södertörns högskola
UNIVERSITY COLLEGE

SÖDERTÖRNS HÖGSKOLA (UNIVERSITY COLLEGE)

STIG BLOMSKOG

# A formal analysis of a conventional job evaluation system

WORKING PAPER 2007:5

REPORT
2007-09-13

**A formal analysis of a conventional job evaluation system***

By

Stig Blomskog

Södertörns högskola, University College

Box 4101 Huddinge

SE-141 89 Sweden

E-mail: Stig.Blomskog@sh.se

Tel : +46(0)8 608 40 52

Fax : +46(0)8 608 44 80

* I wish to thank professor Per-Erik Malmnäs for constructive comments. This study has been funded by the Swedish Council for Working Life and Social Research.

## Abstract

In this paper we analyze the use of numerical information in the context of job evaluation. The analysis is based on the job evaluation system *Steps to Pay Equity,* which is recommended by the European Project on Equal Pay supported by the European commission. The main findings can be summarized as follows. Firstly, in *Steps to Pay Equity* no method is suggested that can be used in order to construct stronger scales than ordinal scales. This implies that rankings of jobs are based on the addition of ordinal scales, which means that the rankings are very unstable for admissible transformations. Secondly, there is no explicit definition or explanation how the weights should be interpreted, something that hampers an assessment about the reasonability of the assigned weights. Thirdly, the convention to classify jobs on predefined levels can give rise to heavy deformations of relevant differences between jobs, which means that received rankings of jobs are unjustified guidance for impartial pay setting. We suggest a possible remedy by illustrating the use of a specific multi-attribute evaluation model.

## 1. Introduction

Job evaluations have become important instruments in order to discover wage discrimination by gender. Job evaluation means that a set of jobs is compared with respect to an overall evaluation of demand and difficulties that are associated with the jobs. In Equal Pay Acts for EU-member countries it is stated that the demands and difficulties for things such as skills, responsibility, effort and working conditions have to be considered when jobs are compared in order to reveal indications of gender biased pay structures at various workplaces. In practical applications these four main-criteria are divided into various numbers of sub-criteria or factors. In order to make evaluative comparisons of such multidimensional items tractable it is common that numerical models are applied. Usually weighted sums of scores are used as measure of the overall values of jobs. The scores represent the evaluation of the jobs with respect to each factor. This type of numerical models is popular and has a wide application in many other evaluation contexts where multi-dimensional items are to be evaluated.

The paper has two purposes. Firstly, we will analyze the way numerical information is applied in conventional job evaluation methods. It is important to use numerical information in a formally correct way. Otherwise, the result in terms of numerical information gives a biased representation of actual difference between jobs as regards demands and difficulties. Secondly, we will discuss an evaluation method that can be used in order to avoid deformation of actual differences between jobs with regards demands and difficulties that occur in application of conventional job evaluation methods. Such deformations mean that there is weak correspondence between actual and relevant differences between jobs and the result of conventional job evaluation methods. In other words ranking of jobs based on conventional job evaluation methods gives a poor guidance for a gender-neutral pay setting of jobs.

The paper is organized as follows. The next section describes a representative job evaluation system that serves as the basis for the analysis. The third section contains definitions used in the subsequent analysis. The fourth section contains an analysis of the numerical specification of the representative job evaluation model. In the fifth section we discuss the occurrence of deformations. Further, we demonstrate a remedy of deformations by applying a Multi Criteria Evaluation Model, the PRIME-model, which is constructed by Salo and Hämäläinen (2001).[1] Appendix contains a more detailed description of the PRIME-model.

## 2. A representative job evaluation system

The analysis is based on a job evaluation system, named *Steps to Pay Equity,* recommended by the European Project on Equal Pay, which is supported by European Commission (Harriman and Holm 2001). We assume that the system and its way to use numerical information is representative for many job evaluation systems the purpose of which is to reveal indications of a gender biased pay structure.

The system *Steps to Pay Equity* can briefly be described as follows. Eight criteria or factors are recommended as grounds for an evaluative comparison of a set of jobs. Each factor is divided as default into five levels, which are scored from 1 to 5. Definitions of the factors are shown in Appendix B. Each factor is assigned a weight in percent, which intends to express the importance of the factor according to the user. In the subsequent analysis we use the term decision maker (DM) to refer to users or to persons responsible for the evaluation. The job evaluation process starts with establishing job descriptions of all jobs, which serves as basic information in the job evaluation process. Each job is then classified on a level of each factor that the DM judge to best fit the job description. In a final step all factors are assigned weights in percent, which means that 100 percent are distributed among the factors according to the DM's assessment of the relative importance of the factors. An example of assigning weights to the eight criteria is presented in figure 1.

[1] PRIME is an acronym for Preference Ratios in Multiattribute Evalaution.

**Figure 1**: Factors and weights

| Factors | Weight % |
|---|---|
| SKILL | 50 |
| 1. Education/experience | 20 |
| 2. Problem solving | 15 |
| 3. Social skills | 15 |
| RESPONSIBILITY FOR | 40 |
| 4. material resources and information | 10 |
| 5. people | 10 |
| 6. planning, development, results, work management | 20 |
| WORKING CONDITIONS | 10 |
| 7. Physical conditions | 5 |
| 8. Mental conditions | 5 |

*Source: Steps to Pay Equity,* see Harriman and Holm (2001).

Based on the classification of the jobs on defined levels for each factor each job is then assigned a total score in terms of a weighted sum of scores, which represent the evaluation of jobs with respect to (w. r. t.) each factor.

Thus in *Steps to Pay Equity* as well as in most job evaluation systems ranking of jobs w. r. t. an overall evaluation of demands and difficulties are represented by a weighted sum of scores, which can be formally stated as follows:

$$J_A \succ_{v(1-n)} J_B \Leftrightarrow \sum w_i v_i(J_A) > \sum w_i v_i(J_B)$$
$$J_A \sim_{v(1-n)} J_B \Leftrightarrow \sum w_i v_i(J_A) = \sum w_i v_i(J_B),$$

where "$\succ_{v(1-n)}$" = of more value w. r. t. an overall evaluation of all factors *1* to *n*.

"$\sim_{v(1-n)}$" = of equal value w. r. t. an overall evaluation of all factors *1* to *n*.

"$\succ_{v(i)}$" = of more value w. r. t. an evaluation of factor *i*.

"$\sim_{v(i)}$" = of equal value w. r. t. an evaluation of factor *i*.

$v_i(J_A) =$ the score assigned to job *A* representing an evaluation of factor *i*.

$v_i(J_B) =$ the score assigned to job *B* representing an evaluation of factor *i*.

$w_i$ = weight of factor *i*.

It should be pointed out that an additive value model implies that the overall value order on jobs is a weak order and that factorial independency holds, i.e. each factor contributes to the overall value independent of the values of other factors. Both conditions can be questioned in the context of job evaluations, which means that an additive value model is an unwarranted representation of the overall value of jobs and its associated relations "of more value" and "of equal value". However, in the subsequent analysis we assume that both conditions are valid.[2]

## 3. Definition of concepts used in the analysis

In the analysis in the next section we will use definitions of numerical scale types. In this case the scales or measures are value functions, $v(x)$, which represent an evaluation of a factor, where $x$ represent a value or a level of an arbitrary factor. Definitions of ordinal, interval and ratio scales are stated in terms of admissible transformations as follows (see Roberts 1979, p. 65).

The value function $v(x)$ expresses an:

*ordinal scale* if and only if each admissible alternative $v'(x)$ has the form $f(v(x))$ with $f$ strictly increasing.

*interval scale* if and only if each admissible alternative $v'(x)$ has the form $\alpha v(x)+\beta$ with $\alpha > 0$.

*ratio scale* if and only if each admissible alternative $v'(x)$ has the form $\alpha v(x)$ with $\alpha > 0$.

A value function as $v(x) = v_1(x_1) + v_2(x_2)$ is an:

*Additive conjoint measurement scale* if and only if all admissible transformations are all positive transformations, where $\alpha > 0$, $\beta_1$, $\beta_2$, as:

$v_1^*(x_1) = \alpha v_1(x_1) + \beta_1$ and $v_2^*(x_2) = \alpha v_2(x_2) + \beta_2$.

An additive conjoint measurement scale presupposes that the partial value functions, $v_1(x_1)$ and $v_2(x_2)$ are conjointly constructed (see e.g. Roberts 1979, chapter 5.4).

We also state a definition of meaningful use of numerical information as:

A statement involving numerical scales is meaningful

*if and only if*

the truth remains unchanged under all admissible transformation of all scales involved (see Roberts 1979, p. 71).

Finally, we will discuss a possible way to give an operational definition of the key notions "of more value" and "of equal value" used in the context of job evaluation. The suggested definition will be used in section 5 as a way to avoid deformation of actual differences between jobs due to the convention illustrated in *Steps to Pay Equity* for classifying jobs on predefined levels for each factor. The basis for the suggested operational definitions is the observation that the purpose of an evaluative comparison of jobs w. r. t. demands and difficulties is to give reasons for pay differentials between jobs. Thus, a DM accepting that two jobs are of equal value w. r. t. an overall evaluation seems also to have to accept that both jobs should be equally paid, i.e. the normative consequences of applying the notion "of equal value" in the context of job evaluation is that the two jobs should be equally paid.[3]

Of course, there can be other reasons not considered in the job evaluation process that have weight in the final decision about setting the pay of jobs. In the subsequent definitions we assume that the DM ignores other relevant reasons for the decision about the pay structure on jobs. In other words, when using the suggested definitions it is assumed that *ceteris paribus* conditions hold, i.e. all other factor and considerations are assumed to be equal for the jobs involved in a specific evaluation. The operational definitions of the key notions can be stated as follows. Firstly, a ranking of jobs w. r. t. an evaluation of a factor $i$ is defined as:

$$J_1 \succ_{v(i)} J_2$$

*if and only if*

[2] For an extensive discussion of necessary and sufficient conditions related to additive value model see e.g. Fishburn (1970) or Wakker (1989).

[3] Evaluative words as "of equal value" can be analysed as intermediary concepts. A function of intermediary concepts is to couple descriptive grounds to normative consequences. In the context of job evaluation this corresponds to the function of the concept "of equal value" to couple demands and difficulties of jobs to norms about pay settings. For an extensive discussion of intermediary concepts, see Lindahl and Odelstad (1996).

considering the differences between the two jobs w. r. t. factor *i* the DM finds it reasonable that the pay of *job 1* should be higher than the pay of *job 2*.

Secondly, ranking of pair of jobs w. r. t. an evaluative difference between the jobs w. r. t. factor *i* is defined as:

$$\langle J_1 J_2 \rangle \succ_{dv(i)} \langle J_3 J_4 \rangle$$

*if and only if*

considering the differences w. r. t. factor *i* the DM finds it reasonable that the pay differential between *job 1* and *job 2* should be larger than the pay differential between *job 3* and *job 4*.

This operational definition of ranking of pairs of jobs w.r.t. evaluative differences is similar to an operational definition of "strength-of- preference" in terms of the concept "the willingness to pay" (see Fishburn 1970).

Illustrating the use of the operational definitions by a simple example concludes this section. Assume that four jobs are to be evaluated w. r. t. *Skills*, which is measured in period of training as presented in Table 1.

**Table 1**: Jobs and required skills

| *Jobs* | $J_1$ | $J_2$ | $J_3$ | $J_4$ |
|---|---|---|---|---|
| *Skills* | 2 years | 1 year | 6 months | 3 months |

Considering the difference between the four jobs w. r. t. period of training the DM finds it reasonable to pay the jobs differently, i.e. the following ranking is implied by the operational definition:

$$J_1 \succ_{v(Skills)} J_2 \succ_{v(Skills)} J_3 \succ_{v(Skills)} J_4 \,,$$

which can be represented by an ordinal value function as:

$$v(J_1) > v(J_2) > v(J_3) > v(J_4) > 0 \,.$$

Secondly, considering the differences between the jobs w. r. t. skills the DM find it reasonable that the pay differential between *job 1* and *job 2* should be larger than the pay differential between *job 2* and *job 3*, and further, the pay differential between *job 2* and *job 3* should be larger than the pay differential between *job 3* and *job 4*, i.e. the following ranking of pair of jobs is implied by the operational definition:

$$\langle J_1 J_2 \rangle \succ_{dv(Skills)} \langle J_2 J_3 \rangle \succ_{dv(Skills)} \langle J_3 J_4 \rangle ,$$

which can be represented by an ordered metric scale as:

$$v(J_1) - v(J_2) > v(J_2) - v(J_3) > v(J_3) - v(J_4) > 0 .$$

If we assume that the DM can rank all pairs of jobs a higher ordered metric scale is established as for example:

$$v(J_1) - v(J_4) > v(J_1) - v(J_3) > v(J_1) - v(J_2) > v(J_2) - v(J_4) > v(J_2) - v(J_3) > v(J_3) - v(J_4) > 0 .$$

Ordered metric scales will be used in section 5 to discuss a possible way to avoid deformations of actual and relevant differences between jobs which occur due to classifications of jobs on predefined levels.[4]

## 4. An analysis of the specification of numerical job evaluation models

### *4.1. Introduction*

In this section we will analyze the construction of scores and weights as recommended in *Steps to Pay Equity*. The analysis is based on simple examples where we assume that only two factors are relevant for an evaluative comparison of jobs. First, we analyze the construction of levels of each factor, which are represented by scores. In the next subsection we analyze the construction of weights.

### *4.2. Construction of levels and scores*

Each factor is divided into a number of levels, which are ranked by qualitative value judgments as in Table 2. Scores represent the ranking of levels. Each job is then classified on a level of each factor that best fits the job description concerning demands and difficulties. The result of the job classification is that each job is assigned a score representing the evaluation w. r. t. each factor.

**Table 2:** Qualitative judgment and rating of levels

| *Levels of a criterion* | *Qualitative judgment* | *Scores* |
|---|---|---|
| $L_i^5$ | *Very high demand* | $v_i(L_i^5) = 5$ |
| $L_i^4$ | *High demand* | $v_i(L_i^4) = 4$ |
| $L_i^3$ | *Normal demand* | $v_i(L_i^3) = 3$ |
| $L_i^2$ | *Low demand* | $v_i(L_i^2) = 2$ |
| $L_i^1$ | *Very low demand* | $v_i(L_i^1) = 1$ |

The ranking of levels represented by the scores can formally be expressed as:

$L_i^5 \succ_{v(i)} L_i^4 \succ_{v(i)} L_i^3 \succ_{v(i)} L_i^2 \succ_{v(i)} L_i^1$, where "$\succ_{v(i)}$" = of more value w. r. t. factor *i*.

If the scores are interpreted as ordinal scales the ranking of levels can be represented by all strictly increasing transformations of the scores in Table 2 as for example:

$$[v_i(L_i^5)]^2 = 25,\ [v_i(L_i^4)]^2 = 16, [v_i(L_i^3)]^2 = 9, [v_i(L_i^2)]^2 = 4, [v_i(L_i^1)]^2 = 1 .$$

In job evaluation systems as *Steps to Pay Equity* there is no discussion about what type of scales the scores are supposed to be. Obviously, the scores are on ordinal scales, but ordinal scales cannot be used in order to define value differences between ranked levels. For example, by using the scores in Table 2 the following numerical statements about the value difference can be defined:

$$\frac{v_i(L_i^5) - v_i(L_i^1)}{v_i(L_i^3) - v_i(L_i^1)} = 2 \quad \text{or} \quad v_i(L_i^5) - v_i(L_i^4) = v_i(L_i^3) - v_i(L_i^2) .$$

But an admissible transformation of the ordinal scale implies that:

[4] Ordered metric scales were introduced by Coombs (1950). For a theoretical analysis of ordered and higher ordered metric scales, see Luce and Suppes (1964). See also Siegel (1964).

$$\frac{[v_i(L_i^5)]^2-[v_i(L_i^1)]^2}{[v_i(L_i^3)]^2-[v_i(L_i^1)]^2}=3 \quad \text{or} \quad \left[v_i(L_i^5)\right]^2-\left[v_i(L_i^4)\right]^2>\left[v_i(L_i^3)\right]^2-\left[v_i(L_i^2)\right]^2.$$

Obviously, scores interpreted as ordinal scales cannot be used to define value differences between the levels. However, in job evaluation systems as *Steps to Pay Equity* there is no methodological discussion about in what way stronger scales than ordinal scales can be constructed. It seems that the scores representing the rank-order of levels is assumed to be an interval scale but this should be justified. Assuming that the scores given in Table 2 is an interval scale is to stipulate that the value differences between adjacent levels are equal, i.e. an equal spaced interval scale is not justified.

Further, as is well known, an additive value model, where the partial value functions are ordinal scales, cannot meaningfully define an overall ranking of jobs, which can be illustrated by a simple example. We assume that two jobs are classified on levels as: $\left\langle L_1^5, L_2^1\right\rangle$ and $\left\langle L_1^3, L_2^3\right\rangle$, respectively. The levels are assigned scores as in Table 2 and both factors have equal weight, which implies that both jobs are of equal value since:

$$v_1(L_1^5)+v_2(L_2^1)=v_1(L_1^3)+v_2(L_2^3)=6,$$

But an admissible transformation of the ordinal scales implies that:

$$[v_1(L_1^5)]^2+[v_2(L_2^1)]^2=26>[v_1(L_1^3)]^2+[v_2(L_2^3)]^2=18.$$

Thus the ranking of the two jobs depends on the preferred ordinal scale, which means that the result of job evaluations depends on what numbers are preferred to represent the ranking of levels. Or in other words the convention to represent ranking of levels by a scale from 1 to 5 determines the overall ranking of jobs with respect to an overall evaluation of demand and difficulties. A reasonable requirement on an evaluation method is that the result in terms of a rank-order should not depend on an arbitrary and conventional application of numbers, which is to use numerical information in formally incorrect and misleading way.

In order to justify that the scores are on an interval scale, i.e. contain information about value differences between levels, a method has to be suggested when the levels are defined. We demonstrate a simple method, so called mid-value splitting, that can be applied in order to construct five levels consistent with an interval scale (see Keeney and Raffia, 1993). First, the

DM defines the highest and lowest level designated as in Table 2. Secondly, the DM constructs a level ordered between the highest and lowest ranked levels that is consistent with the qualitative judgment about value differences as:

$$\left\langle L_i^5, L_i^3 \right\rangle \sim_{dv(i)} \left\langle L_i^3, L_i^1 \right\rangle, \text{ where "} \sim_{dv(i)} \text{" = of equal value difference w. r. t. factor } I.$$

It means that the DM judge that the value difference between the fifth and the third level is equal to the value difference between the third level and the first and lowest ranked level. Thirdly, the fourth and the second level are constructed in such way as is consistent with the following value judgments:

$$\left\langle L_i^5, L_i^4 \right\rangle \sim_{dv(i)} \left\langle L_i^4, L_i^3 \right\rangle \text{ And } \left\langle L_i^3, L_i^2 \right\rangle \sim_{dv(i)} \left\langle L_i^2, L_i^1 \right\rangle.$$

Finally, as a check of consistency the DM has to accept the following value judgment:

$$\left\langle L_i^4, L_i^2 \right\rangle \sim_{dv(i)} \left\langle L_i^3, L_i^1 \right\rangle.$$

If the definition of the five levels is consistent with this procedure the scores in Table 2 is on an interval scale. But for practical reasons such a scaling procedure might not be feasible. One such reason is that many of the factors relevant in job evaluations are on closer examination, constituted by a number of sub-factors. For example in *Steps to Pay Equity* the factor "*Social skills*" is defined as follows:

> *"Measured by: communication, co-operation, contacts, cultural understandings, empathy, service."*

Apart from the ambiguous meaning of the sub-factors it seems hard to believe that a DM can in a sensible way construct levels of such a multidimensional factor which are consistent with an equally spaced interval scale.

But in addition to such practical difficulties there are also more principal problems concerning the construction of levels consistent with an interval scale. Such a construction presupposes that levels can be realized in such a way that value differences between levels are equally spaced. But at least for some types of factors, the associated levels are constituted beforehand, e.g. educational requirements measured in terms of period of training. There is no reason to believe that differences in period of training between jobs would be consistent with an equally spaced interval scale used for the purpose of evaluating jobs.

Finally, even if it were possible to construct levels consistent with an interval scale classification of jobs into predefined levels might give rise to an extensive deformation of the actual differences between jobs as regards demands and difficulties. This means that the result of a job evaluation in terms of weighted sum of scores assigned to the jobs might have a very weak correspondence with the actual and relevant differences between jobs. A possible remedy of such deformations is discussed in section 5.

### *4.3. Construction of weights*

The recommendation in *Step to Pay Equity* about how weights should be assigned is evident from the following quotation:

> "Weighting different factors against each other and deciding their impact on the result is referred to as weighting. The assigning of weights may have a significant impact on the final result. Users must, on the basis of their own specific objectives, determine what weight to attach to the various factors. Different companies have different values depending upon the focus and goals of the operations and what work is performed. This will be expressed in the weight given to the various factors in *Steps to Pay Equity*. The individual company is best equipped to make such assessments. It may be beneficial to test different alternatives and arrive at a satisfactory weighting level through discussion. Such discussions may be based on different documents describing policy and guidelines for the operations, documents which express important values within the company."[5]

However, in the quotation there is no explicit definition of the key notion "weight". It seems that the authors do not distinguish between weights as numerical scale constants and weights as qualitative relations about importance. A reasonable interpretation is that the assigned weights in some sense are intended to represent the DM's opinion about the importance of the factors in terms of influence on the ranking of jobs. But it is easy to see that numerical weights per se cannot reasonably represent a DM's intuition about the relative importance of various factors in terms of influence on the ranking of jobs. The point can be illustrated by an example. Assume two factors are relevant for an evaluation of jobs. The factors are divided into five levels as in Table 2 and constructed consistent with an interval scale. Assume that

[5] See Harriman and Holm (2001 p. 12).

one job is classified on levels as: $\left\langle L_1^5, L_2^1 \right\rangle$ and a second job is classified on levels as: $\left\langle L_1^1, L_2^5 \right\rangle$. Assume that the DM assesses both factors to be of equal importance, which we represent by the weights as $w_1 = w_2$, where $w_1 + w_2 = 1$. The specified job evaluation model implies that both jobs are of equal value since:

$$0.5 \cdot v_1(L_1^5) + 0.5 \cdot v_2(L_2^1) = 0.5 \cdot v_1(L_1^1) + 0.5 \cdot v_2(L_2^5).$$

Thus, if the DM accepts that both factors are of equal importance the DM also accepts that both jobs are of equal value. But an admissible transformation of one of the scales as $v_2^* = 2 \cdot v_2$ implies that: $v_1(L_1^5) + v_2^*(L_2^1) < v_1(L_1^1) + v_2^*(L_2^5)$, which is obviously inconsistent with the fact that the DM accepts both jobs to be of equal value. But it is of course easy to adjust the weights in order to reestablish that both jobs are of equal values as follows:

$$w_2 = \frac{w_1}{2}, \text{ which implies}$$

$$0.66 v_1(L_i^5) + 0.33 v_2^*(L_i^1) = 0.66 v_1(L_1^1) + 0.33 v_2^*(L_2^5).$$

Obviously, the weights are to be interpreted as a scaling constant, which has to be adjusted in order to make the ranking of jobs independent of admissible scale transformations. Thus, the weights cannot per se represent a DM's assessment about the importance of various factors in terms of influence on the ranking of jobs. The notion "importance" has to be defined in terms of qualitative judgments about relative influence of the factors on the overall ranking of jobs.

We suggest a definition of the relations "of more importance" and "of equal importance", which is common in Multi Criteria Decision Analysis (see von Winterfeld and Edwards 1986). The definition is based on a DM's judgment about relative value differences between the highest and lowest ranked levels for two factors. In a two-factor model definitions are as follows:

*factor 1* $\sim_I$ *factor 2* if and only if $[L_1^5, L_1^1] \sim_{dv(1-2)} [L_2^5, L_2^1]$

and

*factor 1* $\succ_I$ *factor 2* if and only if $[L_1^5, L_1^1] \succ_{dv(1-2)} [L_2^5, L_2^1]$, where "$\sim_I$" = "of equal importance" and "$\succ_I$" = "of more importance".

"$\sim_{dv(1-2)}$" = "the value difference between the highest and lowest ranked level w. r. t. *factor 1* is equal to the corresponding value difference w. r. t. *factor 2*".

"$\succ_{dv(1-2)}$" = "the value difference between the highest and lowest ranked level w. r. t. *factor 1* is larger then corresponding value difference w. r. t. *factor 2*".

Thus, if we assume that both factors are of equal importance and that the scores as in Table 2 are interval scales the following equality holds:

$$v_1(L_1^{i+1}) - v_1(L_1^i) = v_2(L_2^{k+1}) - v_2(L_2^k).$$

However, an admissible transformation of the scale of the second factor as: $v_2^* = \alpha_2 v_2 + \beta_2$, implies that in order to correctly represent the equality between adjacent levels the second factor has to be multiplied by $\frac{1}{\alpha_2}$, which implies:

$$v_1(L_1^{i+1}) - v_1(L_1^i) = \frac{1}{\alpha_2}\left[\alpha_2(v_2(L_2^{j+1}) - v_2(L_2^j))\right].$$

Thus, if we start from interval scales as in Table 2 and assume that the factors are of equal importance implies that:

$\frac{w_2}{w_1} = 1$, whereas for the transformed scale the relation between weights is:

$$\frac{w_2}{w_1} = \frac{1}{\alpha_2}.$$

Thus, even if the suggested definition of importance is applied the weights are to be interpreted as scaling constants that have to be adjusted for admissible transformation of scales.

It should be noted that the relations "of more importance" and "of equal importance" defined as above are of course highly unstable relations. If the number of levels is changed in one or more of the factors then it might be necessary to adjust the weights. However, according to recommendations in the *Steps to Pay Equity* weights should not be adjusted due to a change in the number of levels as is evident in the quotation.

> "In the basic version of *Steps to Pay Equity* each factor has five levels of difficulty. Some users may prefer to increase or reduce the number of levels in one or more of the factors. *The weighting of factors is not affected by the changing of the number of levels. However, the calculation of the number of points per level is affected*."[6] (Emphasis added)

But this recommendation can give rise to counter-intuitive results when the number of levels is changed, which we illustrate in the following example. We assume that the DM assesses both factors to be of equal importance, i.e.

$$[L_1^5, L_1^1] \sim_{dv(1-2)} [L_2^5, L_2^1] .$$

We assume that the DM decides to add to the first factor a sixth level ranked higher than the fifth level, i.e.

$$L_1^6 \succ_{v(1)} L_1^5 .$$

It seems reasonable to assume that the value difference between the highest and lowest ranked level w. r. t. the first factor has become larger than the corresponding value difference w. r. t. the second factor, i.e.

$$[L_1^6, L_1^1] \succ_{v(1-2)} [L_2^5, L_2^1] .$$

By the definition suggested above the first factor has now become more important than the second factor, which means that the numerical weights have to be adjusted from $w_1 = w_2$ to

[6] See Harriman and Holm (2001, p. 12).

$w_1 > w_2$. But according to the recommendation in *Steps to Pay Equity* the weights should not be adjusted, which implies:

$$[L_1^5, L_1^1] \prec_{dv(1-2)} [L_2^5, L_2^1].$$

Thus, the value difference between the fifth and the first level w. r. t. *factor 1* has become less then the corresponding value difference w. r. t. *factor 2*. In other words, this means that the actual influence of the first factor on the overall ranking has decreased due to an extension of a sixth level. Of course, even if an added level does not extend the range of a factor these counter-intuitive results emerge, which can be illustrated as follows. We assume that two jobs are classified as: $\left[L_1^3, L_2^2\right]$ and $\left[L_1^2, L_2^3\right]$, respectively, and the factors are scaled as in Table 2 and have equal weights, which implies:

$$v_1(L_1^3) + v_2(L_2^2) = v_1(L_1^2) + v_2(L_2^3) = 5.$$

Assume that the DM decides to add a level, $L_2^*$, to the second factor ranked between e. g. the fifth and the fourth level, i.e.

$$L_2^5 \succ_{v(2)} L_2^* \succ_{v(2)} L_2^4 \succ_{v(2)} L_2^3 \succ_{v(2)} L_2^2 \succ_{v(2)} L_2^1.$$

The recommendation in *Steps to Pay Equity* about how points should be adjusted when the number of levels are increased means that 5 points should now be equally distributed across six levels of *factor 2*, which means that the value difference between adjacent levels of *factor 2* becomes:

$v_2(L_2^{i+1}) - v_2(L_2^i) = \frac{5}{6}$, which in turn implies:

$$v_1(L_1^3) + v_2(L_2^2) = \frac{28}{6} > \frac{27}{6} = v_1(L_1^2) + v_2(L_2^3).$$

This means that *job1* is ranked higher than *job2* even if the definition of the levels and the classification of both jobs have not changed. This feature of the job evaluation system *Steps to Pay Equity* can be summarized as follows:

Firstly, the DM evaluates two jobs classified as above: When the set of jobs is $\{J_1, J_2\}$ the result of the evaluation is: $J_1 \sim_{v(1-2)} J_2$. Secondly, the DM decides to add a level because of the feature of a third job that is included in the evaluation. When the set of jobs is $\{J_1, J_2, J_3\}$ the result of the evaluation is: $J_1 \succ_{v(1-2)} J_2$. Thus the construction of scores as recommended in *Steps to Pay Equity* is not consistent with the condition "Independent of irrelevant alternatives", which is regarded as important principle for decision making and evaluation.[7]

But this undesirable feature of a job evaluation system can easily be avoided. One reason for adding levels seems to be that there are jobs that the DM finds difficult to classify on any of the defined levels – the jobs fall between two adjacent levels. In the example above the DM finds it necessary to define a level ranked between the fifth and fourth level in order to make a reasonable classification of the third job. Defining another level can of course improve the accuracy of the classification. But an extension of the number of levels should not change the numerical intervals between the adjacent levels as recommended in *Steps to Pay Equity*. Instead of rescaling the numerical intervals between all adjacent levels as in the example it seems more reasonable to assign scores to the added level as for example as:

$$v_2(L_2^5) - v_2(L_2^*) = v_2(L_2^*) - v_2(L_2^4) = \frac{1}{2}.$$

We end the section by a short comment on an observation that it is common in job evaluations to present the result both in terms of non-weighted as well as in a weighted sum of scores. But a non-weighted sum cannot meaningfully represent a ranking of jobs, even if the scale of each factor is an interval scale, which is illustrated by the following example:

$$v_1(L_1^5) + v_2(L_2^1) = v_1(L_1^3) + v_2(L_2^3).$$

Admissible transformations of the scales are e.g.: $v_1^* = 2v_1$ and $v_2^* = v_2$, i.e. the identity transformation, which implies that:

$$v_1^*(L_1^5) + v_2^*(L_2^1) > v_1^*(L_1^3) + v_2^*(L_2^3).$$

[7] For a discussion see Arrow (1963) or Sen (1970).

One objection that can be raised is that the transformation of both scales has to be similar as:

$v_i^* = \alpha v_i + \beta_i,\ \ \alpha > 0$, where $\alpha$ is equal for both scales, whereas $\beta_i$ is a specific intercept for each scale. But such a more restrictive set of transformation presupposes that the scales are dependent or coordinated in some sense, i.e. the additive two-factor model forms an additive conjoint measurement scale. This means that the DM has made explicit judgments that the value difference between all adjacent levels of both factors are equal, i.e. an equally spaced interval scale is established for all factors, which implies:

$(L_1^{l+1} L_1^l) \sim_{dv(1-2)} (L_2^{k+1} L_2^k)$, which is represented by $v_1(L_1^{l+1}) - v_1(L_1^l) = v_2(L_2^{k+1}) - v_2(L_2^k)$.

But this implies that at the construction of the scale all levels are adjusted in such a way that an equally spaced interval scale can be applied. But this means that the relative importance of the factors is already determined in terms of the relative influence on the overall values, and a weighting process is not called for. In such a case the expression "non-weighted sum" is used in a misleading way.

## 5. A remedy of deformation of actual differences between jobs

The convention in many job evaluations systems, as in *Steps to Pay Equity*, to classify the jobs on predefined levels might give rise to an extensive deformation of actual differences between jobs with respect to demands and difficulties, which means that there might be a very weak correspondence between ranking of jobs according to a job evaluation and actual differences between the jobs. Such a deformation can be illustrated as in Table 3, where four jobs are classified on three different levels for each of three factors. Further, a DM's intuitive evaluation regarding relative value differences between the jobs is represented in Table 3 as relative distances. Thus concerning the first factor the DM considers that the value differences between the first three jobs are relatively small compared to the value difference between the third and the fourth job.

**Table 3**: Classification and intuitive evaluation of jobs

| Levels \ Factors | 1 | 2 | 3 |
|---|---|---|---|
| **3** | $J_1$ | $J_4$ | |
| **2** | $J_2$<br>$J_3$ | $J_2$<br>$J_3$<br>$J_1$ | $J_2$<br>$J_3$<br>$J_1 \sim_{v(3)} J_4$ |
| **1** | $J_4$ | | |

If we assume that the DM assigns equal weights to all factors and that the levels are scored as in Table 3, the following ranking is implied:

$$V(J_1) = 7 > V(J_2) = V(J_3) = V(J_4) = 6 .$$

But an inspection of the DM's intuition about reasonable value differences between jobs w. r. t. each criterion implies that the ranking is obviously counter-intuitive. A more reasonable ranking seems to be that *job 2* is ranked higher than *job 1*, since the value differences between *job 2* and *job 1* w. r. t. the second and third factor are significantly larger than the value difference between *job 1* and job *2* w. r. t. the first factor. The small value difference in favour of *job 1* w. r. t. *factor 1* cannot compensate for the value differences in favour of *job 2* w. r. t. *factor 2* and *factor 3*. For the same reason *job 2* should be ranked higher than *job 4* and *job 3*. Thus, the DM's intuitive evaluation supports a ranking as:

$$J_2 \succ_{v(1-3)} J_3 \succ_{v(1-3)} J_1 \sim_{v(1-3)} J_4 .$$

The deformations of the DM's intuitive evaluation of the jobs due to classifying jobs on predefined levels can easily be avoided by using a more flexible evaluation model. We demonstrate the use of a model, named PRIME, which does not require that the jobs are classified on predefined levels and supports the use of more flexible scales compared to scores as in Table 2. The PRIME-model assumes that the overall value of each job can be represented by an additive value model as:

$$V(J_i) = w_1 \cdot v_1(J_i) + w_2 \cdot v_2(J_i) + w_3 \cdot v_3(J_i)\text{, where } i = 1, 2, 3 \text{ and } 4.$$

The specification of the additive value model proceeds in three steps. We use the example discussed above in order to demonstrate the evaluation process defined by the PRIME-model, which is named an elicitation tour. (An extensive description of the PRIME-model is given in Appendix A.)

Firstly, the DM makes ordinal judgments w. r. t. each factor. The DM's value judgments according to Table 3 can be represented by ordinal value functions, which are normalized as conventional:

*Factor 1:* $1 = v_1(J_1) > v_1(J_2) > v_1(J_3) > v_1(J_4) = 0$

*Factor 2:* $1 = v_1(J_4) > v_1(J_2) > v_1(J_3) > v_1(J_1) = 0$

*Factor 3:* $1 = v_1(J_2) > v_1(J_3) > v_1(J_1) = v_1(J_4) = 0$ .

Secondly, the DM makes cardinal judgments w. r. t. value differences between jobs. According to Table 3 this means that:

*Factor 1*: $v_1(J_3) - v_1(J_4) > v_1(J_2) - v_1(J_3) = v_1(J_1) - v_1(J_2) > 0$

*Factor 2*: $v_2(J_1) - v_2(J_3) > v_2(J_4) - v_2(J_2) = v_2(J_2) - v_2(J_3) > 0$

*Factor 3*: $v_3(J_2) - v_3(J_3) = v_3(J_3) - v_3(J_1) = v_3(J_3) - v_3(J_4) > 0$ .

The DM's intuition about relative value differences between the jobs are represented by equalities and inequalities, which is essentially a representation corresponding to an ordered metric scale as discussed in section 3.

Thirdly, the DM assigns weights to each factor. The PRIME model defines weights as the definition suggested in section 4, which implies by the assumption in the example that:

$$v_1(J_1) - v_1(J_4) = v_2(J_4) - v_2(J_1) = v_3(J_2) - v_3(J_4) \text{ or } w_1 = w_2 = w_3 .$$

The weights are normalized as follows: $w_1 + w_2 + w_3 = 1$, which means that the overall value functions defined in PRIME take values in the numerical interval: $0 \le V(J_i) \le 1$.

The overall value of each job is represented by numerical intervals as follows:

$$V(J_i) = \left[\min(V(J_i); \max(V(J_i)\right].$$

The minimum and maximum values are established by solving two linear programs, i.e. the minimum and the maximum values consistent with constraints imposed by the specification of the value functions in step 1 to 3 described above. The overall value intervals for the four jobs are as in Table 4.

**Table 4:** Overall values and rank

| Jobs | Min value | Max value | Rank |
|---|---|---|---|
| 1 | 0,333 | 0,333 | 3 |
| 2 | 0,777 | 1 | 1 |
| 3 | 0,389 | 0,833 | 2 |
| 4 | 0,333 | 0,333 | 3 |

Overlapping value intervals occurs for job 2 and 3, which means that a pairwise dominance criterion has to be applied. The pair wise dominance criterion is defined as follows:

$$J_k \succ_D J_l \Leftrightarrow \max[V(J_l) - V(J_k)] < 0 \Leftrightarrow \max[\sum_{i=1}^{n} w_i v_i\,(J_l) - \sum_{i=1}^{n} w_i v_i(J_k)] < 0,$$

i.e. job $J_k$ is ranked higher than job $J_l$ if and only if the overall value of $J_k$ exceeds that of $J_l$ for all feasible solutions of the linear constraints implied by the interval-valued statements in the elicitation tour defined by step 1 to 3 above. A non-dominance relation occurs between the two jobs $J_k$ and $J_l$, if the inequality does not hold, i.e. there are overall values consistent with the constraints imposed by the value judgments in step 1 to 3 above implying that: $V(J_l) > V(J_k)$. In other words, we cannot, based on the value judgments in step 1 to 3, distinguish between the two jobs.

Comparing *job 2* and *job 3* by means of the pairwise dominance criterion implies in this case that *job 2* is ranked higher than *job 3*.

Relating the judgments to the consequences in terms of recommended pay differentials might strengthen the judgments regarding comparisons of value difference, specified in step 2. In other words, by making the judgments about relative value differences operational as discussed in section 3 the DM might find it possible to state more precise value judgments. We demonstrate the reasoning by an example. Assume that the first factor corresponds to *Skills* measured in period of training. We assume that the differences between the three highest ranked jobs are three months, respectively, and that the difference between job *3* and job *4* is two years of period of training. From the point of view about reasonable pay differentials, when only the factor *Skills* is considered, the DM is assumed to be able to justify a judgment as:

$$8 \geq \frac{v_1(J_3) - v_1(J_4)}{v_1(J_2) - v_1(J_3)} \geq 2 \, .$$

The interpretation of the DM's intuition is that, when only differences in period of training are considered, the pay differential between job *3* and job *4* can be at least twice and at most eight times the pay differential between job *2* and job *3*. The consequence in this example of strengthening of judgments about relative value differences is that the overall value intervals for *job 2* and *job 3* do not overlap as evident in Table 4, which means that the ranking between the two jobs now become obvious.

**Table 5:** Overall values and rank

| Jobs | Min Value | Max Value | Rank |
|---|---|---|---|
| **1** | 0,333 | 0,333 | 3 |
| **2** | 0,833 | 0,933 | 1 |
| **3** | 0,444 | 0,755 | 2 |
| **4** | 0,333 | 0,333 | 3 |

Thus possible strengthening of an ordered metric scale demonstrated above can generally be expressed as ratio estimates between two value differences as follows:

$U \geq \frac{v_i(J_k) - v_i(J_l)}{v_i(J_m) - v_i(J_n)} \geq L$, where U and L represents an upper and lower limit of the ratio, respectively.

The interpretation of the expression is that from the DM's point of view ratios outside the upper and lower limits are unreasonable.

The PRIME-model supports a similar treatment concerning the DM's assessment of weights. In the example above we assumed equal weight for the factors. But for purpose of demonstration we assume that the DM finds it reasonable to claim that *factor 1* is more important than *factor 2* and *factor 3*, but not more than twice as important. The judgment can be represented as ratios:

$$1 < \frac{v_1(J_1) - v_1(J_4)}{v_2(J_4) - v_2(J_1)} < 2 \text{ And } 1 < \frac{v_1(J_1) - v_1(J_4)}{v_3(J_2) - v_3(J_1)} < 2 \text{ or}$$

$$1 < \frac{w_1}{w_2} < 2 \text{ And } 1 < \frac{w_1}{w_3} < 2 .$$

The assessment of weights can be given the same operational interpretation as for the judgment of value differences within a factor. Thus claiming that *factor 1* is more important but not more than twice as important as *factor 2* means, based on intuition about reasonable pay differentials, that the DM judg that the differences between job *1* and job *4* w. r. t. *factor 1* justify at most twice as large pay differential as the pay differential justified by the difference between job *4* and job *1* w. r. t. *factor 2*. The result of assigning *factor 1* a higher but imprecise relative weight as presented in Table 6 implies that there is no reason to discriminate between job *1* and job *3*, i.e. job *1* and job *3* are judged by the DM to be of equal value.

**Table 6:** Overall values and rank

| Jobs | Min. Value | Max. Value | Rank |
|---|---|---|---|
| **1** | 0,335 | 0,5 | 2 |
| **2** | 0,812 | 0,933 | 1 |
| **3** | 0,444 | 0,766 | 2 |
| **4** | 0,25 | 0,332 | 4 |

The possibility to justify more precise ratios between value differences seems to depend on type of factor that is evaluated. In the example above we used the factor *Skills*, measured in period of training, in order to demonstrate an intuitive evaluation strategy that can justify more precise ratios between value differences, which in turn means that it is possible to discriminate between more of the jobs considered in an evaluation. In the example, however, the DM could base his or her intuition about reasonable ratios on an objective measure in terms of period of training. But regarding other types of factors as *Social Skills,* defined in section 3, such more precise value judgments stronger than an ordered metric scale might be difficult to defend. Justifying an upper and lower limit for ratio estimates of this type of factors might be beyond the capacity of a DM bearing in mind that the factor is constituted by a number of sub-factors, which in turn have no obvious operational definition similar to *Skills*.

However, we end the section by claiming that based on the example above by using Multi-Attribute Evaluation Models such as PRIME supporting the use of imprecise value information it is possible to avoid an overall ranking of jobs that is the result of an unacceptable deformation of actual differences between jobs with respect to demand and difficulties.[8]

## 5. Summary

In this paper we have analyzed the use of numerical information in the context of job evaluation. The analysis is based on the job evaluation system *Steps to Pay Equity,* which is recommended by the European Project on Equal Pay supported by the European commission. The result of the analysis can be summarized as follows:

Firstly, in *Steps to Pay Equity* no method is suggested to justify stronger scales than ordinal scales that are intended to represent the evaluation of the various factors that describe demands and difficulties in the jobs. As is well known results in terms of rankings of jobs based on addition of ordinal scales are very unstable for admissible transformation of ordinal scales. In other words, the ranking of jobs is dependent on what specific scales are applied.

Secondly, in *Steps to Pay Equity* there is no explicit definition or explanation how the weights should be interpreted. The analysis demonstrates that numerical weights cannot per se represent a DM's opinion about the relative importance of the various factors. Thus, the basis for the DM's judgment concerning the relative importance of factors is unclear and something that hampers a constructive discussion about the reasonability of the assigned weights. Further, the recommendation in *Steps to Pay Equity* not to adjust weights when the number of levels of a factor is changed means that the suggested job evaluation system violates the condition "Independent of Irrelevant Alternatives". In this context it means that a ranking between two jobs can be influenced of an change in the number of defined levels for a specific factor, which is irrelevant to both of them.

Thirdly, the convention applied in *Steps to Pay Equity* to classify jobs on predefined levels gives rise to deformations of actual differences between jobs with respect to demands and difficulties. Thus the ranking of jobs established by *Steps to Pay Equity* can have a weak correspondence to actual and relevant differences between jobs regarding demand and difficulties, which means that the ranking of jobs is in some sense an unjustified guidance for setting fair and gender neutral pay settings. We suggested a possible remedy by illustrating the use of a specific Multi Attribute Evaluation Model, the PRIME model, which supports the use of imprecise value information.

[8] For a more extensive application of the PRIME model in a similar evaluation context that concerns evaluations of employee performance, see Blomskog (2005). For another attempt to model imprecise value information in the context of job evaluations, see Spyridakos (1999).

## References


Arrow, K. J. (1963), *Social choice and individual values*, New York: Wiley, 2nd ed.

Blomskog, S. (2005),"Evaluation of employee performance based on imprecise value judgments – Two experiments", resubmitted for revision to *Journal of Multi-Criteria Decision Analysis*.

Coombs, C. H. (1950), "Psychological scaling without measurement", *Psychological Review,* 57, 145-158.

Fishburn, P.C. (1970), *Utility Theory for Decision making,* John Wiley & Sons, Inc: New York.

Gustafsson, J., Gustafsson, T. and Salo, A. (2000), PRIME Decisions – An Interactive Tool for Value Tree Analysis. V. 1.0, Computer software, System Analysis Laboratory, Helsinki University of Technology. (www.sal.hut.fi/Downloadables).

Gustafsson, J., Salo, A. and Gustafsson, T. (2001), "PRIME Decisions: An interactive Tool for Value Tree Analysis", in M. Köksalan and S. Zionts (eds.), *Multiple Criteria Decision Making in The New Millennium,* Lecture Notes in Economics and Mathematical Systems **507**, Springer-Verlag: Berlin, 165-186

Harriman, A. and Holm, C. (2001), *Steps to Pay Equity – An easy and quick method for the evaluation of work demands.* URL:http://www.equalpay.nu/docs/en/steps_to_pay_equity.pdf (2004-01-13).

Keeney, R. L. and Raiffa, H. (1993), *Decision with Multiple Objectives*, Cambridge University Press.

Lindahl, L. and Odelstad, J. (1996), "Grounds and Consequences in Conceptual Systems", in S. Lindström et al (eds.), *Odds and Ends. Philosophical Essays Dedicated to Wlodek Rabinowitz on the Occasion of His Fiftieth Birthday*, Department of Philosophy: Uppsala.

Luce, R. D. and Suppes, P. (1964), "Preference, Utility and Subjective Probability", in R. Duncan Luce, Robert R. Bush and Eugene Galanter (eds.), *Handbook of Mathematical Psychology,* Vol III., John Wiley and Sons, New York.

Roberts, F.S. (1979), *Measurement Theory with Applications to Decision-making, Utility and the Social Sciences*, Addison-Wesley: London.

Salo, A. and Hämäläinen, R.P. (2001), "Preference Ratios in Multiattribute Evaluation (PRIME) – Elicitation and Decision Procedures Under Incomplete Information", *IEEE Transactions on System, Man and Cybernetics – Part A: Systems and Humans* **31**, 533-545.

Sen, A. K. (1970), *Collective choice and social welfare,* San Francisco: Holden-Day.

Siegel, S. (1964), "A method for obtaining an ordered metric scale", in Messick & Brayfield (eds.), *Decision and Choice,* McGraw-Hiring.

Spyridakos, A., Siskos, Y., Yannacopoulas, D. and Skouris, A. (1999), Multicriteria job evaluation for large organizations. *European Journal of Operational Research* **130**, 375-387.

von Winterfeldt, E. and Edwards, W. (1986), *Decision Analysis and Behavioral Research*. Cambridge University Press: New York.

Wakker, P. (1989), *Additive Representations of Preferences – A New Foundation of Decision Analysis*, Kluwer Academic Publishers, Dordrecht.

## Appendix A: The PRIME-model

### *A.1. PRIME-model*

The PRIME (Preference Ratios in Multiattribute Evaluation)-model is based on multi-attribute value theory. An extensive description of the model and its applications is in Salo and Hämäläinen, 2001. The PRIME-model is implemented by a full-fledged computer software, named PRIME Decisions, which is a decisions aiding tool that offers an interactive decision support. PRIME Decisions can be downloaded from: www.hut.fi\Units\SAL\Downloadables\ (Gustafsson et al, 2000).

In the PRIME-model the overall value of multidimensional alternative $X^l = \langle x_1^l, x_2^l, ....., x_n^l \rangle$, is defined by an additive value model:

$$(1)\ V(X^l) = \sum v_i(x_i^l).$$

By normalization the model can be written as:

$$(2)\ V(X^l) = \sum w_i \cdot v_i^N(x_i^l),$$

where $v_i^N(x_i^l) = \dfrac{v_i(x_i^l) - v_i(x_i^{\min})}{v_i(x_i^{\max}) - v_i(x_i^{\min})}$, and by convention: $v_i(x_i^{\min})$=0 , which implies

that $v_i^N(x_i^l) \in [0,1]$.

$w_i = v_i(x_i^{\max}) - v_i(x_i^{\min})$,

i.e. the attribute weights relate unit increases in normalized value functions to increases in the overall value.

The overall value of an ideal profile, i.e. $P(X^{\max}) = \langle x_1^{\max}, x_2^{\max}, ..., x_n^{\max} \rangle$, is normalized to one, i.e.

$$(3)\ V(X^{\max}) = V\langle x_1^{\max}, ...., x_n^{\max} \rangle = \sum w_i \cdot v_i^N(x_i^{\max}) = \sum_{i=1}^{i=n} w_i = 1.$$

The PRIME Decisions has a feature called *elicitation tour* which guides the DM through a specific sequence of elicitation steps as follows:

**Step 1**: Ordinal value judgments

The DM is asked to rank performance regarding each criterion, which is represented by an ordinal value function:

$$(4)\ v_i(x_i^{\max}) > v_i(x_i^l) > ...... > v_i(x_i^k) > v_i(x_i^{\min}).$$

**Step 2**: Cardinal value judgments

The DM is asked to elicit cardinal judgment regarding value differences between pairs of ranked performances. The imprecise cardinal value judgments are represented as interval-valued statements about ratios estimates regarding two value differences. For instance, a comparison of value difference regarding pairs of adjacent performance can be expressed as ratio estimates:

$$(5)\ L \leq \frac{v_i(x_i^{l+1}) - v_i(x_i^l)}{v_i(x_i^{k+1}) - v_i(x_i^k)} \leq U .$$

The interval [L, U] represents the degree of imprecision of cardinal value judgments regarding the two value differences. However, the PRIME-model supports ratio estimates of value differences regarding arbitrary pairs of performance:

$$(6)\ L \leq \frac{v_i(x_i^k) - v_i(x_i^l)}{v_i(x_i^m) - v_i(x_i^n)} \leq U \text{, given that } v_i(x_i^k) > v_i(x_i^l) \text{ and } v_i(x_i^m) > v_i(x_i^n).$$

**Step 3:** Weight assessment

The DM is asked to assess the weights by:

1) choosing a reference criterion, which is assigned the weight of 100 %.
2) comparing the value difference between the highest and the lowest ranked performance regarding each criterion relative to the corresponding value difference of the reference criterion. The assessments are represented by imprecise ratio estimates as:

$$(7)\ \frac{L}{100} \leq \frac{w_i}{w_{ref}} \leq \frac{U}{100} \Leftrightarrow \frac{L}{100} \leq \frac{v_i(x_i^{\max}) - v_i(x_i^{\min})}{v_{ref}(x_{ref}^{\max}) - v_{ref}(x_{ref}^{\min})} \leq \frac{U}{100},$$

where [L, U] is the numerical interval mapping the degree of imprecision of weight assessments.

The interval-valued statements expressed by the DM in an elicitation tour are translated in to a number of linear constraints. Based on the linear constraints the overall value of each performance profile is represented by a value interval computed from the two linear programs:

$$(8)\ V(X^l) \in \left[ \min \sum_{i=1}^{i=n} w_i v_i (x_i^l), \max \sum_{i=1}^{i=n} w_i v_i (x_i^l) \right] = \left[ \min V(X^l), \max V(X^l) \right].$$

*A.2. Dominance criteria and decision rules*

The PRIME Decisions provide two dominance criteria and several decision rules to help the DM to rank the alternatives, in this case lecturer. The *absolute dominance criterion* is defined as:

$$(9)\ X^k \succ_D X^l \Leftrightarrow \min V(X^k) > \max V(X^l).$$

According to the *absolute dominance criterion* alternative $X^k$ is ranked higher than $X^l$ if the smallest possible value of $X^k$ exceeds the largest possible value of $X^l$. The *absolute dominance criterion* can only be used for pairs of alternatives with nonoverlapping value intervals. In case of overlapping value intervals the *pairwise dominance criterion* has to be applied. The *pairwise dominance criterion* is defined as:

$$(10)\ X^k \succ_D X^l \Leftrightarrow \max[V(X^l) - V(X^k)] < 0 \Leftrightarrow \max[\sum_{i=1}^{i=n} w_i v_i (x_i^l) - \sum_{i=1}^{i=n} w_i v_i (x_i^k)] < 0.$$

According to this criterion alternative $X^k$ is ranked higher than alternative $X^l$ if and only if the overall value of $X^k$ exceeds that of $X^k$ for all feasible solutions of the linear constraints implied by the interval-valued statements in an elicitation tour. A non-dominance relation occurs if the inequality in (10) does not hold, i.e. there are overall values implying that: $V(X^l) > V(X^k)$. The interpretation of a non-dominance relation between an alternative $X^k$ and an alternative $X^l$ is that the DM's value information is not sufficiently precise in order to determine a ranking between the two alternatives. In that case any of the decision rules provided by PRIME decisions can be applied.

In PRIME four decision rules are stated: 1) minimax 2) maximax 3) minimax regret 4) central values. The definition and the performance of the decision rules are discussed in Salo and Hämläinen, 2001. Based on simulations they recommend the minimax regret criterion and application of central values, because they consistently outperform the other ones.

**Appendix B:** The factor plan as defined in *Steps to Pay Equity*

# The Factor plan

The basic version of *Steps to Pay Equity* is based upon MAIN AREAS, factors and aspects

**SKILL**

Factor 1. Education/experience measured by: *number of years of education, occupational experience, further education*
Factor 2. Problem solving measured by: *type of problem, creativity, independence, decision-making, development, versatility*
Factor 3. Social skills measured by: *communication, co-operation, contacts, cultural empathy, service*

**RESPONSIBILITY**

Factor 4. Responsibility for material resources and information measured by: *financial value, what the responsibility entails, independence, sequences*
Factor 5. Responsibility for people measured by: *what the responsibility entails, independence, consequences*
Factor 6. Responsibility for planning, development, results, management measured by: *the focus and scope of the responsibility, independence,*

**WORKING CONDITIONS**

Factor 7. Physical conditions measured by: *physical strain, strain on the senses, unpleasant physical conditions, risk for personal injury or illness*
Factor 8. Mental conditions measured by: *concentration, monotony, availability, trying relationships, stress*