Evaluating the Practical Use of Different Measurement Scales in Requirements Prioritisation

Lena Karlsson
Dept. of Communication Systems
Lund University
Box 118, 221 00 Lund, Sweden
+46 46 222 03 65
lena.karlsson@telecom.lth.se

Martin Höst
Dept. of Communication Systems
Lund University
Box 118, 221 00 Lund, Sweden
+46 46 222 90 16
martin.host@telecom.lth.se

Björn Regnell
Dept. of Communication Systems
Lund University
Box 118, 221 00 Lund, Sweden
+46 46 222 90 09
bjorn.regnell@telecom.lth.se

ABSTRACT

The importance of prioritising requirements is widely recognised. A number of different techniques for prioritising requirements have been proposed, some based on an ordinal scale, others on a ratio scale. Some measurement scales provide more information than others, i.e. the ratio scale is richer than the ordinal scale. This paper aims to investigate the differences between the scales used in prioritisation. This is important since techniques using a richer scale tend to be more time-consuming and complex to use. Thus, there is a trade-off between simple techniques only providing ranks and complex techniques providing information about the relative distance between requirements priorities. The paper suggests an approach to measure the skewness of the ratio distribution and a way to use the cost-value approach on ordinal scale data. Four different empirical data sets were used to verify the suggested approaches. The skewness measure seems feasible to determine in which cases the ratio scale is valuable. It indicates that some of our subjects tend to use the extreme values of the scale while others are more modest. The cost-value approach based on ordinal scale data also seems feasible. The requirements selection decisions based on ordinal scale data agree substantially with the decisions based on ratio scale data.

Categories and Subject Descriptors

D.2.8 [**Software Engineering**]: Metrics – *Process metrics*

General Terms

Management, Measurement, Experimentation

Keywords

Requirements prioritisation, ordinal scale data, ratio scale data, cost-value approach

1. INTRODUCTION

This paper presents two approaches to compare the different scales used in requirements prioritisation. Most prioritisation techniques provide the result on either the ordinal or the ratio scale. Techniques providing the result on a ratio scale include information about the relative distance between requirements and are often more time-consuming and complex to use than techniques based on an ordinal scale [10, 14]. Therefore it is interesting to investigate whether or not the added information is valuable to the decision-maker. Early results of this study were presented in [11].

ISESE'06, September 21–22, 2006, Rio de Janeiro, Brazil.

Requirements prioritisation is an important area in requirements engineering and software development [14, 20, 22]. If an organisation fails to determine the most important requirements, it risks that the developed system does not meet customers' needs and expectations [7]. During the prioritisation activity, requirements are examined and decisions are made regarding implementation. Prioritisation is closely related to release planning, as the most important, yet cost-effective, requirements found during prioritisation are selected for the earliest release. The cost-value approach [8] takes both implementation costs and customer value into account. As software development often have limited resources, it is essential to choose the requirements that give the best return on investment, in terms of customer satisfaction [7].

There are several different techniques for requirements prioritisation. Some techniques result in priorities on an ordinal scale, and provide the ranked order among requirements, e.g. the Numeral assignment [6] and the Planning game [1]. Other techniques provide the result on a ratio scale, and state how much more important one requirement is than another. Examples of these techniques are the Pair-wise comparisons [18], Wiegers' method [20], and the $100 test [13].

Scales that contain more information than others are called richer [3]. Hence, the ratio scale is richer than the ordinal scale as it provides the relative distance between ordered requirements in addition to the ranks. While a higher level of information may be beneficial, it often results in a more complex and time-consuming technique for gathering the information. In a recent experiment conducted with 16 PhD students as subjects, Pair-wise comparisons tended to be at least twice as time-consuming as the Planning game, when prioritising 16 requirements [10]. Another investigation, performed in industry, suggests that Pair-wise comparisons are complex to perform, because users found it difficult to estimate how much more valuable one requirement is than another [14]. Because of the time-consumption and complexity for the user, it would be easier and more efficient to use prioritisation techniques with less rich scales. However, it might not be sufficient as a basis for decision-making.

This paper presents an empirical investigation, which was performed by analysing ratio scale prioritisation results gained from four different experiment assignments. Most of the 36 subjects were students or PhD students taking a requirements engineering (RE) course or a research methodology course. The subjects used Pair-wise comparisons to prioritise requirements, so the resulting priorities are presented on a ratio scale. In two of the cases, the participants were supported by a Requirements Management (RM) tool [21] during prioritisation. The results were analysed after the experiment assignments.

We would like to determine which situations that require the ratio scale and in which situations the ordinal scale is sufficient. Therefore we need a measure that can describe the characteristics of the ratio scale distribution so that results from different prioritisation sessions can be compared. The measure is called *skewness* and is the standard deviation for the difference between a ratio scale distribution and a baseline distribution. In addition we want to investigate how the cost-value approach can be used if prioritisation is performed on an ordinal scale. Therefore we attempt to answer the research questions below.

RQ 1. How can we measure the skewness of a ratio scale prioritisation distribution?

RQ 2. How can the cost-value approach be applied when the priorities are based on ranks?

The paper is outlined as follows. Section 2 describes the different scale types and Section 3 presents some common requirements prioritisation techniques. Section 4 explains the empirical investigation and the data analysis. Section 5 discusses the paper and the validity of the results. Section 6 concludes the paper and provides some ideas for further work.

2. THEORY OF SCALES

In the 50s S.S. Stevens proposed properties of measurement systems and described four different scale types: Nominal, Ordinal, Interval and Ratio, each of which possesses different properties of measurement systems [3]. The scale types are presented in order of richness, i.e. the second one is said to be richer than the first one as all relations in the second one are contained in the first [3]. This section describes these four different scale types in more detail.

2.1 Nominal Scale

The nominal scale is the most primitive of the four scale types and includes some kind of categorisation or classification. All objects are grouped into subgroups and each subgroup is assigned a certain name or number. No object is allowed to belong to more than one subgroup and there is no ordering among the classes and no notion of magnitude associated with the numbers or symbols [3]. Requirements grouped according to which sub systems they concern is an example of nominal classification.

The only statistics to be gathered on this scale is frequency, i.e. the number of objects in each group. The mode can be calculated, but not the median or mean.

2.2 Ordinal Scale

The ordinal scale can be used to enhance the nominal scale with information about the ordering of classes or categories. This is the case in Numeral assignment, when each requirement is classified according to its value and assigned to e.g. the mandatory, desirable, or inessential [6] group. Priorities can also be measured using numbers such as 1, 2, 3, where the requirements with highest priority are assigned a 1. In addition, requirements within the groups can be ranked so that an ordered list of requirements is received. This scheme is used in the Planning game.

The numbers associated with the requirements represent ranking only, so arithmetic operations, such as addition and multiplication, have no meaning [3]. Statistics to be used on ordinal scales are calculation of the median and non-parametric statistics.

2.3 Interval Scale

This scale type carries information about the size of the intervals between the ordered classes, so that we can in some sense understand the jump from one class to another. An interval scale preserves order, as with an ordinal scale, and differences – but not ratios. The interval scale does not have any apparent application in requirements engineering.

2.4 Ratio Scale

The richest of the four scale types is the ratio scale, as it possesses ordering, size of intervals and ratios between entities. There is a zero element, representing a total lack of the attribute and measurement start at zero. This scale type is used in e.g. the Pair-wise comparisons. The ratio scale provides not only ordering of requirements, but also the relative distance between ordered requirements, and states how much more important one requirement is than another. All arithmetic can be applied to classes on this scale. Both parametric and non-parametric statistics can be performed on ratio scale data, and the mean can be calculated.

3. REQUIREMENTS PRIORITISATION

This section describes different requirements prioritisation techniques, which use either the ordinal or the ratio scale. There exist several prioritisation techniques, such as Numeral assignment, Planning game, $100 test, Wiegers' method, and Pair-wise comparisons. The section ends with a description of how the prioritisation result can be illustrated in a cost-value diagram.

3.1 Numeral Assignment

The Numeral assignment technique is based on the principle that each requirement is assigned a symbol representing the requirement's perceived importance. Several variants based on the Numeral assignment technique exist, e.g. classifying requirements as mandatory, desirable or inessential [6]. Another way to classify requirements is to divide them into essential, conditional or optional requirements, as suggested by the IEEE [5]. Furthermore, it would be possible to give each requirement a number e.g. between 1 and 5, where requirements with a 5 are the most important ones [6]. Classifying requirements according to Numeral assignment does not give us information about the relation between the requirements in each class, thus several requirements may appear equally valuable.

3.2 Planning Game

The Planning game is used when planning and deciding what to develop in an Extreme Programming project. When the requirements have been elicited and documented on so called Story cards, they are divided into three different piles: (1) those without which the system will not function, (2) those that are less essential but provide significant business value, and (3) those that would be nice to have [1]. Requirements are also time-estimated. Based on the time estimates, or by choosing the cards and then calculating the release date, the customers prioritise the requirements within the piles and then decide which requirements that should be planned for the next release [15]. Thus, the technique uses a sorting algorithm, similar to Numeral assignment to assign the requirements to one of three piles. Then, the requirements within each pile are compared to each other and ranked in order to achieve a sorted list.

The result of the easy and straightforward Planning game technique is an ordered list of requirements. This means that the requirements are represented on an ordinal scale, without giving any information about how much more important one requirement is than another. Decision-makers need to answer the question "how important is this requirement?" and determine the absolute importance of all requirements. In a recent experiment, the Planning game appeared superior to the Pair-wise comparisons, as it was less time-consuming and most subjects found it easier to use [10].

3.3 $100 Test

In the $100 test, each participant is given $100 in fictional money to distribute between requirements. Each participant is asked to write down on a sheet of paper how much of this money is to be spent on each requirement. Then a facilitator tabulates the results and provides an ordered ranking of requirements [13]. The total amount of money spent on each requirement provides us with a relative difference between the different requirements, i.e. the results are obtained on a ratio scale.

3.4 Wiegers' Method

According to Wiegers' method, the priority of a requirement can be calculated by dividing the value of a requirement with the sum of costs and technical risks associated with implementing it [20]. Typical participants are the project manager, key customer representatives and development representatives. The resulting priorities are on a ratio scale.

3.5 Pair-Wise Comparisons

Pair-wise comparisons involve comparing all possible pairs of requirements in order to determine which of the two requirements is of higher priority, and to what extent. If there are *n* requirements to prioritise, the total number of comparisons to perform is $n(n-1)/2$. For each requirement pair the decision-maker estimates the relation between the requirements on the scale {9, 7, 5, 3, 1}, where 1 represents equal importance and 9 represents one requirements being much more important than the other. This relation results in a dramatically increasing number of comparisons as the number of requirements increases. While Pair-wise comparisons are time-consuming to perform for large sets of requirements, the technique provides a structured way to compose a ratio scale list of requirements, which is difficult to do by hand.

In the investigation by Karlsson et al. [9], the authors concluded that Pair-wise comparisons (there called AHP) was the most promising approach because it is based on a ratio scale, is fault tolerant, and includes a consistency check. Pair-wise comparisons was the only technique in the evaluation that satisfied all these criteria. While it includes a priority distance, i.e. a ratio scale, the other approaches provide only the priority order. However, because of the rigour of the technique, it was also the most time-consuming in the investigation.

In another empirical investigation of prioritisation techniques performed by Lehtola & Kauppinen [14], Pair-wise comparisons was compared to Wiegers' method [20]. The authors conclude that users found it difficult to estimate how much more valuable one requirement is than another and that some users considered pair-wise comparisons as pointless as they felt it would have been easier for them to just select the most important requirements [14].

In a recent experiment, the Pair-wise comparisons were compared to the Planning game regarding time consumption, ease of use and accuracy. It was concluded that the Pair-wise comparisons were at least twice as time-consuming as the Planning game, when prioritising 16 requirements [10].

A commercial RM tool has been developed to support the Pair-wise comparison technique [21]. The number of comparisons is reduced by a certain algorithm and priorities can be presented in different charts and diagrams. There exists several different ways to reduce the number of comparisons [2, 4]. However, Pair-wise comparisons still tends to be more time consuming than ordinal scale techniques when prioritising a large number of requirements.

3.6 Cost-Value Diagrams

The cost-value diagram is used to determine which requirements have a high value-to-cost ratio and which do not [8]. When using the ratio scale there are different ways to determine from the cost-value diagram the requirements with high contribution to the product, i.e. with a high value-to-cost ratio. Similarly, it is possible to find the requirements with a low contribution to the product, i.e. with a low value-to-cost ratio. The cost-value diagram is often divided into three separate areas, marked A, B, and C, below. The two main options to determine these areas are shown in Figure 1. Option (a) is to draw lines so that one third of the requirements end up in each area (A, B and C) and option (b) is to draw lines so that requirements with a value-to-cost ratio higher than 2 end up in area A, and the ones with a value-to-cost ratio lower than 0.5 end up in area C. Then area B will include the requirements in between. Option (a) is utilised in the commercial tool [21] used in the experiment assignments described below, and option (b) is presented in e.g. [8]. In both cases, the lines start in the origin of the diagram and are drawn diagonally through the diagram. Requirements in area A are high contributors and should be implemented as soon as possible, as they are valuable but not expensive to implement. Requirements in area C are low contributors and too expensive to implement regarding their low customer value. Requirements in area B are medium contributors and have to be analysed further.

When using the ordinal scale, the described procedures are not natural. Lines drawn from the origin of the diagram (Figure 1a) is not applicable, as the zero does not have any meaning in the ordinal scale, i.e. no requirement can be ranked as zero. As stated before, it is not valid to use arithmetic such as division and multiplication, and therefore the value-to-cost ratio is not feasible when using ranks (Figure 1b).

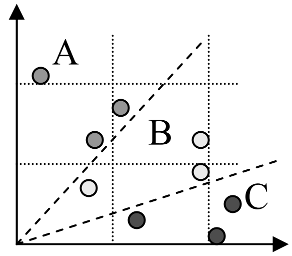

Figure 1a. Option (a): One third of the requirements in each area

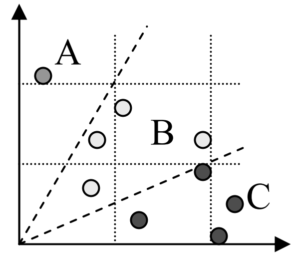

Figure 1b. Option (b): Lines are drawn regarding the value-to-cost ratio

A more feasible option for the ordinal scale would be to divide the graph into a number of squares by drawing vertical and horizontal lines through the graph. Since common techniques such as the Planning game and Numeral assignment involve dividing the requirements into three groups for each criterion, this could also be applied in the cost-value diagram. It would result in nine equally large squares based on the ranks, as shown in Figure 2. We suggest that the requirements in areas denoted A in Figure 2, are the high contributors that should be implemented first due to their high value and low cost. The requirements in areas denoted C are low contributors and should be implemented last or perhaps not at all. Requirements in areas B are medium contributors and need further investigation.

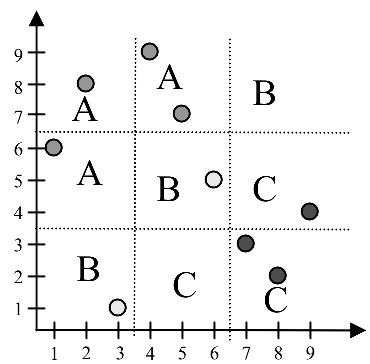

Figure 2. Ordinal scale cost-value diagram

4. METHODOLOGY

This section presents the method used in our empirical investigation, and the analysis of the data. In order to investigate the skewness of the ratio scale distribution (RQ1) we need authentic ratio scale data results from real prioritisation sessions. These data can be compared to a baseline in order to get a measure of the skewness for each subject. Similarly, for the cost-value approach (RQ2), we need to compare ratio scale cost-value diagrams with ordinal scale cost-value diagrams in order to see if the requirements selection differs depending on the scale. It was decided to use the ratio scale data and reduce it to ordinal information to draw the ordinal cost-value diagrams for each subject. Thereby, we can compare ratio and ordinal scale diagrams based on the same data set. These data were obtained from experiments and assignments conducted at the university.

4.1 Data Collection

The research methodology is based on *archive analysis* as the data were produced for purposes other than this research [17]. Four different data sets, containing data from 36 subjects, were investigated, see Table 1. Two of the data sets were obtained in experiment assignments conducted within a Masters' course. The other two sets were obtained from a previous experiment conducted by the authors, in which requirements prioritisation techniques were examined [10].

Table 1. Outline of the four data sets

Data set	Setting	Number of subjects	Number of requirements	Cooperation /Individual
1	Experiment assignment	8	21	Cooperation
2	Experiment assignment	12	21	Cooperation
3	RE experiment	8	8	Individual
4	RE experiment	8	16	Individual

4.1.1 Data set 1: First experiment assignment

The first data set was extracted from an experiment assignment in an optional RE course for Master's students. The purpose of the experiment assignment was to teach the 25 students about the challenges of prioritising requirements and to allow the students to investigate a commercial RM tool [21]. During the assignment, 2-3 students cooperated on one computer, as it is a realistic situation for decision-makers. In total, ten groups of students participated; thus ten priority results were obtained, hereafter called subjects. The subjects were asked to prioritise between 21 mobile phone requirements, such as Chat function, Wireless Application Protocol (WAP), Predictive text input (T9), etc. The tool utilises Pair-wise comparisons as prioritisation technique and has an algorithm that reduces the number of necessary comparisons, without jeopardising the consistency. The resulting priorities are presented as percentages on a ratio scale. The subjects were encouraged to choose two different criteria, one to maximise and one to minimise. Most subjects selected one criterion related to value, and one criterion related to cost. In the analysis after the experiment assignment, two of the ten subjects were removed, as their criteria were not consistent with either value or cost. The ratio scale data from the resulting eight subjects were then examined as described in Section 4.2.

4.1.2 Data set 2: Second experiment assignment

The second data set was obtained in a similar manner as the first one. The same RE course was given a second semester to 26 students. Two of the students worked alone and the rest worked in pairs, resulting in 14 priority results, hereafter called subjects. The same tool was used, as well as the same requirements in the prioritisation task. The subjects were encouraged to select the criteria cost and value but as two of the subjects choose other criteria they were removed from the analysis. As we anticipated the results to be used for research purposes, we also posed some qualitative questions after the session. These qualitative results are presented in Section 4.2.3.

4.1.3 Data set 3 and 4: RE experiment

The third and fourth data sets are based on an experiment conducted to compare two different requirements prioritisation techniques: Planning game and Pair-wise comparisons. The experiment tested the difference between the techniques regarding time consumption, ease of use, and accuracy, see [10]. In this case, the Pair-wise comparisons were performed manually, i.e. no tool was allowed. The 16 subjects worked individually, as the task was performed for experimental purposes. The experiment yielded two different sets of data as half of the subjects prioritised between eight requirements and half of the subjects prioritised between 16 requirements. The requirements were similar to the ones in the RE course, i.e. mobile phone requirements.

4.2 Data Analysis

The priority results from the experiment assignments and the prioritisation experiment were investigated in two different ways. First, the results were analysed to investigate the skewness of the ratio scale priorities. Secondly, the ratio scale data were reduced to ranks, i.e. ordinal scale data, and investigated in cost-value diagrams. Finally, some qualitative results from the second experiment assignment are presented.

4.2.1 Evaluation of skewness

The first research question concerns possibilities of measuring the skewness included in the ratio scale distribution. In this section we define a skewness measure based on the standard deviation for the difference between a ratio scale distribution and a baseline distribution. A more skewed distribution indicates that the person performing the prioritisation has given much larger weights to some of the requirements than others. A less skewed distribution indicates that the differences in priorities between requirements are not very large. In that case it could have been sufficient to use the ordinal scale since the distribution could be approximated with a linear distribution. Therefore, we use the linear distribution as a baseline. The measure of skewness can be used to determine in which situations the ratio scale is needed and when the ordinal scale is sufficient.

A bar chart illustrates the result of a prioritisation through bars of different length. The length of each bar represents the ratio for each requirement, thus the total value of all requirements adds up to 100%. Unlike the cost-value diagram, the bar chart shows the result for one criterion at a time.

The left bar chart in Figure 3 illustrates the bar chart from a ratio scale prioritisation for one of the subjects. The distribution is clearly skewed and the top requirement accounts for approximately 30% of the total value. If we want to evaluate how skewed this distribution is, we can compare it to a baseline bar chart where the difference between adjacent bars is equal all over the chart. This is shown in the right bar chart in Figure 3. Note that the requirements have the same ranks in both charts, but the right one is transformed to the baseline distribution. Imagine if the right one were the result from using a ratio scale technique. Then it would have been more efficient to use an ordinal scale technique with ranking, as the difference in importance between requirements is the same all over the chart. Therefore we use the right bar chart in Figure 3 as the baseline when evaluating the skewness of the ratio scale bar chart.

Procedure. To evaluate the skewness of a ratio scale prioritisation distribution we want a measure of the characteristics of the ratio scale distribution. This measure is obtained by comparing the ratio scale bar chart to the linear one, as shown in Figure 3, in the following manner. Suppose the value of the lowest ranked requirement in the linear bar chart account for k% of the total value. The second lowest ranked requirement would then account for 2*k% of the value. In a similar manner, the highest ranked requirement account for N*k% of the value, if there are N requirements in total.

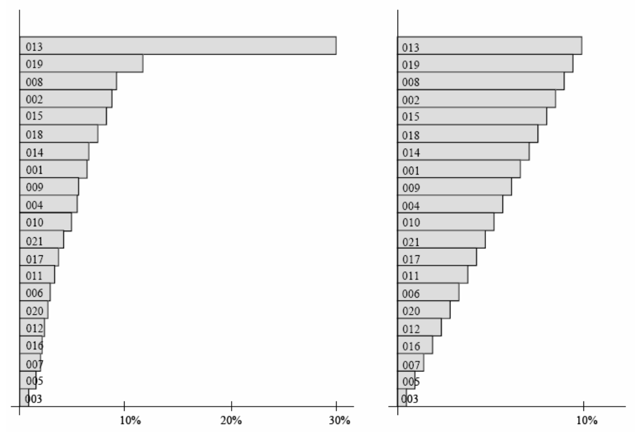

Figure 3. Comparison between skewed and linear bar charts

Thus, multiplying the constant k with the rank yields the linear equivalent to the priority. The difference between adjacent requirements is the same all over the chart, and equal to the constant k%. In order for the total value of all requirements to add up to 100%, the statement below can be used.

$$k + 2k + 3k + ... + Nk = 1$$

Using this equation, it is possible to calculate the constant k:

Eq 1. $$k = \frac{1}{1 + 2 + 3 + ... + N}$$

From these assumptions, we can calculate the skewness of the ratio scale data by calculating how much the ratio bar deviates from the constant k multiplied by its rank. This difference was calculated for each requirement and then the standard deviation, i.e. the skewness, was calculated as follows:

Eq 2. $$Skewness = \sqrt{\frac{\sum (diff\_each\_req)^2}{N - 1}}$$

Since the values of the ratio scale prioritisation often are presented in percentages, we have chosen to use percentages in the calculations.

Example. For the example in Figure 3, the highest ranked requirement accounts for 28.7% of the value according to the ratio scale. To calculate the value of the highest ranked requirement in the linear distribution, we need to calculate the constant k with Equation 1, where N=21.

$$k = \frac{1}{1 + 2 + 3 + ... + 21} = 0.00433$$

Thus, the highest ranked requirement accounts for N*k%, i.e. 21*0.00433=9.1%. Consequently, the difference between the ratio scale distribution and linear distribution is 28.7%–9.1%=19.6% for this requirement. This difference is summed up for all the requirements and used in Equation 2 to calculate the skewness. In this example the skewness is 4.6%.

Result. The skewness was calculated for all subjects with Equation 2 and is presented in the tables below.

Table 2. Skewness for dataset 1

	A1	A2	A3	A4	A5	A6	A7	A8	*Av*
Value (%)	4.6	1.3	1.5	1.0	2.0	4.3	1.3	2.7	*2.3*
Cost (%)	4.0	3.9	2.5	0.5	2.3	4.8	1.1	1.2	*2.5*

As seen in Table 2, Subject A4 has the smallest skewness for both criteria and thus the least skewed distribution.

Table 3. Skewness for data set 2

	B1	B2	B3	B4	B5	B6	
Value (%)	1.1	1.9	1.9	1.9	1.5	1.0	
Cost (%)	1.0	4.3	1.2	1.7	0.9	2.4	
	B7	**B8**	**B9**	**B10**	**B11**	**B12**	***Av***
Value (%)	1.3	1.1	1.2	2.9	1.0	2.8	*2.0*
Cost (%)	1.6	2.1	1.2	1.1	1.0	1.5	*1.6*

As Table 3 indicates, subject B11 has a small skewness for both criteria, while e.g. subject B2 has a rather large skewness. The average skewness is smaller for the subjects in the second experiment assignment than in the first for both criteria.

Table 4. Skewness for dataset 3

	C1	C2	C3	C4	C5	C6	C7	C8	*Av*
Value (%)	3.5	2.3	3.6	2.7	6.2	3.2	3.1	2.2	*3.4*
Cost (%)	2.0	5.5	2.3	2.4	5.8	1.3	3.7	2.0	*3.1*

Among these subjects, C5 has the most skewed distribution.

Table 5. Skewness for dataset 4

	D1	D2	D3	D4	D5	D6	D7	D8	*Av*
Value (%)	1.7	0.9	1.5	0.6	1.9	1.0	3.0	2.1	*1.6*
Cost (%)	1.2	3.5	1.5	0.3	0.9	1.5	1.0	1.9	*1.5*

Subject D4 has the lowest skewness among the subjects in Table 5.

Analysis. The average skewness is rather equal for both criteria within each of the four data sets. Thus, the skewness seems to be independent of the criteria, at least for these criteria and these data sets. The size of the average skewness varies between the data sets. This can be explained by the different numbers of requirements in the different settings, and the difference in application, i.e. manual as opposed to tool-supported techniques.

The skewness seems to vary among the subjects. However, most subjects have rather similar skewness values for both used criteria. In fact, the correlation between the skewness for value and the skewness for cost is 0.37, p=0.028, see also Figure 4. This could imply that some people tend to use more extreme values during prioritisation regardless of criterion. Others tend to be more modest and use the smaller values no matter which criterion. These subjects could perhaps have been satisfied with an ordinal scale technique.

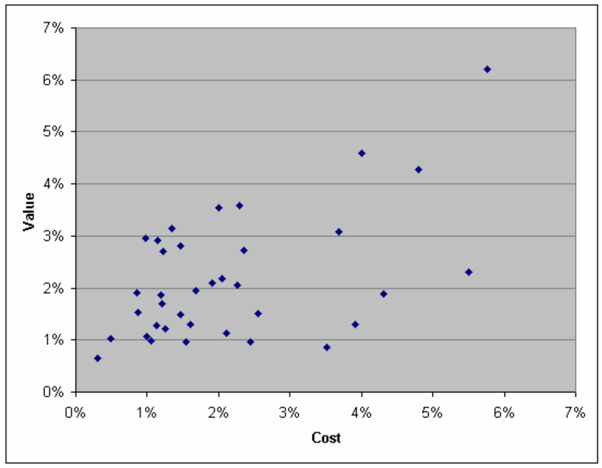

Figure 4. Correlation between skewness for value criteria and cost criteria

This section has shown a possible way to calculate the skewness of the ratio distribution. This was used to investigate if certain subjects use the ratio scale potential more than others, which is confirmed by a slight correlation in our empirical data. It was also investigated if certain criteria tend to get a more skewed distribution. However, the average skewness is similar for both criteria within each data set.

4.2.2 Evaluation of ordinal scale cost-value diagram

The second research question regards the possibility to apply the cost-value approach when priorities are based on ranks. Usually, cost-value diagrams are used to visualise ratio scale priorities by percentages, see e.g. [8]. It may, however, be possible to draw similar diagrams with ordinal scale data, based on the ranks instead of the ratios, as described in Section 3.6. The cost-value diagram is used as decision-support when selecting the most appropriate set of requirements for a release. We want to investigate whether the same requirements would be selected for implementation when using ordinal data as when using ratio data in the cost-value diagram.

Procedure. Two sets of cost-value diagrams were drawn based on our empirical data, one with ratio scale data and one with ranks. The ratio scale diagram corresponds to the example in Figure 1a, i.e. divided in three equally large areas, and the ordinal scale diagram corresponds to the example in Figure 2. The ratio scale diagram was used as a baseline in the investigation. We wanted to see whether the ordinal scale diagram would point out the same requirements as the ratio scale diagram did. Therefore, we marked each requirement in the ordinal scale cost-value diagram with different symbols depending on where in the ratio diagram it appeared, see Figure 5. Each requirement with high contribution in the ratio scale diagram was marked with a circle; each requirement with medium contribution in the ratio scale diagram was marked with a square; and each requirement with low contribution in the ratio scale diagram was marked with a triangle. Thereby it was possible to compare the cost-value diagram based on ranks, with the original ratio cost-value diagram.

The ordinal scale cost-value diagram was divided into nine equally large areas, similar to the example in Figure 2. We assume that the three areas to the upper left, called A in Figure 2, should ideally contain high contributors according to the ratio scale and should be marked with a circle, if the ordinal scale was reflecting the ratio scale perfectly. Similarly, the three areas to the lower right, called C in Figure 2, should ideally contain low contributors according to the ratio scale and should be marked with a triangle. The middle areas, called B in Figure 2, should contain medium contributors, and be marked with a square.

Next we can calculate the level of agreement between the ordinal scale cost-value diagram and the original ratio scale diagram. The *Kappa value* (K) can be used to assess the agreement between a set of raters who have assigned a set of objects to one of several categories [19]. In this case, the objects are the requirements and the categories are the three areas A, B and C. The two "raters" are the categorisation based on ranks and the categorisation based on ratios.

The Kappa value is calculated according to the following:

Eq 3. $$K = \frac{P(A) - P(E)}{1 - P(E)}$$

where P(A) denotes the proportion of the times the raters agree and P(E) denotes the expected agreement that would be present by chance if all ratings were made randomly [19]. Kappa values close to 1 represent very high agreement, while Kappa values close to, or below, 0 represent no agreement. The strength of agreement is classified by e.g. Landis and Koch [12] as described in Table 6.

Table 6. Landis and Koch Kappa statistics [12]

Kappa statistic	Strength of agreement
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

Example. The cost-value diagram in Figure 5 is used as an example of how the requirements are distributed over the diagram when the ordinal scale is used. From this diagram it is possible to count the number of "correct" and "incorrect" requirements in each area.

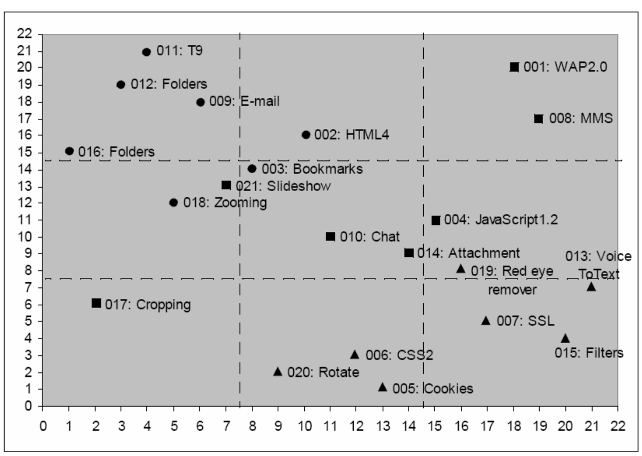

Figure 5. Example of cost-value diagram based on ranks

For the example in Figure 5, it is visible that most requirements in area B are marked with a square, but some requirements have ended up in the wrong area. In this case, prioritising with an ordinal scale would miss one high priority requirement, and would accidentally discard one medium priority requirement. However, most requirements actually end up in the correct area.

Results. The tables below present the number of high (H), medium (M), and low (L) requirements in each of the three areas A, B, and C for each subject. The ideal categorisation, corresponding to the one in which the ranks reflect the ratios, would be that all requirements are in the correct area, i.e. area A contains high contributors, area B medium contributors and area C low contributors. The last row in the tables below presents the Kappa value, i.e. the measure of the agreement between the priorities based on ranks and the priorities based on ratio scale.

Table 7. Requirements in different areas of the cost-value diagram for data set 1

	A1	A2	A3	A4	A5	A6	A7	A8	*Av*
Area A	3H	6H 2M	6H 1M	6H	7H 2M	5H 2M	7H 1M	6H 3M	
Area B	4H 7M 3L	1H 4M 1L	1H 5M	1H 6M 1L	2M 1L	2H 3M 2L	3M 1L	1H 2M 2L	
Area C	4L	1M 6L	1M 7L	1M 6L	3M 6L	2M 5L	3M 6L	2M 5L	
K	0.50	0.64	0.79	0.79	0.57	0.43	0.64	0.43	*0.60*

As can be seen in Table 7, subject A3 and A4 have the same Kappa value, as they have the same number of correctly placed requirements. Subject A6 and A8 have lower agreement because fewer requirements are correctly placed. The Kappa values for data set 1 represent moderate to substantial agreement.

Table 8. Requirements in different areas of the cost-value diagram for data set 2

	B1	B2	B3	B4	B5	B6	
Area A	7H 1M	4H 1M	6H	5H 1M	4H	5H 1M	
Area B	6M	3H 5M 3L	1H 6M	2H 6M 2L	3H 6M 5L	2H 4M 3L	
Area C	7L	1M 4L	1M 7L	5L	1M 2L	2M 4L	
K	0.93	0.43	0.86	0.64	0.36	0.43	
	B7	**B8**	**B9**	**B10**	**B11**	**B12**	*Av*
Area A	6H 1M	6H 3M	6H	5H	3H 1M	7H 2M	
Area B	1H 5M 2L	1H 4M 1L	1H 6M	2H 7M 1L	4H 6M 2L	3M	
Area C	1M 5L	6L	1M 7L	6L	5L	2M 7L	
K	0.64	0.64	0.86	0.79	0.50	0.71	*0.65*

As Table 8 indicates, the Kappa values vary for data set 2 between fair agreement for subject B5 and almost perfect agreement for subject B1.

Table 9. Requirements in different areas of the cost-value diagram for data set 3

	C1	C2	C3	C4	C5	C6	C7	C8	*Av*
Area A	2H	3H	3H	2H	3H	1H	3H	3H	
Area B	1H 2L	1M	2M	1H 2M 1L	2M	2H 2M 2L	2M	1M	
Area C	2M 1L	1M 3L	3L	2L	3L	1L	3L	1M 3L	
K	0.07	0.80	1	0.64	1	0.30	1	0.80	*0.72*

For data set 3, the Kappa values vary a lot between the subjects. Subject C1 has only slight agreement, subject C6 has fair agreement and subjects C3, C5, and C7 have perfect agreement between the ordinal and ratio scale diagrams.

Table 10. Requirements in different areas of the cost-value diagram for data set 4

	D1	D2	D3	D4	D5	D6	D7	D8	*Av*
Area A	5H 1M	5H 2M	5H 2M	5H	5H 1M	5H 1M	4H 2M	4H	
Area B	2M 1L	2M 1L	3M	5M	3M	3M 1L	1H 4M	1H 5M 1L	
Area C	3M 4L	2M 4L	1M 5L	1M 5L	2M 5L	2M 4L	5L	1M 4L	
K	0.54	0.54	0.72	0.91	0.72	0.63	0.72	0.72	*0.69*

For data set 4, the Kappa value varies between 0.54 and 0.91, which corresponds to moderate, substantial or almost perfect agreement.

Analysis. It seems as the areas A, B, and C are valid to use to distinguish the high, medium and low priority requirements, at least with the data sets in this investigation. The Kappa values correspond to, at least, fair agreement in all cases, except when only eight requirements were prioritised in data set 3. The small number of requirements makes each requirement affect the Kappa value more than in cases with a larger number of requirements. The Kappa values for data set 3 vary more than for the other data sets.

The average Kappa value for all data sets are 0.67, which reflects substantial agreement.

The conclusion drawn from this section is that it is possible to achieve similar decision-support with ordinal scale prioritisation techniques as with ratio scale prioritisation techniques. For most subjects in the investigation, only a few requirements are judged differently when basing the decision on the ordinal scale. Therefore, it would be possible to use the ordinal scale cost-value diagram as presented here when selecting requirements for a release. This is valuable as most ordinal prioritisation techniques are easier and faster to use than techniques with ratio scale results.

4.2.3 Opinions on scales

After the second experiment assignment, the subjects answered some questions regarding how they felt about the prioritisation technique and the scale. Approximately half of the subjects found it easy to decide which degree on the ratio scale to select, and half found it difficult. Motivations for finding it easy included that the scale was intuitive and the subjects felt sure about the domain. Subjects finding it difficult motivated it with lack of domain knowledge and some said that it was easy to choose which requirement that was more valuable, or expensive, but to decide how much was more difficult. Five of the subjects also stated that it would have been easier and faster to omit the ratio scale and only decide "more than" or "less than". This would speak in favour of the ordinal scale. However, as some subjects pointed out, it would be more fault intense and yield less information.

When asking the subjects whether they used the extreme values on the prioritisation scale, almost half of the subjects said that they did not. Their motivations include insecurity due to lack of domain knowledge, and that "you never know if another requirement is even better". It is possible that when decision-makers are insecure about the domain, it may be sufficient to use the ordinal scale, since the extreme values of the ratio scale are not used.

5. DISCUSSION

The main validity issue is the generalisability. As we have only investigated students and PhD students prioritising rather independent mobile phone requirements it is difficult to generalise to industrial cases. The skewness measure and the ordinal cost-value diagram need industrial validation in order to rely on the results with more certainty. The issue of participant and observer bias is reduced in this study as it is an archive analysis and neither the students nor the researchers planned to use the results for this purpose and could hence not affect them.

The data show a slight correlation ($r=0.37$, $p=0.028$) between the skewness for value and the skewness for cost. We need more data in order to determine whether the skewness for different criteria correlates. Similarly, it requires more investigation of the cost-value diagram in order to determine if it is valuable for decision-makers to use the ordinal scale. For most of our empirical data there appears to be a substantial agreement between the ratio scale cost-value diagram and the ordinal scale cost-value diagram. However, it requires use by decision-makers to see whether the ordinal scale diagram is sufficiently accurate. It shall also be noted that it is not necessarily the case that using an ordinal scale requirements prioritisation technique would give the same cost-value diagram as the one obtained by reducing the ratio scale data to ordinal scale data, which was done in this study. Further studies need to be made in order to determine if this transformation of data is an appropriate approximation of the results from ordinal scale prioritisation. An alternative approach to obtain comparable data on different scales would be to use data obtained from different techniques, one based on ordinal scale and one based on ratio scale. However, the difference in prioritisation techniques would then influence the result more than the difference in scales, which is the issue we try to investigate here. A comparison between a ratio scale techniques and an ordinal scale technique is presented in [10].

The skewness may be used to compare different sets of data in order to determine which one that benefits more from the ratio scale. In our case, we compared the skewness between different subjects and between the criteria cost and value. There are other criteria as well that could be investigated similarly, such as risk, effort, business opportunity, etc. It would also be possible to compare prioritisation data from different domains, or different types of requirements, to see if certain domains benefit more from the ratio scale than others. The skewness is rather easy to calculate using a spreadsheet, and it determines the skewness for each criterion separately. Instead of the linear baseline used for calculating the skewness, it would be possible to create a baseline based on the average distribution for the investigated subjects. In that manner, the skewness would then vary positively or negatively from the baseline distribution.

The cost-value diagram based on ranks can be used when the prioritisation technique results in ranks, which is the case when using e.g. the Planning game or Numeral assignment. For our investigation we chose to divide the diagram into nine equally large areas to simulate the high, medium, and low priority groups often resulting from e.g. Numeral assignment. However, in a real case, these groups might not be equally large and thus the nine areas might be of different sizes. Alternative ways to divide the cost-value diagram could result in other conclusions but this has not been investigated in this paper.

6. CONCLUSIONS

This paper has presented two ways to analyse and compare ordinal and ratio scale data: (1) Evaluation of skewness, and (2) Evaluation of ordinal scale cost-value diagram. It is important to be able to compare data from different scales since there is a trade-off between using a richer, but more time-consuming, scale and a less rich, but faster, scale when prioritising requirements.

We have presented a skewness measure that can be used to characterise the ratio scale data, see RQ 1. The skewness can be used to compare the characteristics of the ratio scale for different subjects, criteria or types of requirements. Based on our empirical data we conclude that some subjects are more inclined to use the extreme values of the ratio scale than others. However, the skewness of the ratio scale does not seem to be affected by the different criteria, at least not cost and value. The practical implication of the presented results suggest that decision-makers can use the skewness measure in a pre-test before prioritisation to determine whether a ratio scale technique is worth the effort (due to high skewness) or if it is sufficient with an ordinal scale technique (due to low skewness). If an ordinal scale prioritisation technique is used, it is valuable for decision-makers to be able to use the cost-value approach. Therefore, we have suggested how the cost-value diagram can be used to find the most valuable requirements, when priorities are on an ordinal scale, see RQ 2. The empirical ratio scale data and the ranks of the same data were compared using cost-value diagrams. The decision regarding which requirements to select in the data sets used in this study would be rather similar basing the diagram on ranks instead of ratios.

This paper has attempted to answer questions about how the difference between the scales can be calculated and illustrated. However, we cannot answer the even more interesting question about when the ratio information is worth the extra effort and when ranking of requirements may be enough. Below we have listed a number of ways to investigate this further:

- Conduct a case study where both ordinal and ratio scale prioritisation techniques are applied and evaluated by practitioners.
- Set up a controlled experiment where several different criteria are used, for example value, cost, effort, risk etc., to investigate if some criteria need the ratio scale more than others. Similarly, different types of requirements can undergo the same investigation.
- Perform an interview survey where several experienced practitioners answer questions regarding when the ratio scale is necessary and which characteristics that can be used to decide if the ratio scale should be used or not.
- Simulate priorities from different known distributions to compare with data from real ratio scale prioritisations to see which distribution that is most similar to the real ones – and to see if there are some distributions that can benefit more from the ratio scale than others.

7. ACKNOWLEDGEMENTS

This work is partly funded by VINNOVA in the ITEA project MERLIN. The authors would like to thank all participants in experiments and assignments and Dr. Vincenzo Gervasi for valuable comments on an earlier version of the paper.

8. REFERENCES

[1] Beck, K., *Extreme Programming Explained*, Addison-Wesley, 1999.

[2] Carmone, F.J., Kara, A., and Zanakis, S.H., A Monte Carlo Investigation of Incomplete Pairwise Comparison Matrices in AHP, *European Journal of Operational Research,* Vol 102, pp. 538-553, 1997.

[3] Fenton, N.E., and Pfleeger, S.L., *Software Metrics - A Rigorous and Practical Approach*, PWS Publishing Company, 1997.

[4] Harker, P.T., Incomplete Pairwise Comparisons in the Analytic Hierarchy Process, *Mathl. Modelling*, Vol 9, pp. 837-848, 1987.

[5] IEEE Std. 830-1998, IEEE Recommended Practice for Software Requirements Specifications, 1998.

[6] Karlsson, J., Software Requirements Prioritizing, *Proceedings of the International Conference of Requirements Engineering (ICRE'96)*, pp. 110-116, 1996.

[7] Karlsson, J., and Ryan, K., Supporting the Selection of Software Requirements, *Proceedings of the 8th workshop on Software Specification and Design*, pp. 146-149, 1996.

[8] Karlsson, J., and Ryan, K., A Cost-Value Approach for Prioritizing Requirements, *IEEE Software*, pp. 67-74, 1997.

[9] Karlsson, J., Wohlin, C., and Regnell, B., An Evaluation of Methods for Prioritizing Software Requirements, *Information and Software Technology*, pp. 939-947, 1998.

[10] Karlsson, L., Berander, P., Regnell, B., and Wohlin, C., Requirements Prioritisation: An Experiment on Exhaustive Pair-Wise Comparisons versus Planning Game Partitioning, *Proceedings of the 8th Int. Conference on Empirical Assessment in Software Engineering (EASE'04)*, Edinburgh, Scotland, 2004.

[11] Karlsson, L., Regnell, B., Comparing Ordinal and Ratio Scale Data in Requirements Prioritisation, *Proceedings of the 3rd Int. Workshop on Comparative Evaluation in Requirements Engineering (CERE'05)*, Paris, France, 2005.

[12] Landis, J.R., Koch, G.G., The measurement of observer agreement for categorical data, *Biometrics*, Vol. 33, No. 1, pp. 159-174, 1977.

[13] Leffingwell, D., and Widrig, D., *Managing Software Requirements - A Unified Approach*, Addison-Wesley, 2000.

[14] Lehtola, L., and Kauppinen, M., Empirical Evaluation of Two Requirements Prioritization Methods in Product Development Projects, *Proceedings of European Software Process Improvement Conference*, Trondheim, Norway, 2004.

[15] Newkirk, J. W., and Martin, R. C., *Extreme Programming in Practice*, Addison-Wesley, 2001.

[16] Phillips-Donaldsson, D., 100 Years of Duran, *Quality Progress*, Vol. 37, pp. 25-39, 2004.

[17] Robson, C., *Real World Research*, Blackwell, 1997.

[18] Saaty, T. L., *The Analytical Hierarchy Process*, McGraw-Hill, 1980.

[19] Siegel, S., and Castellan, J.N., *Nonparametric Statistics for the Behavioral Sciences*, Second Edition, McGraw-Hill, 2000.

[20] Wiegers, K., *Software Requirements*, Microsoft Press, 1999.

[21] http://www.telelogic.com/corp/products/focalpoint/index.cfm (visited February 28th, 2006).

[22] Zave, P., Classification of Research Efforts in Requirements Engineering, *Proceedings of the 2nd Int. Symposium of Requirements Engineering*, pp. 214-216, 1995.