Two Experiments Comparing Two Four-Step EPMcreate-Based Creativity Techniques for Requirements Elicitation[*]

Andrea Herrmann [0000-0002-4234-8422][1], Luisa Mich [0000-0002-0018-6883][2], and Daniel M. Berry [0000-0002-6817-9081][3]

[1] Herrmann & Ehrlich, Stuttgart, Germany
`herrmann@herrmann-ehrlich.de`
[2] Department of Industrial Engineering, University of Trento, Italy
`luisa.mich@unitn.it`
[3] Cheriton School of Computer Science, University of Waterloo, Canada
`dberry@uwaterloo.ca`

Abstract. The Elementary Pragmatic Model Creativity Technique, a.k.a EPM-create, is a method for creative requirements discovery. It includes 16 steps corresponding to all the combinations of two stakeholders' viewpoints. The feasibility and effectiveness of the 16-step process have been confirmed by a number of experiments. The need to reduce the number of steps was observed in the very first experiments run in 2003. To address that problem a four-step creativity technique, POEPMcreate (Power-Only EPMcreate), was defined and tested. This paper describes an experiment comparing two four-step techniques, POEPMcreate and a new technique, ROSEPMcreate (Redundant, Odd Step EPMcreate), resembling traditional requirements elicitation with brainstorming. The results, even if preliminary, seem to indicate that an analyst's work experience and native language influence the number and innovativeness of requirements he or she generates. The paper also offers Kano categories as a way to evaluate the innovativeness of generated requirements.

Keywords: Elementary pragmatic model; creativity technique; viewpoint; creativity process; requirements elicitation.

1   Introduction

Creativity techniques can be successfully applied to elicit requirements. The literature offers a variety of approaches and research about them. Among them, we can cite [1], [2], [3] and the papers of the CreaRE workshops, a REFSQ satellite (https://sites.google.com/site/creare2018, https://sites.google.com/site/creare2017).

---

[*] This paper describes an experiment repeated with different subjects in a different location in order to confirm the first experiment's conclusions. Therefore, this paper reuses background, context-setting, and experiment-description text as well as data from our previous paper [10] in order to make this paper self-contained and complete.

One of the parameters that can be used to classify these techniques is the number of phases or steps characterizing their creative elicitation processes.

This paper focuses on variations of the Elementary Pragmatic Model Creativity Technique, a.k.a. EPMcreate, which provides a 16-step process for creative requirements elicitation [4]. According to EPMcreate, requirements are ideas to fulfill needs of stakeholders having different viewpoints, and the viewpoints of a pair of different stakeholders can be combined in 16 different ways, suggesting a 16-step elicitation process. A number of experiments have confirmed the feasibility and the effectiveness of EPMcreate [4], [5]. The most recent experiments have investigated if it is possible to adopt a process based on a subset of 16 steps of EPMCreate [5], [6]. Given the high number of these subsets [6], it is necessary to choose the steps to be included in a proposed new technique, using sound selection criteria. For the first technique with a reduced number of steps, the adopted criterion was that of coverage, that is, using only the steps corresponding to the four not overlapping combinations of two stakeholders' viewpoints. Since these four steps are those named for the first four powers of two, the technique was named Power-Only EPMcreate (POEPMcreate) [5]. Please consult references [4] and [5] for descriptions of the techniques.

In this paper, POEPMcreate is compared with ROSEPMcreate (Redundant, Odd Step EPMcreate), another four-step variant of EPMcreate. ROSEPMcreate includes activities that − albeit in different ways − are usually included in the traditional requirements elicitation sessions. An experiment was designed and run twice with students working in teams of two, playing the role of analysts combining the two designated viewpoints. The results allow comparing results from two different populations and suggest new issues to be addressed in future experiments.

In the rest of the paper, Section 2 illustrates the steps of EPMcreate, the bases for the four-step techniques compared in the experiment. Section 3 describes the experiment. The main results and their analysis are given in Section 4. Section 5 describes the threats to the validity of the results, and Section 6 concludes the paper.

2   The 16-Step Creativity Technique

Each of the 16 steps of EPMcreate corresponds to a different combination of the viewpoints of two stakeholders. These combinations can then be denoted using a Boolean function of two binary variables [7]. The 16 Boolean functions are binary connectives, and the most known of them are the logical AND, OR, and XOR, which in terms of requirements, correspond to the situations in which the analyst is trying to generate requirements that are needed by both the identified stakeholders (AND); by any of them (OR); and by one or the other but not by both of them, (XOR), i.e., by either of them. Also, the always-false and always-true functions are included in the EPMcreate process as the first and the last steps, respectively. See Table 1 in our previous paper [10]. In general, EPMcreate can be considered a conceptual framework that supports the generation of requirements elicitation techniques with any number, from 1 through 16, of EPMcreate's steps. Given that testing all possible combinations

of steps is impossible, the problem is to find sound principles with which to select for a technique steps that are more adequate, effective, or feasible.

3 The Experiment: Two 4-Step Techniques

We wanted to empirically validate that ROSEPMcreate, a new four-step variant of EPMcreate, could be used to support requirements elicitation. To this end, we run an explorative experiment to compare ROSEPMcreate to POEPMcreate, whose effectiveness has already been empirically validated [5]. POEPMcreate was defined to not include any redundancies, i.e., each of the four regions of the Venn diagram, the atoms of the Boolean algebra for two variables [7], is explored only once. It thus consists of Steps 1, 2, 4, and 8 of EPMCreate. ROSEPMcreate consists of Steps 3, 5, 7, and 15. Each of Steps 3 and 5 focuses on the needs of one stakeholder, even as these needs might coincide with the needs of the other stakeholder. Next, Step 7 focuses on the needs of either or both stakeholders. Finally, Step 15 is a catch-all step in which the analyst is invited to think of any possible requirement idea. The overlap of the foci of Steps 3 and 5 is in distinction to that of each of Steps 2 and 4, in POEPMcreate, that focuses on the needs of one stakeholder to the exclusion of the needs of the other. Observe that the focus of Step 7 overlaps the focus of each of Steps 3 and 5, and that the focus of Step 15 overlaps the focus of each of Steps 3, 5, and 7. In addition, Step 15 corresponds to the question, "Anything else?" and to one of the principles of brainstorming [8], "Do not worry about feasibility or any other issue." The redundancy of the steps and that the name of each step is an odd number suggests the name of Redundant, Odd Steps EPMcreate (ROS-EPMcreate) for the new technique.

3.1 Experiment Design and Execution

The experiment participants were asked to generate requirements to design a meeting planner system. The chosen system fulfills two main needs: being simple enough (1) to not require deep domain knowledge and (2) to allow running the entire experiment during a single lecture slot. We avoided naming a specific product to avoid influencing the results of the experiment. The two stakeholder (SH) groups were designated to be:

- SH1 = the organizer of a meeting
- SH2 = a participant of the meeting being organized

The four steps for POEPMcreate were described to the participants as:

1. Please find requirements that are needed by SH1 as well as by SH2.
2. Please find requirements that are needed by SH1, but not by SH2.
3. Please find requirements that are needed by SH2, but not by SH1.
4. Please find requirements that are needed neither by SH1 nor by SH2.

The four steps for ROSEPMcreate were described to the participants as:

1. Please find requirements that are needed by SH1.
2. Please find requirements that are needed by SH2.
3. Please find requirements that are needed by either SH2, SH1, or both.
4. Please find requirements, independent of whether SH1 or SH2 need them.

Experiment 1 (Exp1) was conducted at the Hochschule für Technik in Stuttgart (HFT, <https://www.hft-stuttgart.de>) during the course "Software Project Management". The voluntary participants were students studying business informatics in the second year and can be considered as having the experience comparable to that of junior analysts. They had no specific requirements engineering training before, but they were familiar with the meeting planning problem, because they had to arrange group meetings, in order to carry out their group course assignments. Experiment 2 (Exp2) took place at the University of Heidelberg during the course "Requirements Engineering". The voluntary participants were students from different semesters and fields of study, but most of them were Master's degree students from Informatics. The prerequisite for this course was a previous course, "Introduction to Software Engineering". The experiment took place at the end of the course; the participants can be considered to be on the knowledge level of the IREB CPRE Foundation Level.

Exp1 had 22 participants, and Exp2 had 16. The 22 and 16 participants were assigned to 11 and 8 two-person teams by distributing cards with team names "Pa", "Pb", "Ra", "Rb", etc. (with "P" or "R" identifying the team's technique, POEPMcreate or ROS-EPMcreate, respectively, and the small letter making the name unique) randomly to the participants. Each team applied its technique's four steps in the specified order and did each step for 10 minutes, for a total of 40 minutes. Each team was given two short questionnaires after finishing the last step. The first questionnaire asked the team to self-evaluate the requirements it had generated and the learnability of its technique; the second gathered data about ages, the years of experience in software development, and native languages of the team's members, to allow us to check if these characteristics affected the results of the experiment. Altogether, each experiment lasted 60 minutes.

3.2   Instructions to the Team and the Questionnaires

The instructions given to the participants were the following (translated from the original instructions in German, the language of the course):

Your task consists in identifying requirements for software which supports organizing meetings. These requirements can describe functionalities, but also non-functional requirements, or concern the graphical design or technical properties. Please, try to be as creative as possible, i.e., to develop new ideas.

In addition, the experimenter said in German to all participants, "Find as many requirements as possible." Each team was then asked to apply the steps of its assigned technique, giving its ideas in free-form text, using the description of its four steps given in Section 3.1 and Venn diagrams like the one in Fig. 1. The last page of experiment sheet gave the questionnaires shown in Table 1 and Table 2. In addition to the

tables, there was a free-text field, in which the team could add any further comment it had.
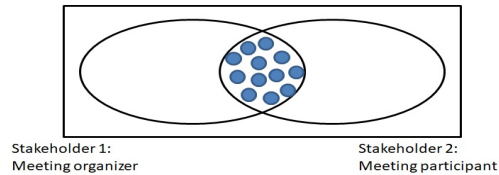

Stakeholder 1:
Meeting organizer

Stakeholder 2:
Meeting participant

Fig. 1. Step 1 of POEPMcreate: Requirements that both SH1 and SH2 need.

Table 1. Opinion questionnaire (Translated from German)

	No	Rather no	Partly	Rather yes	Yes
Do you believe that your requirements are complete?					
Do you believe that your requirements are innovative?					
Did you find the technique easy to apply?					

Table 2. Personal profile questionnaire (Translated from German)

	Participant 1	Participant 2
How old are you (in years)?		
Do you already have work experience in the SW domain? (yes/no)		
What is your native language?		

4 Experiment Results

The requirements for all teams were merged and then were sorted alphabetically, in order to help us find repeated requirements (doublets). When we found a set of repetitions of the same requirement, as some were better worded, we chose the one with the clearest wording to be the primary requirement, with the others being considered duplicates. Also, having a merged list of requirements allowed the authors to evaluate each requirement in an unbiased way without knowing which team, and which technique, generated it. The raw data can be found at
<https://cs.uwaterloo.ca/~dberry/FTP_SITE/tech.reports/HMBrawData/>.

4.1 Participants' Profile

The teams in Exp1and Exp2 are profiled in Table 3 (Remark: One ROSEPMcreate participant did not answer the profile questions.) The teams with only native speakers of German are Pb, Pd, Pe, Pg, Pi, Rb, Rc, and Rg through Rj. The teams with only non-native speakers of German are Pf, Rd, and Re. We expected that a participant's native language would not be relevant, as all the participants speak German fluently. The teams with participants who have work experience are Pg, Pj, Ra, Rb, Rc, Rg, and Ri.

4.2 Numbers of Requirements

All told, 282 requirements were found in Exp1 by 11 teams, and 298 requirements were found in Exp2 by 8 teams, for a total of 580. Among the 580 were 487 functional requirements, 90 non-functional requirements, and 3 non-requirements. After identifying 329 duplicates, 248 individual requirements were left. Of these 248 individual requirements, 205 were functional and 43 were non-functional. The number of duplicates seems low if we consider that a meeting planner is not an exotic tool. In the rest of this paper, unless otherwise explicitly stated, a quoted number of requirements includes duplicates. On average, each team in Exp1 found 25.6 requirements, and each team in Exp2 found 37.3 requirements. This difference between the experiments has a statistical significance of 97% according to a one-sided t-test. So, the Exp2 teams found significantly more requirements than the Exp1 teams, perhaps because the Exp2, Heidelberg participants had more RE experience from previous courses than did the Exp1, Stuttgart participants. Also, a participant's work experience and native language might have made a difference. On average, Exp2 participants were 2.3 years older, had more work experiences, and more likely spoke German natively than Exp1 participants.

In Exp1, teams Ra through Rc, 4 out of 10 participants have *work experience* in the software domain, while in teams Pa through Pf, there is no participant with such work experience. So, in Experiment 1, 18% of the participants have work experience. In Exp2, each of the four teams, Pg, Pj, Rg, and Ri, has a participant with work experience, and the other four teams, Ph, Pi, Rh, and Rj, have no participant with experience. So, in Exp2, half of the teams and one third of the participants have work experience. The teams with work experience produced significantly more requirements (two-sided t-test, 90% confidence level). This can partly explain why more requirements were generated in Exp2.

Trying to be creative in a foreign *language* does not necessarily mean that one creates fewer ideas. However, it is possibly more difficult to write them down. In fact, comparing the number of requirements, a one-sided t-test showed that the teams with two native-speaker participants found significantly (95%) more requirements than the teams, Pf, Rd, and Re, with two non-native speakers. This result is probably not an effect of only language because the Exp1 teams Pf, Rd, and Re are from Stuttgart, and in general, Stuttgart teams noted fewer requirements than did Heidelberg teams. When we compare Pf, Rd, and Re with only the native-speaker teams in Stuttgart, the difference was no longer statistically significant.

Table 3. Profile of the participants in Exp1 and Exp2

	Exp1		Exp2	
	Teams Pa – Pf POEPMcreate	Teams Ra – Re ROSEPMcreate	Teams Pg – Pj POEPMcreate	Teams Rg – Rj ROSEPMcreate
No. of participants	12	10	8	8
Average age in years	24.3	22.9	24.8	27,3
No. of participants with work experience in SW domain	0	4	2	3
No. of participants whose native language is German	8	5	6	7

4.3 Numbers of Requirements per Team, Technique, and Step

Table 4 through Table 7 show the number of requirements that each team generated in each step. For each technique, we can see that (1) the teams overall and (2) all but two teams found more requirements in Steps 1 and 2 than in Steps 3 and 4. This difference is more pronounced for ROSEPMcreate than for POEPMcreate, and is not surprising, as ROSEPMcreate includes redundancies, i.e., Steps 3 and 4 ask for requirements for viewpoints that are covered partially by previous steps.

Table 4. No. of requirements generated per team and step with POEPMcreate in Exp1

Step \Team	Pa	Pb	Pc	Pd	Pe	Pf	Total	#/team
1	10	11	9	6	12	5	53	8.833
2	9	9	9	5	8	5	45	7.500
3	3	2	4	6	8	6	29	4.833
4	1	3	6	3	5	6	24	4.000
Total	23	25	28	20	33	22	151	25.17

Table 5. No. of requirements generated per team and step with POEPMcreate in Exp2

Step \Team	Pg	Ph	Pi	Pj	Total	#/team
1	31	6	7	9	53	13.25
2	16	6	11	11	44	11
3	3	6	5	8	22	5.5
4	14	1	5	12	32	8
Total	64	19	28	40	151	37.75

Table 6. No. of requirements generated per team and step with ROSEPMcreate in Exp1

Step \Team	Ra	Rb	Rc	Rd	Re	Total	#/team
1	13	15	11	9	8	56	11.2
2	8	16	6	6	7	43	8.6
3	1	9	0	3	1	14	2.8
4	2	8	0	1	7	18	3.6
Total	24	48	17	19	23	131	26.2

Table 7. No. of requirements generated per team and step with ROSEPMcreate in Exp2

Step \Team	Rg	Rh	Ri	Rj	Total	#/team
1	9	6	18	12	45	11.25
2	11	9	16	12	48	12
3	6	5	6	6	23	5.75
4	7	5	6	13	31	7.75
Total	33	25	46	43	147	36.75

The variance among the teams using any one technique is large. For either technique, the most productive team generated twice as many requirements than the least productive team. ROSEPMcreate teams generated slightly more requirements than POEPMcreate teams. However, in comparing the teams, the steps, and the techniques with respect to the numbers of requirements, no difference was found to be statistically significant, using a χ^2-test with a significance level of 90%, probably due to the large variances among the teams' numbers of generated requirements. These results

suggest that ROSEPMcreate is feasible and that it is at least as effective as POEP-Mcreate.

4.4 Kano categories

To investigate the innovativeness of the requirements, we tried categorizing them by applying to them the Kano categorization [9], which distinguishes three different categories of requirements, according to the users' expectations and the state of the art:

- Basic requirements: When these requirements are satisfied, the customer is indifferent, because the customer expects these requirements to be satisfied. When they are not satisfied, however, the customer is extremely dissatisfied.
- Performance requirements: The more that these requirements are satisfied, the more the customer is satisfied.
- Delighters: These are characteristics that are not required and are thus unexpected. Therefore, if they are missing, the customer is not dissatisfied, but if they are realized, the customer is positively delighted.

Since generating a basic or performance requirement that an application usually has is not very creative, for evaluation of the quality of ideas, we considered only the delighters to be innovative. A creativity technique should support the stakeholders in generating delighters. The first two authors discussed the merged list of generated requirements and compared them to a known meeting planner, allowing them to determine by consensus the numbers of basic requirements and delighters. Among the requirements, there were 193 basic requirements, 175 performance requirements, and 179 delighters. Then we checked which technique better supported generating delighters. Table 8 presents the results of this analysis. In Exp1, teams using POEP-Mcreate generated slightly more delighter requirements than did teams using ROSEPMcreate. In Exp2, the opposite was true. Averaged over both experiments, each technique led to 9.4 delighters per team. According to a χ^2-test at a significance level of 95%, the difference between POEPMcreate's and ROSEPMcreate's distributions of the requirements among the Kano categories is not statistically significant. In Exp1, each team found an average of 6.8 delighters, and in Exp2, each team found an average of 13 delighters. This difference between the experiments is statistically significant at a significance level of 95% with a one-sided t-test. The difference between teams with work experience and those without is significant according to a two-sided t-test at a significance level of 95%: Teams with work experience generated more delighter requirements and produced more requirements: they found an average of 15 delighters, while teams without work experience found an average of 5.4 delighters.

Table 8. Numbers of requirements, categorized according to the Kano categories.

Kano category	All Rs	All Rs (no duplicates)	With POEPMcreate	With ROSEPMcreate
Basic Rs	193 (35.3%)	61 (26.5%)	94 (34.9%)	99 (35.6%)
Performance Rs	175 (32.0%)	74 (32.2%)	81 (30.1%)	94 (33.8%)
Delighters	179 (32.7%)	95 (41.3%)	94 (34.9%)	85 (30.6%)

5   Threats to Validity

The results of the 10 teams applying POEPMcreate and of the 9 teams applying ROSEPMcreate vary a lot. Therefore, the statistical significance and the statistical power of the results are low: the differences observed between the two techniques, although easy to explain, cannot be confirmed to be statistically significant. Some might consider that the participants' being instructed to be creative to be a threat in that it affects the behavior of the participants in applying the techniques. However, since the techniques are creativity techniques, this effect is desired for both techniques. Thus, no one technique obtained any advantage over the other. Also, the standard instructions for many creativity techniques includes the exhortation to be creative. Tables 4 through 7 show that in each technique, Steps 3 and 4 generate fewer requirements than Steps 1 and 2. Also, results show that in each technique, Steps 1 and 2 generate requirements that do not belong to the right stakeholders. These observations together suggest that the techniques were not applied as expected and that the student participants found it difficult to withhold requirements not belonging to the perspective defined for this step. This reality needs to be explored in future work. The teams applying POEPMcreate are not equivalent to the teams applying ROSEPMcreate. Teams applying ROSEPMcreate had more work experience. However, by participants' ages and numbers of native German speakers, they are equivalent.

The Exp2 teams, found more requirements altogether and more delighters than the Exp1 teams. These differences were found to be statistically significant. The participant profiles show that the teams in Exp2 have more native speakers. Being a native speaker and having work experience are two factors that were found to lead to significantly more requirements. These differences probably influenced the results, both in the numbers of requirements generated and in the teams' perceptions of the techniques applied. Finally, there is the standard threat of having used students as participants rather than professional analysts. However, it is reasonable to expect that all university students have had experience with meeting planners, brainstorming, and being creative. Thus, this experiment did not demand of the participants more knowledge than they already knew, and the participants can be considered as being representative of both end users of a meeting planner, the organizer and the attenders.

One threat common with all experiments is that the instructions to participants can be misunderstood.

6   Conclusion

EPMcreate assumes that a requirements idea generation session follows 16 steps during which the analysts generate requirements ideas by focusing on all 16 combinations of two stakeholders' viewpoint. The need for a lighter weight creativity process requiring fewer steps led first to the development of POEPMcreate, including only the 4 steps that cover the disjoint 4 regions of the 2-variable Venn diagram. This paper describes an experiment designed to test the effectiveness of ROSEPMcreate, another variant of EPMcreate, based on a different set of 4 steps covering the same regions

with some redundancy. The results of Exp1 and Exp2 seem to indicate that ROSEP-Mcreate is at least as feasible and as effective as POEPMcreate, thus confirming the findings from just Exp1 [10]. The experiment results show also that changing viewpoint combinations when applying POEPMcreate or ROSEPMcreate helps generate requirement ideas. With each change, the teams generated additional requirements, although not always belonging to the correct viewpoint. The experiments show also that an analyst's work experience and native language influence the number and innovativeness of requirements the analyst generates. The paper also offers Kano categories as a way to evaluate the innovativeness of generated requirement ideas in experiments. Therefore, it is possible to say that the techniques helped the teams also to generate innovative requirements, the delighters in the Kano categorization.

Future work should focus on identifying the factors that determine the success of a technique, e.g., varying (a) the stakeholders whose viewpoints are explored, (b) the order of the steps, and (c) the amount of redundancy in the viewpoint combinations used in the steps.

References

1. Saha, S.K., Selvi, M., Büyükcan, G., Mohymen, M.: A systematic review on creativity techniques for requirements engineering. Proc. Int. Conf. on Informatics, Electronics & Vision (ICIEV), Dhaka, 34–39 (2012). doi: 10.1109/ICIEV.2012.6317443.
2. Lemos, J., Alves, C., Duboc, L., Rodrigues, G.N.: A systematic mapping study on creativity in requirements engineering. Proc. 27th Ann. ACM Symp. on Applied Computing (SAC'12). ACM, New York, NY, USA, 1083–1088 (2012). doi: 10.1145/2245276.2231945.
3. Mahaux, M., Nguyen, L., Mich, L., Mavin, A.: A framework for understanding collaborative creativity in requirements engineering: Empirical validation. Proc. 4th Int. Work. on Empirical Requirements Engineering (EmpiRE), Karlskrona, 48–55 (2014).
4. Mich, L., Anesi, C., Berry, D.M.: Applying a pragmatics-based creativity-fostering technique to requirements elicitation. Requirements Engineering. 10(4), 262–275 (2005).
5. Sakhnini, V., Mich, L., Berry, D.M.: The effectiveness of an optimized EPMcreate as a creativity enhancement technique for Web-site requirements elicitation. Requirements Engineering. 17(3), 171–186 (2012). doi: 10.1007/s00766-011-0133-0.
6. Mich, L., Sakhnini, V., Berry, D.M.: Downsizing EPMCreate to lighter creativity techniques for requirements elicitation. Proc. 6th Work. on Creativity in Requirements Engineering, (CreaRE) at REFSQ. (2017). http://ceur-ws.org/Vol-1796/creare-paper-2.pdf.
7. Preparata, F.P., Yeh, R.T.Y.: Introduction to Discrete Structures for Computer Science and Engineering. Addison-Wesley Longman, Boston (1973).
8. Osborn, A.F.: Applied Imagination: Principles and Procedures of Creative Problem Solving (3rd Revised Edition). Charles Scribner's Sons, New York (1963).
9. Kano, N.: Attractive quality and must-be quality. Journal of the Japanese Society for Quality Control. 4, 39–48, (1984).
10. Herrmann, A., Mich, L., Berry, D.M.: Creativity techniques for requirements elicitation: Comparing four-step EPMcreate-based processes. Proc. 7th Int. Work. on Empirical Requirements Engineering (EmpiRE). pp. 1–7. IEEE (2018).