evolution of open-source projects in software engineering courses", in *ICSM 2006*, pp. 136-144.
- [36] P. Makkonen, "Is web-based seminar an effective way of learning in adult education?", in *HICSS 2003*, pp. 155-163.
- [37] K. Axelsson, U. Melin, and T. Wedlund, "Student activity in seminars - designing multi-functional assessment events", in: *ITiCSE 2005*, pp. 93-97.
- [38] B. Kitchenham, S.L. Pfleeger, L.M. Pickard, P.W. Jones, D.C. Hoaglin, K. El Emam, and J. Rosenberg, "Preliminary guidelines for empirical research in software engineering", *IEEE Transactions on Software Engineering*, vol. 28(4), pp. 721-734, 2002.
- [39] J. Koskinen, "Automated Transient Hypertext Support for Software Maintenance", *Jyväskylä Studies in Computing*, vol. 4, Univ. of Jyväskylä, Jyväskylä, Finland, 2000.
- [40] A. Abran, I. Silva, and L. Primera, "Field studies using functional size measurement in building estimation models for software maintenance", *Journal of Software Maintenance and Evolution: Research and Practice*, vol. 14(1), pp. 31-64, 2002. {III,EE:EFFE,E,34,4}.
- [41] H. Agrawal, R. Demillo, and E. Spafford, "Debugging with dynamic slicing and backtracking", *Software - Practice and Experience*, vol. 23(6), pp. 589-616, 1993. {II,TE:SLIC,C,28,3}.
- [42] A. Andrews, S. Ghosh, and E. Choi, "A model for understanding software components", in *ICSM 2002*, pp. 359-368. {III,TA:PCMP,T,10,2}.
- [43] G. Arango, I. Baxter, P. Freeman, and G. Pidgeon, "TMM: Software maintenance by transformation", *IEEE Software*, vol. 3(3), pp. 27-39, 1986. {II,TE:RVEN,C,13,3}.
- [44] V. Arunachalam, and W. Sasso, "Cognitive processes in program comprehension: An empirical analysis in the context of software reengineering", *Journal of Systems and Software*, vol. 34(3), pp. 177-189, 1996. {I,TA:PCMP,E,13,1}.
- [45] E. Baniassad, and G. Murphy, "Conceptual module querying for software reengineering", in *ICSE 1998*, pp. 64-73. {III,TE:RVEN,C,10,3}.
- [46] R. Banker, S. Datar, and C.F. Kemerer, "A model to evaluate variables impacting the productivity of software maintenance projects", *Management Science*, vol. 37(1), pp. 1-18, 1991. {I,EE:EFFE,T,18,4}.
- [47] R. Banker, S. Datar, C.F. Kemerer, and D. Zweig, "Software complexity and maintenance costs", *Communications of the ACM*, vol. 36(11), pp. 81-94, 1993. II,EE:METR,T,14,4}.
- [48] R. Banker, and S. Slaughter, "Project size and software maintenance productivity: Empirical evidence on economics of scale in software maintenance", in *International Conference on Information Systems (ICIS 1994)*, pp. 279-289. {III,EE:EFFE,E,11,4}.
- [49] B. Bellay, and H. Gall, "A comparison of four reverse engineering tools", in *WCRE 1997*, pp. 2-11. {II,TE:RVEN,E,10,2}.
- [50] T. Biggerstaff, B. Mitbender, and D. Webster, "The concept assignment problem in program understanding", in *ICSE 1993*, pp. 482-498. {I,TA:PCMP,T,17,2}.
- [51] D. Binkley, and K.B. Gallagher, "Program slicing", *Advances in Computers*, vol. 43, pp. 1-50, 1996. {II,TE:SLIC,S,50,3}.
- [52] D. Boehm-Davis, R. Holt, and A. Schultz, "The role of program structure in software maintenance", *International Journal of Man-Machine Studies*, vol. 36(1), pp. 21-63, 1992. {II,TA:MNTB,T,43,2}.
- [53] L. Briand, T. Langley, and I. Wiczorek, "A replicated assessment and comparison of common software cost modeling techniques", in *ICSE 2000*, pp. 377-386. {III,EE:COST,E,10,3}.
- [54] M. Brown, and J. Hershberger, "Color and sound in algorithm animation", *Computer*, vol. 25(12), pp. 52-63, 1992. {I,TE:PVIS,T,10,2}.

- [55] I. Burnstein, and K. Roberson, "Automated chunking to support program comprehension", in *IWPC 1997*, pp. 40-49. {II,TE:RVEN,C,10,2}.
- [56] D. Caivano, F. Lanubile, and G. Visaggio, "Software renewal process comprehension using dynamic effort estimation", in *ICSM 2001*, pp. 209-218. {III,EE:EFFE,T,10,4}.
- [57] F. Calzolari, P. Tonella, and G. Antoniol, "Dynamic model for maintenance and testing effort", in *ICSM 1998*, pp. 104-112. {III,EE:EFFE,T,9,4}.
- [58] M. Capretz, and M. Munro, "Software configuration management issues in the maintenance of existing systems", *Software Maintenance*, vol. 6(1), pp. 1-14, 1994. {II,TE:RVEN,T,15,2}.
- [59] G. Catalin-Roman, and K. Fox, "A taxonomy of program visualization systems", *Computer*, vol. 26(12), pp. 11-24, 1993. {I,TE:PVIS,S,14,2}.
- [60] T. Chan, S. Chung, and T. Ho, "An economic model to estimate software rewriting and replacement times", *IEEE Transactions on Software Engineering*, vol. 22(8), pp. 580-598, 1996. {III,EE:COST,T,19,5}.
- [61] N. Chapin, J. Hale, K. Khan, J. Ramil, and W.-G. Tan, "Types of software evolution and software maintenance", *Journal of Software Maintenance and Evolution: Research and Practice*, vol. 13(1), pp. 3-30, 2001. {II,TA:MTPR,T,28,1}.
- [62] T.Y. Chen, and C.K. Low, "Error detection in C++ through dynamic data flow analysis", *Software - Concepts and Tools*, vol. 18(1), pp. 1-13, 1997. {II,TA:DEBG,T,13,3}.
- [63] Y.-F. Chen, G. Fowler, E. Koutsofios, and R. Wallach, "Ciao: A graphical navigator for software and document repositories", in *ICSM 1995*, pp. 66-75. {II,TE:RVEN,C,10,3}.
- [64] Y.-F. Chen, E. Gansner, and E. Koutsofios, "A C++ data model supporting reachability analysis and dead code detection", *IEEE Transactions on Software Engineering*, vol. 24(9), pp. 682-694, 1998. {III,TE:RVEN,C,13,4}.
- [65] D. Coleman, B. Lowther, and P. Oman, "The application of software maintainability models in industrial software systems", *Journal of Systems and Software*, vol. 29(1), pp. 3-16, 1995. {III,EE:EFFE,T,14,3}.
- [66] S. Comella-Dorda, K. Wallnau, R. Seacord, and J. Robert, "A survey of black-box modernization approaches for information systems", in *ICSM 2000*, pp. 173-183. {II,TE:REEN,S,11,2}.
- [67] M. Consens, A. Mendelzon, and A. Ryman, "Visualizing and querying software structures", in *ICSE 1992*, pp. 138-156. {III,TE:PVIS,C,19,3}.
- [68] J. Cook, and J. Dage, "Highly reliable upgrading of components", in *ICSE 1999*, pp. 203-212. {I,TE:RVEN,T,10,3}.
- [69] C. Corritore, and S. Wiedenbeck, "Mental representations of expert procedural and object-oriented programmers in a software maintenance task", *International Journal of Human-Computing Studies*, vol. 50(1), pp. 61-84, 1999. {I,TA:PCMP,E,24,2}.
- [70] J.H. Cross II, E. Chikofsky, and C.H. May Jr, "Reverse engineering", *Advances in Computers*, vol. 35, pp. 199-254, 1992. {I,TE:RVEN,S,56,2}.
- [71] S. Davies, "The nature and development of programming plans", *International Journal of Man-Machine Studies*, vol. 32(4), pp. 461-481, 1990. {III,TA:PCMP,T,21,1}.
- [72] A. De Lucia, M. Di Penta, S. Stefanucci, and G. Venturi, "Early effort estimation of massive maintenance processes", in *ICSM 2002*, pp. 234-237. {II,EE:EFFE,E,4,4}.
- [73] S. Dietrich, and F. Calliss, "A conceptual design for a code analysis knowledge base", *Software Maintenance*, vol. 4(1), pp. 19-36, 1992. {II,TE:RVEN,T,17,2}.
- [74] M. Ernst, J. Cockrell, W. Griswold, and D. Notkin, "Dynamically discovering likely program invariants to support program evolution", *IEEE Transactions on Software Engineering*, vol. 27(2), pp. 99-123, 2001. {III,TE:RVEN,T,25,3}.
- [75] R. Fanta, and V. Rajlich, "Reengineering object-oriented code", in *ICSM 1998*, pp. 238-246. {II,TE:RVEN,T,9,1}.
- [76] A. Fasolino, D. Natale, A. Poli, and A. Quaranta, "Metrics in the development and maintenance of software: An application in a large scale environment", *Journal of Software Maintenance: Research and Practice*, vol. 12(6), pp. 343-355, 2000. {III,EE:METR,E,13,3}.
- [77] R. Ferenc, A. Beszedes, M. Tarkainen, and T. Gyimothy, "Columbus - reverse engineering tool and schema for C++", in *ICSM 2002*, pp. 172-181. {II,TE:RVEN,C,10,3}.
- [78] L. Foreman, and S. Zweben, "A study of the effects of control and data flow testing strategies", *Journal of Systems and Software*, vol. 21(3), pp. 215-228, 1993. {III,TA:MTPR,E,14,2}.
- [79] J. French, J. Knight, and A. Powell, "Applying hypertext structures to software documentation", *Information Processing & Management*, vol. 33(2), pp. 219-231, 1997. {II,TE:RVEN,C,13,1}.
- [80] S. Fäustle, M. Fugini, and E. Damiani, "Retrieval of reusable components using functional similarity", *Software - Practice and Experience*, vol. 26(5), pp. 491-530, 1996. {III,TE:RVEN,T,40,3}.
- [81] G. Gannod, and B. Cheng, "A framework for classifying and comparing software reverse engineering and design recovery techniques", in *WCRE 1999*, pp. 77-88. {I,TE:RVEN,S,12,2}.
- [82] E. Gellenbeck, and C. Cook, "An investigation of procedure and variable names as beacons during program comprehension", in *ESP 1991*, pp. 65-81. {III,TA:PCMP,E,17,1}.
- [83] V. Gibson, and J. Senn, "System structure and software maintenance performance", *Communications of the ACM*, vol. 32(3), pp. 347-358, 1989. {III,TA:MNTB,T,12,2}.
- [84] M. Goedicke, and U. Zdun, "Piecemeal legacy migrating with an architectural pattern language: A case study", *Journal of Software Maintenance and Evolution: Research and Practice*, vol. 14(1), pp. 1-30, 2002. {III,TE:RVEN,E,31,3}.
- [85] N. Gorla, A. Benander, and B. Benander, "Debugging effort estimation using software metrics", *IEEE Transactions on Software Engineering*, vol. 16(2), pp. 223-231, 1990. {II,EE:EFFE,T,9,4}.
- [86] J. Grass, "Object-oriented design archaeology with CIA++", *Computing Systems*, vol. 5(1), pp. 5-66, 1992. {II,TE:RVEN,C,62,2}.
- [87] W. Griswold, and D. Notkin, "Automated assistance for program restructuring", *ACM Transactions on Software Engineering and Methodology*, vol. 2(3), pp. 228-269, 1993. {I,TE:REEN,C,42,4}.
- [88] W. Griswold, and D. Notkin, "Architectural tradeoffs for a meaning-preserving program restructuring tool", *IEEE Transactions on Software Engineering*, vol. 21(4), pp. 275-287, 1995. {III,TE:REEN,C,12,5}.
- [89] A. Hassan, and R. Holt, "Architecture recovery of web applications", in *ICSE 2002*, pp. 349-359. {II,TE:RVEN,T,11,3}.
- [90] K. Heisler, Y. Kasho, and W. Tsai, "A reverse engineering model for C programs", *Information Sciences*, vol. 68, pp. 155-189, 1993. {II,TE:RVEN,T,35,3}.
- [91] S. Henry, and M. Humphrey, "Object-oriented vs. procedural programming languages: Effectiveness in program maintenance", *Journal of Object-Oriented Programming*, vol. 6(3), pp. 41-49, 1993. {I,TA:MNTB,E,9,2}.
- [92] S. Horwitz, and T. Reps, "The use of program dependence graphs in software engineering", in *ICSE 1992*, pp. 392-411. {II,TE:SLIC,T,20,5}.
- [93] M. Interesse, "Preventive actions for residual Y2K defects: A case study", *Journal of Software Maintenance: Research and Practice*, vol. 12(1), pp. 1-23, 2000. {I,TA:PCMP,E,24,2}.
- [94] K. Jambor-Sadeghi, M. Ketabchi, J. Chue, and M. Ghiassi, "A systematic approach to corrective maintenance", *The Computer Journal*, vol. 37(9), pp. 764-778, 1994. {III,TA:DEBG,C,15,3}.
- [95] M. Jørgensen, "An empirical study of software maintenance tasks", *Software Maintenance*, vol. 7(1), pp. 27-48, 1995. {III,TA:MTPR,E,22,2}.
- [96] M. Jørgensen, "Experience with the accuracy of software maintenance task effort prediction models", *IEEE Transactions on Software Engineering*, vol. 21(8), pp. 674-681, 1995. {III,EE:EFFE,E,8,4}.
- [97] M. Jørgensen, and D. Sjøberg, "Impact of experience on maintenance skills", *Journal of Software Maintenance: Research and Practice*, vol. 14(2), pp. 123-146, 2002. {I,TA:PCMP,E,24,2}.
- [98] D. Kafura, and G. Reddy, "The use of software complexity metrics in software maintenance", *IEEE Transactions on Software Engineering*, vol. 13(3), pp. 335-343, 1987. {III,EE:METR,S,9,2}.
- [99] M. Kajko-Mattsson, "Problem management maturity within corrective maintenance", *Journal of Software Maintenance and Evolution: Research and Practice*, vol. 14(3), pp. 197-227, 2002. {I,TA:DEBG,T,31,2}.
- [100] M. Kamkar, "An overview and comparative classification of program slicing techniques", *Journal of Systems and Software*, vol. 31(3), pp. 197-214, 1995. {III,TE:SLIC,S,18,3}.
- [101] Y. Kataoka, T. Imai, H. Andou, and T. Fukaya, "A quantitative evaluation of maintainability enhancement by refactoring", in *ICSM 2002*, pp. 576-585. {III,EE:METR,E,10,5}.

- [102] C.F. Kemerer, "An empirical validation of software cost estimation models", *Communications of the ACM*, vol. 30(5), pp. 416-429, 1987. {III,EE:COST,E,14,3}.
- [103] C.F. Kemerer, "Software complexity and software maintenance: A survey of empirical research". *Annals of Software Engineering*, vol. 1, pp. 1-22, 1995. {III,EE:METR,S,23,2}.
- [104] B. Kitchenham, G. Travassos, A. Mayrhauser, F. Niessink, N. Schneidewind, J. Singer, S. Takada, R. Vehviläinen, and H. Yang, "Towards an ontology of software maintenance", *Journal of Software Maintenance: Research and Practice*, vol. 11(6), pp. 365-389, 1999. {II,TA:MTPR,T,25,2}.
- [105] J. Koskinen, "Experimental evaluation of hypertext access structures", *Journal of Software Maintenance and Evolution: Research and Practice*, vol. 14(2), pp. 83-108, 2002. {II,TE:RVEN,E,26,3}.
- [106] U. Kölsch, and J. Laschewski, "A framework for object-oriented reverse engineering of legacy information systems", *International Journal of Software Engineering and Knowledge Engineering*, vol. 9(1), pp. 27-54, 1999. {III,TE:RVEN,T,28,2}.
- [107] D. Lanning, and T. Khoshgoftaar, "Modeling the relationship between source code complexity and maintenance difficulty", *Computer*, vol. 27(9), pp. 35-40, 1994. {III,TA:MNTB,E,6,3}.
- [108] M. Lehman, D. Perry, and J. Ramil, "Implications of evolution metrics on software maintenance", in *ICSM 1998*, pp. 208-217. {II,EE:METR,E,10,3}.
- [109] S. Letovsky, "Cognitive processes in program comprehension", in *ESP 1986*, pp. 58-79. {I,TA:PCMP,T,22,1}.
- [110] P. Linos, and V. Courtois, "A toolset for maintaining hybrid C++ programs", *Journal of Software Maintenance: Research and Practice*, vol. 8(6), pp. 389-420, 1996. {I,TE:RVEN,C,32,1}.
- [111] L. Marcoccia, "Building infrastructure for fixing the year 2000 bug: A case study", *Journal of Software Maintenance: Research and Practice*, vol. 10(5), pp. 333-352, 1998. {I,TA:PCMP,T,20,1}.
- [112] A. von Mayrhauser, "Maintenance and evolution of software products", *Advances in Computers*, vol. 39, pp. 1-49, 1994. {II,TA:MTPR,T,50,1}.
- [113] A. von Mayrhauser, and S. Lang, "A coding scheme to support systematic analysis of software comprehension", *IEEE Transactions on Software Engineering*, vol. 25(4), pp. 526-540, 1999. {I,TA:PCMP,T,15,2}.
- [114] A. von Mayrhauser, and A.M. Vans, "Program understanding models and experiments", *Advances in Computers*, vol. 40, pp. 1-38, 1995. {II,TA:PCMP,T,39,1}.
- [115] A. von Mayrhauser, and A.M. Vans, "Identification of dynamic comprehension processes during large scale maintenance", *IEEE Transactions on Software Engineering*, vol. 22(6), pp. 424-437, 1996. {I,TA:PCMP,E,14,2}.
- [116] A. von Mayrhauser, and A.M. Vans, "Program understanding during software adaptation tasks", in *ICSM 1998*, pp. 316-325. {I,TA:PCMP,E,10,2}.
- [117] A. von Mayrhauser, A.M. Vans, and A. Howe, "Program understanding behaviours during enhancement of large-scale software", *Journal of Software Maintenance: Research and Practice*, vol. 9(5), pp. 299-327, 1997. {III,TA:PCMP,E,29,2}.
- [118] A. Mehta, and G. Heineman, "Evolving legacy system features into fine-grained components", in *ICSE 2002*, pp. 417-427. {III,TA:MTPR,T,11,3}.
- [119] E. Merlo, P.Y. Gagne, J.F. Girard, K. Kontogiannis, P. Panagaden, and R. De Mori, "Reengineering user interfaces", *IEEE Software*, vol. 12(1), pp. 64-73, 1995. {II,TE:REEN,C,10,3}.
- [120] T. Moher, D. Mak, B. Blumenthal, and L. Leventhal, "Comparing the comprehensibility of textual and graphical programs: The case of Petri nets", in *ESP 1993*, pp. 137-161. {III,TA:PCMP,E,25,2}.
- [121] J. Munson, and S. Elbaum, "Code Churn: A measure for estimating the impact of code change", in *ICSM 1998*, pp. 24-31. {I,EE:EFFE,C,8,5}.
- [122] G. Murphy, D. Notkin, W. Griswold, and E. Lan, "An empirical study of static call graph extractors", *ACM Transactions on Software Engineering and Methodology*, vol. 7(2), pp. 158-191, 1998. {III,TE:RVEN,E,34,3}.
- [123] J. Niere, W. Schäfer, J. Wadsack, L. Wendehals, and J. Welsh, "Towards pattern-based design recovery", in *ICSE 2002*, pp. 338-348. {III,TE:RVEN,T,11,3}.
- [124] F. Niessink, and H. van Vliet, "Two case studies in measuring software maintenance effort", in *ICSM 1998*, pp. 76-85. {III,EE:EFFE,E,10,3}.
- [125] F. Niessink, and H. van Vliet, "Software maintenance from a service perspective", *Journal of Software Maintenance: Research and Practice*, vol. 12(2), pp. 103-120, 2000. {II,TA:MTPR,T,18,1}.
- [126] P. Oman, and C. Cook, "Typographic style is more than cosmetic", *Communications of the ACM*, vol. 33(5), pp. 506-520, 1990. {I,TA:MNTB,T,15,2}.
- [127] P. Oman, and C. Cook, "A programming style taxonomy", *Journal of Systems and Software*, vol. 15(3), pp. 287-301, 1991. {I,TA:MNTB,T,15,2}.
- [128] P. Paul, and A. Prakash, "A query algebra for program databases", *IEEE Transactions on Software Engineering*, vol. 22(3), pp. 202-217, 1996. {III,TE:RVEN,C,16,5}.
- [129] T. Pearce, and P. Oman, "Maintainability measurements on industrial source code maintenance activities", in *ICSM 1995*, pp. 295-303. {II,EE:METR,E,9,3}.
- [130] N. Pennington, "Stimulus structures and mental representations in expert comprehension of computer programs", *Cognitive Psychology*, vol. 19(3), pp. 295-341, 1987. {III,TA:PCMP,E,47,1}.
- [131] M. Polo, M. Piattini, and F. Ruiz, "Using code metrics to predict maintenance of legacy programs: A case study", in *ICSM 2001*, pp. 202-208. {III,EE:METR,E,7,4}.
- [132] R.S. Pressman, "Software process and project metrics", in *Software Engineering - A Practitioner's Approach (5th ed.)*, pp. 79-111, 2001, McGraw-Hill. {III,EE:METR,S,23,3}.
- [133] R.S. Pressman, "Technical metrics for software", in *Software Engineering - A Practitioner's Approach (5th ed.)*, pp. 507-538, 2001, McGraw-Hill. {III,EE:METR,S,30,3}.
- [134] R.S. Pressman, "Technical metrics for object-oriented systems", in *Software Engineering - A Practitioner's Approach (5th ed.)*, pp. 653-669, McGraw-Hill. {II,EE:METR,S,17,3}.
- [135] B. Price, R. Baecker, and I. Small, "A principled taxonomy of software visualization", *Journal of Visual Languages and Computing*, vol. 4(3), pp. 211-266, 1993. {I,TE:PVIS,S,56,2}.
- [136] G. Ramalingam, and T. Reps, "A theory of program modifications", in *Proceedings of the Colloquium on Combining Paradigms for Software Development*, Springer-Verlag, 1992, pp. 137-152. {I,TA:MTPR,T,16,5}.
- [137] R. Riecken, J. Koenemann-Belliveau, and S. Robertson, "What do expert programmers communicate by means of descriptive commenting", in *ESP 1991*, 177-195. {II,TA:PCMP,T,19,1}.
- [138] R. Rist, "Plans in programming: Definition, demonstration, and development", in *ESP 1986*, pp. 28-45. {III,TA:PCMP,T,18,1}.
- [139] E. Rose, and I. Eriksson, "Development and maintenance costs: Measures of software maintainability", in *Global & Multiple Criteria Optimization and Information Systems Quality*, Turku, Finland: Åbo Akademi, 1998, pp. 21-35. {III,EE:COST,T,15,3}.
- [140] I. Sahin, and M. Zahedi, "Policy analysis for warranty, maintenance, and upgrade of software systems", *Journal of Software Maintenance and Evolution: Research and Practice*, vol. 13(6), pp. 469-493, 2001. {III,EE:EFFE,T,25,4}.
- [141] S. Schach, and A. Tomer, "A maintenance-oriented approach to software construction", *Journal of Software Maintenance and Evolution: Research and Practice*, vol. 12(1), pp. 25-45, 2000. {III,TA:MNTB,T,21,3}.
- [142] C. Seaman, "The information gathering strategies of software maintainers", in *ICSM 2002*, pp.141-149. {III,TA:MTPR,E,9,1}.
- [143] F. Sheldon, K. Jerath, and H. Chung, "Metrics for maintainability of class inheritance hierarchies", *Journal of Software Maintenance and Evolution: Research and Practice*, vol. 14(3), pp. 147-160, 2002. {II,EE:METR,T,14,3}.
- [144] J. Singer, T. Lethbridge, N. Vinson, and N. Anquetil, "An examination of software engineering work practices", in *CASCON 1997*, pp. 209-223. {III,TA:MTPR,E,15,2}.
- [145] H.M. Sneed, "Estimating the costs of software maintenance tasks", in *ICSM 1995*, pp. 168-181. {III,EE:COST,E,19,3}.
- [146] H.M. Sneed, "Planning the reengineering of legacy systems", *IEEE Software*, vol. 12(1), pp. 24-34, 1995. {III,TE:REEN,E,11,3}.
- [147] M.-A. Storey, K. Wong, P. Fong, D. Hooper, K. Hopkins, and H. Müller, "On designing an experiment to evaluate a reverse engineering tool", in *WCSE 1996*, pp. 31-40. {III,TE:RVEN,C,10,2}.
- [148] M.A. Storey, K. Wong, and H. Müller, "How do program understanding tools affect how programmers understand programs?", in *WCSE 1997*, pp. 12-21. {II,TA:PCMP,E,10,2}.
- [149] E. Swanson, and C. Beath, "Departmentalization in software development and maintenance", *Communications of the ACM*, vol. 33(6), pp. 658-667, 1990. {II,TA:MTPR,T,10,2}.

- [150] T. Systä, K. Koskimies, and H. Müller, "Shimba - an environment for reverse engineering Java software systems", *Software - Practice and Experience*, vol. 31(4), pp. 371-394, 2001. {III,TE:RVEN,C,24,3}.
- [151] L. Tahvildari, and K. Kontogiannis, "A software transformation framework for quality-driven object-oriented re-engineering", in *ICSM 2002*, pp. 596-605. {II,TE:REEN,T,10,4}.
- [152] B. Teasley, "The effects of naming style and expertise on program comprehension", *International Journal of Human-Computing Studies*, vol. 40(5), pp. 757-770, 1994. {III,TA:PCMP,E,14,2}.
- [153] J. Teng, S. Jeong, and V. Grover, "Profiling successful reengineering projects", *Communications of the ACM*, vol. 41(6), pp. 96-102, 1998. {II,TE:REEN,E,7,2}.
- [154] P. Tonella, G. Antoniol, R. Fiutem, and F. Calzolari, "Reverse engineering 4.7 million lines of code", *Software - Practice and Experience*, vol. 30(2), pp. 129-150, 2000. {I,TE:RVEN,E,22,3}.
- [155] R. Turver, and M. Munro, "An early impact analysis technique for software maintenance", *Journal of Software Maintenance: Research and Practice*, vol. 6(1), pp. 35-52, 1994. {II,TE:RVEN,E,18,3}.
- [156] G. Venkatesh, "Experimental results from dynamic slicing of C programs", *ACM Transactions on Programming Languages and Systems*, vol. 17(2), pp. 197-216, 1995. {II,TE:SLIC,E,20,3}.
- [157] I. Vessey, "Expertise in debugging computer programs: A process analysis", *International Journal of Man-Machine Studies*, vol. 23(5), pp. 459-494, 1985. {II,TA:DEBG,E,36,2}.
- [158] I. Vessey, "Toward a theory of computer program bugs: An empirical test", *International Journal of Man-Machine Studies*, vol. 30(1), pp. 23-46, 1989. {I,TA:DEBG,T,24,2}.
- [159] B. Weide, W. Ogden, and S. Zweben, "Reusable software components", *Advances in Computers*, vol. 33, pp. 1-65, 1991. {I,TA:MNTB,S,66,2}.
- [160] S. Wiedenbeck, "Beacons in computer program comprehension", *International Journal of Man-Machine Studies*, vol. 25(6), pp. 697-709, 1986. {I,TA:PCMP,T,13,2}.
- [161] S. Wiedenbeck, V. Ramalingam, S. Sarasammo, and C. Corritore, "A comparison of the comprehension of object-oriented and procedural programs by novice programmers", *Interacting with Computers*, vol. 11(3), pp. 255-282, 1999. {II,TA:PCMP,E,28,2}.
- [162] N. Wilde, A. Chapman, and R. Richardson, "The extensible dependency analysis tool set: A knowledge base for understanding industrial software", *International Journal of Software Engineering and Knowledge Engineering*, vol. 4(4), pp. 521-534, 1994. {I,TE:RVEN,C,14,3}.
- [163] W. Yang, "Identifying syntactic differences between two programs", *Software - Practice and Experience*, vol. 21(7), pp. 739-755, 1991. {II,TE:RVEN,C,17,3}.
- [164] D. Yeh, and J.-W. Jeng, "An empirical study of the influence of departmentalisation and organizational position on software maintenance", *Journal of Software Maintenance and Evolution: Research and Practice*, vol. 14(1), pp. 65-82, 2002. {III,TA:MTPR,E,18,2}.
- [165] Rui Yin, and R. Keller, "Program comprehension by visualization in contexts", in *ICSM 2002*, pp. 332-341. {II,TE:PVIS,C,10,2}.
- [166] Ying Zou, and K. Kontogiannis, "Migration to object oriented platforms: A state transformation approach", in *ICSM 2002*, pp. 530-539. {III,TE:RVEN,T,10,3}.
- [167] *The ECTS Grading Scale*, Academic Office, Univ. of Warwick, UK, 2002. Available: <http://www2.warwick.ac.uk/academic-office/undergrad/ects/grading/> [Accessed Nov. 28, 2006].
- [168] J. Paakki, A. Salminen, and J. Koskinen, "Automated hypertext support for software maintenance", *The Computer Journal*, vol. 39(7), pp. 577-597, July 1996.
- [169] J. Koskinen, A. Salminen, and J. Paakki, "Hypertext support for the information needs of software maintainers", *Journal of Software Maintenance and Evolution: Research and Practice*, vol. 16(3), pp. 187-215, May 2004.
- [170] J. Koskinen, J.J. Ahonen, T. Tilus, H. Sivula, I. Kankaanpää, H. Lintinen, and P. Juutilainen, "Developing software evolution estimation methods for software industry" in *Proceedings of the 3rd IASTED International Conference on Software Engineering (SE 2006)*, ACTA Press, 2006, pp. 323-328.
- [171] J. Koskinen, J.J. Ahonen, I. Kankaanpää, H. Sivula, and T. Tilus, "Developing support for industrial software evolution decision making", *Communications in Dependability and Quality Management*, vol. 11(1), pp. 98-106, March 2008.

Received: February 17, 2009

Revised: May 12, 2009

Accepted: May 13, 2009

© Jussi Koskinen; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.