A Controlled Experiment of a Method for Early Requirements Triage Utilizing Product Strategies

Mahvish Khurum[1], Tony Gorschek[1], Lefteris Angelis[2], and Robert Feldt[1]

[1] Blekinge Institute of Technology, Department of Systems and Software Engineering,
S-372 25, Ronneby, Sweden
mkm@bth.se, tgo@bth.se, rfd@bth.se
[2] Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
lef@csd.auth.gr

Abstract. [Context and motivation] In market-driven product development of software intensive products large numbers of requirements threaten to overload the development organization. It is critical for product management to select the requirements aligned with the overall business goals, product strategies and discard others as early as possible. Thus, there is a need for an effective and efficient method that deals with this challenge and supports product managers in the continuous effort of early requirements triage [1, 2] based on product strategies. This paper evaluates such a method – A Method for Early Requirements Triage Utilizing Product Strategies (MERTS), which is built based on the needs identified in literature and industry. **[Question/problem]** The research question answered in this paper is "If two groups of subjects have a product strategy, one group in NL format and one in MERTS format, will there be a difference between the two groups with regards to effectiveness and efficiency of requirements triage?" The effectiveness and efficiency of the MERTS were evaluated through controlled experiment in a lab environment with 50 software engineering graduate students as subjects. **[Principal ideas/results]** It was found through results that MERTS method is highly effective and efficient. **[Contribution]** The contribution of this paper is validation of effectiveness and efficiency of the product strategies created through MERTS method for requirements triage, prior to industry trials. A major limitation of the results is that the experiment was performed with the graduate students and not the product managers. However, the results showed that MERTS is ready for industry trials.

Keywords: Market driven requirements engineering, requirements triage, product strategies, MERTS, experiment, effectiveness and efficiency.

1 Introduction

Due to the emergence of markets for off-the-shelf or packaged software [3, 4], market-driven development is gaining increased interest in comparison to customer-specific system development [5, 6]. As a consequence a shift in focus is occurring, affecting software development in general and requirements engineering in particular [6]. In

contrast to traditional requirements engineering, requirements in market-driven requirements engineering (MDRE) to a large extent come from internal sources such as developers, marketing, sales teams, support, bug reports, as well as from external sources such as different users, customers from different and multiple market segments, and competitors [7]. The result is a large and continuous flow of requirements that threaten to overload the development organization [5]. This has two major implications. One, the product and domain knowledge reside largely with the development company itself. For example a developer of robotics products with many of e.g. car manufacturers as customers probably knows more about robotics than any one customer.

Two, the risk and cost of development is carried by the development organization, meaning that the potential revenues depend on selecting the "right" requirements for implementation. The selection accuracy is the main success criteria for the development organization, and being able to perform the selection in a scalable and cost effective way is crucial to avoid overloading. Which requirements to select is a trade-off between different strategic factors such as key-customer requirements and long-term aspects and innovation efforts. All of these factors, and more, need to be explicitly stated and weighed together to reach an optimal strategy for the company, which can then be used for selecting the "right" requirements for implementation.

However, while industry managers regard strategy formulation and use as the most important aspect of technology management [8], strategy formulation is usually performed ad-hoc, and a systematic approach for formulating strategies is often missing in practice [9]. Even if the formulation of strategies was pursued, the factors affecting strategy formulation differ between different stakeholders. Strategic and middle management and technical experts all need to share one vision. Strategic managers often overlook the technical perspective, and technical experts can be unaware of or overlook the strategic managers' perspective. As a result of these challenges, identified both in academia and through industry case studies, a Method for Early Requirements Triage and Selection (MERTS) [10] was developed to combine both strategic and technical perspectives for the formulation of product strategies that are good-enough to be used for early requirements triage and selection.

This paper presents an experiment testing some key aspects of this method, following a stepwise plan to validate MERTS prior to industry piloting.

MERTS has two main purposes. First, it acts as a stepwise guide to creating product strategies taking both strategic and technical views into account thus following a systematic way of agreeing on a joint plan. Secondly, the strategies resulting from MERTS can be used by product managers to effectively perform requirements triage and requirements selection in a reasonable amount of time as spending initial 10 minutes on triage versus 10 hours is super critical for industry. The experiment aims at testing the second purpose of MERTS. Thus, the main purpose of the experiment is to assess the efficiency and effectiveness of requirements triage utilizing strategy formulated and formatted using MERTS prior to industry piloting. Thus, this experiment is considered as a lab validation following the research approach suggested by Gorschek et al. [11] aimed at producing useable and useful research results and successful technology transfer.

Before describing the experiment and experiment results an introduction to MERTS is given in Section 2. Section 3 details the experiment design. Section 4 lists

the validity threats. Section 5 contains preparation and execution details. Section 6 presents the results and analysis, and finally Section 7 presents the conclusions drawn and plans for further work.

2   MERTS Background

MERTS is centered on ensuring that the five strategic questions for a product are answered explicitly [10]. Fig. 1 gives an overview of MERTS and the three main parts of the method. The goal of MERTS is to offer a clear method detailing how to reach consensus and a homogenous understanding of a product strategy. The product managers using the method are required to follow these three parts. Each part has several steps (see Fig. 1).

Part One – Early Requirements Triage. This part provides steps to create an initial product strategy for use in requirements triage.

   A. Specify. In order to explicitly state the goals and objectives of a product, it is important to specify the directions of movement for the product deduced from the organization's mission statement. Thus it is important to answer the three strategic questions ((1) Where we want to go?, (2) How to get there?, (3) What will be done?) for each product.
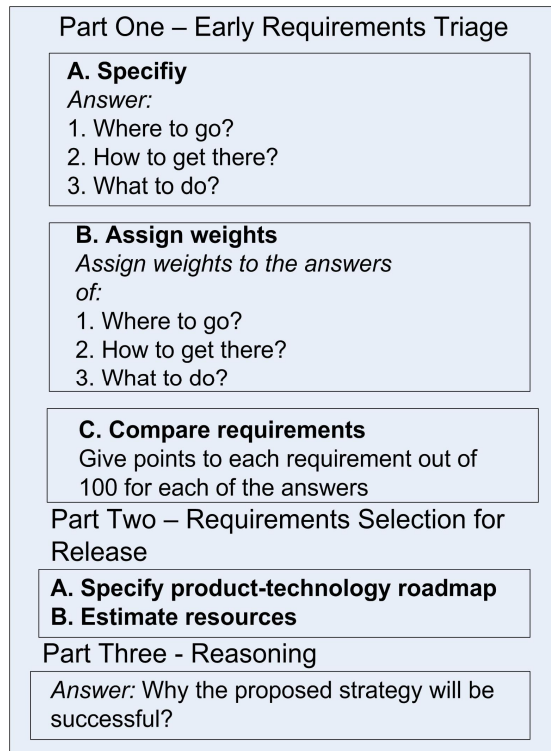
> **Part One – Early Requirements Triage**
>
> > **A. Specifiy**
> > *Answer:*
> > 1. Where to go?
> > 2. How to get there?
> > 3. What to do?
>
> > **B. Assign weights**
> > *Assign weights to the answers of:*
> > 1. Where to go?
> > 2. How to get there?
> > 3. What to do?
>
> > **C. Compare requirements**
> > Give points to each requirement out of 100 for each of the answers
>
> **Part Two – Requirements Selection for Release**
>
> > **A. Specify product-technology roadmap**
> > **B. Estimate resources**
>
> **Part Three - Reasoning**
>
> > *Answer:* Why the proposed strategy will be successful?

Fig. 1. MERTS Steps

The output of this step is an explicit understanding of goals and objectives associated with a specific product which can be used to perform requirements triage and selection for individual products.

To answer (A.1) "**Where to go**" the organization's directions of movement have to be clearly stated. An organization can have one or many directions of movement. For example, shareholders' revenue, profit, growth, and market share [10]. The answer to this question depends on identified directions of movement and their relative importance.

The answer to (A.2) "**How to get there**" will bind the strategy in terms of customer segments and competition targeted and differential advantage of the individual product providing a unique selling point. In order to answer this question there is a need to specify:

- Different customer segments targeted by a specific product, e.g. one customer segment can be the Asian market and another can be the European market. By explicitly specifying customer segments, relative priorities can also be assigned, helping in the selection of requirements. Customer segments can be defined either on a higher level of abstraction or refined depending on the needs of the organization.

- Competitors to a product to show which ones to target. This enables features provided by relevant competitors to be included in the product offering. Just as customer segments, competitors can also be prioritized relatively, giving more importance to features provided by high priority competitors.

- Differential advantage(s) of the product that makes it unique in the market place in relation to competitors. The differential advantage can be based on any one (or combination) of technology, pricing, strategic alliances and non-functional requirements. These can also be prioritized in relation to each other depending on their importance to offering the advantage. By identifying the differential advantages and prioritizing them, it is possible to ensure that all requirements are weighted against them and requirements providing unique differential advantages are not missed.

For the answer to (A.3) "**What to do**" a more management centered perspective can be used, focusing on product pricing, promotion, distribution, and service. However, since MERTS is targeted towards early requirements triage and selection, answers to this question will focus on the abstract technical considerations of a requirement. Some of the possible considerations rated highest by the technical experts during the interviews have been taken as example here, i.e. innovation, core assets, architecture stability, market-pull, technology-push, customization flexibility, and use of COTS [10]. Priorities can be assigned to each of these factors showing their relative importance with respect to each other.

B. Assign Weights. The answers from Step 1 are assigned weights. The rule is to assign weights to each of the factors based on their relative importance in a way that total weight remains 100. This way has been reported to be one of easiest and quickest prioritization methods [17].

C. Compare Requirements. The total weights of all the requirements are compared against a threshold to select or reject each of the requirements.

The first three steps of MERTS should be performed at product management level supporting the triage of requirements (aiding in the selection). The purpose of step 2 (Assign weights) is not requirements prioritization which is usually associated with

early project activities during release planning. The points assigned to each requirement, against each factor or sub-classification, show the level of strategic alignment.

Part Two – Requirements Selection for Release. After a set of requirements (deemed to be aligned with the strategy) have been selected, the question in focus is "when to get there". To answer this following two steps are required.

A. Specify product-technology roadmap. It has been emphasized in literature [12] to chalk out a product-technology roadmap to get an overview of the relationship between product releases (product evolvement) and successive technology generations. This means specifying what a product tends to achieve along the time axis in term of its evolvement and technology trends. This enables placement of requirements in appropriate intervals planned in a roadmap. For example, if a requirement requires expertise in some new technology to be explored in the future and this has been planned in the roadmap, the requirement can be postponed or depending on the urgency of the requirement, the roadmap can be altered.

B. Estimate resources. In order to determine the feasibility of the requirements, the organization needs to explicitly state financial and effort allowances against each interval in the roadmap. Several methods can be used to estimate cost, effort and time, e.g. feature points, function points, lines of code, and methods like e.g. COCOMO [13] can be used to support the efforts. An alternative could be to perform estimates based on previous development efforts. Additionally, requirements prioritization techniques [14] can be used to plan releases for the product.

Part Three – Strategy Rationale. Once the strategic questions have been answered, it is important to document the reasoning behind the decisions. This way if the decisions (and indirectly the answers) result in success (of a product) replication can be achieved, and the organization has good examples to follow for future efforts.

In addition, the strategy formulated through MERTS should be used to share product and organizational visions across the organization. In its simplest form it can mean writing a paragraph explaining the reason behind the answers, keeping in view the organization's long term goals, financial plans, technology trends and marketing trends.

In order to implement MERTS method, requirements need to be comparable to the strategies formulated. The reasoning is that MERTS is meant to assist in early requirements triage and selection. In case of requirements being too technical or too detailed method usage will not be efficient because it will be difficult to compare detailed technical requirements with strategies as strategies are formulated on a higher level of abstraction. Even if there is some process to compare detailed technical requirements with strategies they will still be too many detailed requirements to be compared against strategies. Often many detailed requirements form one product level feature/requirement therefore it is pointless to compare every detailed requirement against the strategies. Any method for abstracting the requirements can be used, e.g. the RAM model by Gorschek and Wohlin [6], as long as it produces requirements on an abstraction level comparable to product strategies.

3 Design of the Controlled Experiment

The usefulness and the usability of MERTS depend on several concepts that need to be assessed; the one studied in this controlled experiment is that it should be possible

to perform triage of a new incoming requirement based on its alignment with the MERTS strategy. This means that a MERTS strategy should be usable in an efficient and effective manner for performing requirements triage (formally stated as research questions in Section 3.3).

3.1 Context

The experiment was conducted in an academic setting, with the help of 50 engineering graduate students at Blekinge Institute of Technology. It was conducted as a mandatory although non-graded exercise at the end of a 7.5 ECTS merits master's course in research methodology. Participation was mandatory and despite the ethical issues of forcing subjects to participate in a study, it was believed that the experiment had several pedagogical benefits in the course. The students were instead given the option to exclude their individual results from the study, an option not utilized by any student. The intended users of MERTS, however, are product managers with several years of experience in a specific domain and product. In the experiment, the subjects have no training in using MERTS, they possess limited domain knowledge, are under time pressure, and most of them have not seen the product strategies or the requirements before. There is thus a considerable gap between the intended target group and the sample used in this experiment. The subjects in this experiment can be expected to adapt a more surface oriented approach to the problem than product managers. We argue that this works to our advantage, since any results that we evaluate are likely to stem from the instrumentation and the use of MERTS, rather than previous experiences of the subjects in the study. If MERTS proves to be usable in the experiment, it would indicate that it is able to decrease the dependency on individual's experience, product knowledge, and methodology.

3.2 Subjects

The group of experiment subjects using MERTS strategy for requirements triage had an average of 1.46 years of industrial experience, and only 3 out of 25 subjects had seen project strategies or performed requirements triage before. The subjects in the group using a natural language (NL) strategy for requirements triage had an average experience of 2.23 years, and 4 subjects out of 25 had seen product strategies in some form in their companies and 5 had performed requirements triage previously. This information was gathered through a post-experiment questionnaire; the groups were formed randomly.

3.3 Research Questions

The main research question is RQ which is described below along with associated hypotheses and independent/dependant variables.

RQ: If two groups of subjects have a product strategy, one group in NL format and one in MERTS format, will there be a difference between the two groups with regards to effectiveness and efficiency?

Hypotheses
Null hypothesis, $H_{o\ Effectiveness}$: The use of MERTS strategy for requirements triage is not significantly different from NL strategy with regards to effectiveness.

Alternative hypothesis, H_1 Effectiveness: The use of MERTS strategy for requirements triage is significantly different from NL strategy with regards to effectiveness.

Null hypothesis, H_o Efficiency: The use of MERTS strategy for requirements triage is not significantly different from NL strategy with regards to efficiency.

Alternative hypothesis, H_1 Efficiency: The use of MERTS strategy for requirements triage is significantly different from NL strategy with regards to efficiency.

Variables Selection: This experiment has the following independent variables:

Independent variables
Product strategy formatted according to MERTS or according to NL.
Dependant variables
The dependant variables are *effectiveness* and *efficiency* measured through:
1. Effectiveness: Number of correct requirements triage decisions.
2. Efficiency: Time taken (in minutes) to perform triage on all requirements.

The definition and hypotheses for finding an answer to RQ depict that the design is: *one factor with two treatments*. The factor is the product strategy and treatments are NL and MERTS.

3.4 Design and Instrumentation

Prior to the experiment execution one round of validation (pre-test) was performed to refine the experiment design and instrumentation. The pre-test was constructed and evaluated with the help of 4 colleagues at Blekinge Institute of Technology. Based on the experience from this pre-test, the experiment package was revised. Specifically, the initial presentation describing the study and running through an example was refined to make it more concrete for describing the motivation of the triage decisions taken by the subjects.

The subjects were divided randomly into two groups, with one treatment per group. The experiment consisted of two parts that both ran consecutively without breaks. The first part was a preparatory lecture where the concepts of requirements triage and MERTS/NL were introduced, together with a presentation of the experiment, research instruments and an example of how to take triage decisions and provide motivations. The second part of the experiment consisted of filling the forms. All other artifacts like the requirements and forms were the same.

During the experiment, the following instruments were used:

- Each subject was given an example of how to perform triage using either NL or MERTS (depending on the group).

- Each subject was given either NL or MERTS formatted strategy for the experiment. The product strategy detailed the goals of a new version of a mobile phone targeted for entertainment-oriented users in the Asian market. The level of information in the two strategies was the same with respect to goals and objectives, targeted customers and competitors, differential advantages and technical considerations. NL strategy was formulated based on example strategies given in literature. Industrial experience of authors with real product strategies was also beneficial to ensure that NL formatted strategy was as close as possible to industry practice. The MERTS strategy however, as

prescribed by the MERTS method, had weights assigned to each of the factors stated in the strategy which was absent in NL strategy because in traditional NL strategies the weights to each of the factors is not explicitly given in numbers rather stated as subjective statements.

   - The requirements set contained 13 product and 18 feature level requirements. For example, messages communication, music playing, enhanced imagining, enhanced display, availability, usability, browsing, connectivity, and so on. The requirements were constructed to be of moderate to good quality based on industry observation. The appropriateness of the requirements and other instruments was also validated in the pre-test. It is important to understand that in lab experimentation, it is not possible to have a large number of requirements for triage. There is a limited amount of time where subjects have to understand the method and then apply it for requirements triage. The aspects of effectiveness and efficiency as evaluated in the experiment are however related to using MERTS strategies vs. NL strategies. The relative efficiency and effectiveness is the goal.

   - Each requirement in the set has at least two levels: product and feature, and often also divided into functions. Each requirement was formatted and specified using the following attributes; Unique Id, Product level requirement, Feature level requirement, Function level requirement, Component level requirement (in some cases) and Comments.

   - The instrumentation had a *Decision* column next to every feature level requirement with two options: Accept and Reject. For every triage decision the experiment subject had to specify a rationale behind the triage (Accept or Reject) decision. It was emphasized during the experiment training that the motivation had to be deduced from the product strategy and not personal judgments and opinions.

   Last in the experiment, each subject had to answer the questions at the end of experiment as a post-test. The experiment materials (NL strategy, MERTS strategy, example requirements and the post-test) is not included in the paper as space does not allow, but can be obtained online at http://www.bth.se/tek/aps/mkm.nsf/pages/merts-experimentation.

4   Validity Evaluation

Internal validity. This threat can have a huge impact on the experiment results if the data collection forms and other instruments are poorly designed. To ensure that the research instruments, including the posed question, are of a good quality, one pre-test with the material was executed before the "live" round. Moreover, all subjects received the same introductory lecture, and were given the same material in the same order. It is thus unlikely that the instrumentation and the implementation of the treatment influenced the results unduly. That being said, since we used the answers of human subjects as measures, the gathered measures are of course not 100% repeatable.

   To alleviate author's bias towards MERTS while designing the experiment, a senior researcher (the second author) not involved in the creation of MERTS, was actively involved in the design of the experiment to avoid favoritism towards MERTS.

Construct validity. To reduce the risk of evaluation apprehension among the test subjects, they were told that they would be graded on their efforts, but not on the number of correct decisions.

External validity. To ensure the external validity and the ability to generalize the results, we use a requirements specification from a fairly mature domain. As discussed in Section 3.2, the knowledge and experience of the participants is less than that of the target audience (e.g. product managers in industry). To reduce this gap, a system from a domain that is familiar to the subjects was used.

The correlation analysis between the total number of correct triage decisions and the industrial experience show that there was no significant difference between performance of subjects with more industry experience and those with less experience (both for the group using MERTS strategy and group using NL strategy). Thus, the two groups were homogenous in terms of industry experience.

As the intended target of MERTS (e.g. product managers) would have not only a better requirement and domain understanding, but also more experience in triage, it can be argued that the ability to use MERTS (and the potential positive results obtained in the experiment) should be transferrable to industry practice. Moreover, experimentation using state-of-the-art research (well-structured method MERTS in this case) also has learning/training benefits for future professionals.

In this study paper printouts were used, which may impact the readability and the ease by which the participants may access the information. Hence, any positive effects are also here transferrable to the target audience and the target environment as the use of tools may increase usability.

5   Operation

The subjects were not aware of the aspects intended for study, and were not given any information regarding research questions in advance. They were aware of the fact that it was a controlled experiment in the area of requirements engineering that was a part of their research methodology course. The experiment ran over a period of three hours, and the subjects were randomly divided into two groups seated in two different rooms. Introduction to the experiment was given during these three hours in the form of a brief slide show presentation. In this presentation basic concepts of product strategy and requirements triage were explained along with examples.

The mean time to conduct the experiment was around 60 minutes when using MERTS strategy, the shortest time spent was around 33 minutes and the longest was 107 minutes. The group using NL strategy had a mean time of around 33 minutes, the shortest time spent was 17 minutes and the longest was 50 minutes.

6   Results and Analysis

6.1   Testing $H_{o\ Effectiveness}$

In each group 18 feature level requirements were given to the subjects and they had to decide which of these are to be selected/rejected in accordance with the product strategy (either MERTS or NL). According to the experiment design 10 feature level

requirements were to be selected and 8 rejected based on the product strategies. During this analysis, answers that were in line with the study design and aptly motivated were treated as "correct". If an answer is in line with the study design but missing a proper motivation (that is the motivation is not based on the given product strategy) or if the answer is not in line with the study design, the answer is considered "incorrect".

Table 1 shows the mean, standard deviation, skewness and kurtosis for the total number of correct decisions for all the 18 feature level requirements for the two strategies: MERTS and NL respectively. The results show that the average number of correct decisions using MERTS (Mean = 17.72) is more than double the average number of correct decisions using the NL (Mean = 6.22).

Table 1. Statistics for total number of correct decisions for MERTS and NL strategies

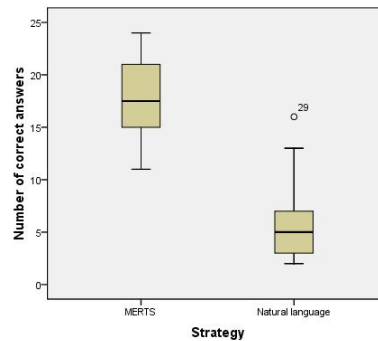| MERTS		Natural Language	
Statistic	**Value**	**Statistic**	**Value**
Mean	17.72	Mean	6.22
Median	17.50	Median	5.00
Std. deviation	4.456	Std. deviation	4.124
Skewness	-.143	Skewness	1.180
Kurtosis	-1.257	Kurtosis	0.639

Fig. 2. Boxplots for total number of correct decisions using the two strategies

Confirmed complementary view is offered by the boxplots in Fig. 2 where the greater number of correct triage decisions using MERTS strategy is evident. Through the boxplots, an outlier (marked with a small circle and 29) was identified which is discussed below.

Table 2. Normality tests for total number of correct decisions

Strategy	Kolmogorov-Smirnov		Shapiro-Wilk	
	Statistic	Sig.	Statistic	Sig.
MERTS	0.158	0.200	*0.923*	0.146
NL	0.203	0.048	0.862	0.013

The skewness and kurtosis values for the total number of correct triage decisions show that the distributions seem to differ from the normal distribution. To check normalization prior to the application of an appropriate statistical test, normality tests were performed on the given data and the results are shown in Table 2. It can been in

Table 2 that the total number of correct triage decisions for MERTS do not differ significantly from the normal distribution (Significance = 0.20 > 0.05) but the distribution of the total number of correct triage decisions for NL is not normally distributed (Significance = 0.048 < 0.05). Based on this result the Mann-Whitney U test was used to compare if the two sample distributions (of total number of correct decisions using MERTS and NL strategies) come from the same population.

Looking at overall effectiveness of MERTS versus NL strategy the bar chart in Fig. 3 confirms that MERTS was more effective for triage decisions than NL.
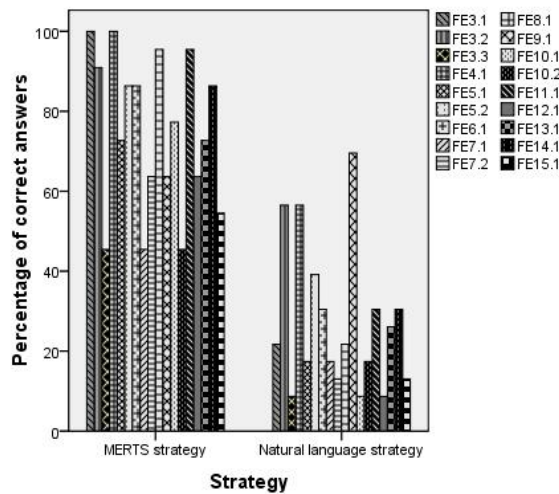

Fig. 3. Percentage of correct decisions in relation to strategy used

Additionally, the visual inspection of cross tabulations in Table 3 shows that the percentages of correct decisions for MERTS are significantly higher than the correct decisions for NL. For example, looking at second row it is possible to see that 22 subjects (62.9%) using MERTS strategy made a correct triage decisions for FE3.2 whereas only 13 subjects (37.1%) using NL strategy made a correct decisions. The difference of the percentages of correct and incorrect decisions between the two groups was tested with the chi-square test. For all the requirements the difference is significant at 0.025 except the requirement 9.1. This is the same requirement for which an outlier has been identified in Fig. 2. The reason behind this is that for this particular requirement the total number of correct decisions both for MERTS and NL are equal (16 correct decisions).

Requirement 9.1 is shown below with its related Function and Product level requirements. This requirement is easy to relate to in both strategy formulations, and also during the requirements engineering course at the university, students were given an example of a very similar requirement. In perfect hindsight it was not surprising that 50% subjects got this requirement decision correct both in MERTS strategy and NL strategy. The conclusion drawn after analysis was that the use of this particular requirement in the experiment was less than optimal.

Requirement 9.1.
Product: PR9: Usability Internationally **Feature: FE9.1: The mobile shall support multiple languages.** Function: FN9.1.1: The mobile shall provide Swedish language support FN9.1.2: The mobile shall provide Chinese language support FN9.1.3: The mobile shall provide Japanese language support.

Table 3. Significance using Chi-Square Test

Requirement	Sig.	MERTS				Natural Language			
		Incorrect		*Correct*		*Incorrect*		*Correct*	
		Count	%	Count	%	Count	%	Count	%
FE3.1	0.000	0	0	25	83.3	20	100	5	16.7
FE3.2	0.002	2	14.3	22	62.9	12	85.7	13	37.1
FE3.3	0.003	13	36.1	11	84.6	23	63.9	2	15.4
FE4.1	0.000	0	0	25	65.8	12	100	13	34.2
FE5.1	0.000	7	25	17	81	21	75	4	19
FE5.2	0.000	3	15.8	21	70	16	84.2	9	30
FE6.1	0.000	3	14.3	22	75.9	18	85.7	7	24.1
FE7.1	0.024	13	38.2	11	73.3	21	61.8	4	26.7
FE7.2	0.000	9	29	15	83.3	22	71	3	16.7
FE8.1	0.000	1	4.8	24	82.8	20	95.2	5	17.2
FE9.1	1.000	9	50	16	50	9	50	16	50
FE10.1	0.000	6	22.2	18	90	21	77.8	2	10
FE10.2	0.047	14	42.4	11	73.3	19	57.6	4	26.7
FE11.1	0.000	1	5.3	23	26.7	18	94.7	7	23.3
FE12.1	0.000	10	31.2	15	88.2	22	68.8	2	11.8
FE13.1	0.002	8	29.6	17	73.9	19	70.4	6	26.1
FE14.1	0.000	3	15	22	75.9	17	85	7	24.1
FE15.1	0.001	11	33.3	14	82.4	22	66.7	3	17.6

Finally to confirm the results, the Mann-Whitney U test is applied in order to check the significance of the results. Significance less than 0.001 was attained, indicating that there is a significant difference between the means of the two groups. The null hypothesis: $H_{oEffectiveness}$ is rejected and $H_{1Effectiveness}$ is confirmed, i.e. using MERTS is significantly different from NL for requirements triage with regards to effectiveness. To conclude, the use of MERTS strategy for requirements triage is superior to NL strategy with regards to effectiveness.

6.2 Testing $H_{0\ Efficiency}$

Fig. 4 shows the mean, standard deviation, skewness and kurtosis values for the time taken by the 50 subjects using the MERTS and NL strategies. The results show that average time taken using MERTS (Mean = 60.12) is double the average time taken using NL (Mean = 33.44). The outlier identified in Fig. 4 contributes to the large mean and standard deviation for the triage time taken using MERTS (Std. Deviation 19.10).

Table 4. Total time taken (minutes) for MERTS and NL strategies

MERTS		Natural Language	
Statistic	**Value**	**Statistic**	**Value**
Mean	60.12	Mean	33.44
Median	59.00	Median	34.00
Std. deviation	19.10	Std. deviation	9.10
Skewness	0.93	Skewness	-0.06
Kurtosis	0.62	Kurtosis	-0.92
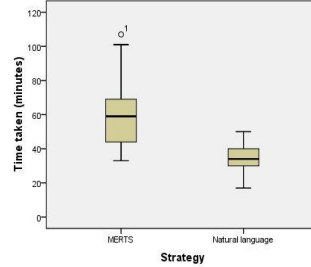
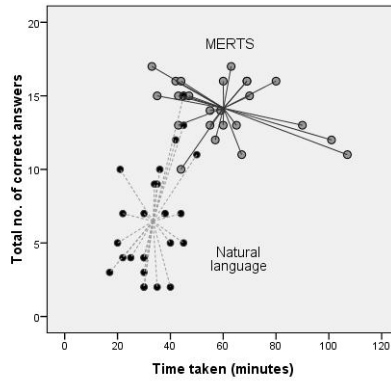**Fig. 4.** Boxplots for time taken (minutes) for the two strategies

Fig. 5. Number of correct decisions versus time taken

Fig. 5 shows a scatter plot of the points representing each of the subject's responses in two dimensions (the dependant variables), the total number of correct decisions and the total time taken (in minutes). The spikes show the distance from the centroid. A clear difference between the two treatments can be seen. MERTS is characterized by long times and greater number of correct decisions whereas NL is characterized by shorter times and fewer correct decisions.

However, an analysis of the ratios of the total number of correct decisions in relation to total time taken using MERTS strategy (ratio value = 0.2355) and NL strategy (ratio value = 0.1937) shows that the time taken to perform correct triage decisions utilizing MERTS is only 0.12 times more than the time to perform correct triage decisions utilizing NL.

Using MERTS strategy, the number of correct decisions far outweigh the number of correct decisions using NL, thus it can safely be stated that MERTS has a fairly equivalent efficiency compared to NL, even if at a first glance MERTS may seem much more resource demanding.

Nevertheless, the subjects in the experiment that used MERTS did spend more time in total, and per correct decision, even if the latter was only marginal. A potential

explanation could be that the subjects using MERTS had to explicitly write a motivation/explanation referring to the strategy for every answer. This qualification of their decisions was not present on the NL side to the same extent as the NL strategy formatting was less exact the motivations were more of the character "could be good to have". The main motivation for demanding a thorough mapping between answers (choosing to accept or dismiss a requirement) and the MERTS formulated strategy was to enable decision traceability, a added value of MERTS that is not a part of the evaluation presented in this paper.

This might not explain the entire time difference, but at least parts of it. Using Mann-Whitney U, significance less than 0.001 was attained, indicating that there is a significant difference between the means of the two groups. This means that the null hypothesis $H_{o \ Efficiency}$ is rejected, and thus $H_{1 \ Efficiency}$ is confirmed i.e. i.e. using MERTS is significantly different from NL for requirements triage with regards to efficiency. However, it cannot be concluded that use of MERTS for correct triage decisions is superior to the use of NL strategy with regards to efficiency. If the hypothesis was formulated as efficiency per correct answer, and if the time taken to write explicit qualification for the MERTS group was taken into consideration we feel confident that MERTS would be as efficient as NL, if not more.

7   Conclusions

MERTS is intended to aid product managers in performing requirements triage effectively and efficiently in a repeatable manner providing traceable decisions.

In this experiment the effectiveness and efficiency of requirements triage using MERTS was compared to using NL formulated strategies, which is the norm in industry. The main motivation of the experiment was to validate MERTS prior to industry piloting as such real industry tests require valuable and hard to obtain resources.

The experiment subjects were given 18 feature level requirements and asked to accomplish a considerable amount of work in a relatively short amount of time. The subjects were expected to form an understanding of the concept of product strategy and requirements triage, understand the domain (read and understand the requirements) and then take decisions whether to include or exclude a specific requirement based on the strategy supplied. The subjects were offered very little training, and they also possessed little prior knowledge regarding the domain compared to the level of a product manager in industry. Considering these aspects and the total number of correct decisions that resulted in using MERTS we feel that it is safe to draw the conclusion that MERTS is far superior to NL when it comes to strategy formulation and utilization for the purpose of requirements triage. The only potential drawback is that MERTS seems to be more resource intensive to use, although per correct answer we think that MERTS is at least as efficient as the NL option. Moreover, MERTS is essentially a systematic method for thinking and making decisions and that is why it takes more time but avoids errors. This systematic work is missing when using NL strategies.

The characteristics of industry are also relevant as real-life requirements triage utilizing product strategies would probably be easier for product managers than for the subjects in this controlled experiment. In industry, requirements triage and selection is not performed in isolation, regular meetings as well as official and unofficial conversations and discussions help in sharing views and reaching consensus. The benefit of MERTS is the ability to document the strategies (and the consensus) in a way that offers explicit decision support for all decision makers when performing requirements triage.

Considering these aspects, the results revealed through this experiment appear even more promising. In addition, product managers in industry are well versed in both their specific domain and in the field of requirements engineering. Given this, the use of MERTS would likely ensure even greater effectiveness and efficiency than was observed during the controlled experiment presented in this paper.

References

1. Davis, A.M.: The art of requirements triage. IEEE Computer 36, 42–49 (2003)
2. Simmons, E.: Requirements Triage: What Can We Learn from a "Medical" Approach? IEEE Software 21, 86–88 (2004)
3. Carmel, E., Becker, S.: A process model for packaged software development. IEEE Transactions on Engineering Management 42, 50–61 (1995)
4. El Emam, K., Madhavji, N.H.: A field study of requirements engineering practices in information systems development. In: Proceedings of the Second IEEE International Symposium on Requirements Engineering, pp. 68–80. IEEE Computer Society, Los Alamitos (1995)
5. Karlsson, L., Dahlstedt, Å., Nattoch Dag, J., Regnell, B., Persson, A.: Challenges in Market-Driven Requirements Engineering - an Industrial Interview Study. In: Proceedings of the Eighth International Workshop on Requirements Engineering: Foundation for Software Quality (REFSQ 2002), pp. 101–112. Universität Duisburg-Essen, Essen (2003)
6. Gorschek, T., Wohlin, C.: Requirements Abstraction Model. Requirements Engineering journal 11, 79–101 (2006)
7. Potts, C.: Invented Requirements and Imagined Customers: Requirements Engineering for Off-the-Shelf Software. In: Proceedings of the Second IEEE International Symposium on Requirements Engineering, pp. 128–130. IEEE, Los Alamitos (1995)
8. Scott, G.M.: Top priority management concerns about new product development, vol. 13. The Academy of Management Executive (1999)
9. Krishnan, V., Karl, T.U.: Product Development Decisions: A Review of the Literature. Manage. Sci. 47, 1–21 (2001)
10. Khurum, M., Aslam, K., Gorschek, T.: MERTS – A method for early requirements triage and selection utilizing product strategies. In: APSEC 2007, Nagoya, Japan (2007)
11. Gorschek, T., Garre, P., Larsson, S., Wohlin, C.: A Model for Technology Transfer in Practice. IEEE Softw. 23, 88–95 (2006)
12. Kappel, T.A.: Perspectives on roadmaps: how organizations talk about the future. Journal of Product Innovation Management 18, 39–50 (2001)
13. Fenton, N.E., Pfleeger, S.L.: Software Metrics - A Rigorous & Practical Approach. International Thomson Computer Press (1996)
14. Berander, P.: Evolving Prioritization for Software Product Management. APS, PhD. Blekinge tekniska hogskola (2007)