An evaluation of the weighting method in a gender-neutral job evaluation tool recommended by the International Labor Office (ILO)

Stig Blomskog

PESO Research Report 3 (2015)
School of Social Sciences
Södertörn University

södertörns högskola
SÖDERTÖRN UNIVERSITY

The PESO Research Report series publishes original research from the School of Social Sciences, Södertörn University. The reports have undergone peer review and are published after approval from the School's Publication Committee.

Abstract

Gender-neutral job evaluation has become a key method for confirming the presence of value discrimination when accounting for job-related factors such as required skills, responsibility levels, effort and working conditions, and for correcting for a gender-biased pay setting. However, this extensive use of gender-neutral job evaluation tools makes it important to examine the validity of these tools.

The purpose of this report is to evaluate the validity of a weighting method stated in a gender-neutral job evaluation tool that is recommended by the International Labor Office (ILO). The purpose of the ILO tool is to function as a general and worldwide guideline for gender-neutral job evaluations.

The evaluation starts from the basic validity requirement that a weighting method has to be based on a correct interpretation of the weights in *additive value models*, which are used as "measures" of the value of jobs. The conclusion of the evaluation is that the ILO weighting method does not fulfill this basic validity requirement. The conclusion is reached in the following way:

First, I show that the meaning of the weights in additive value models is to determine so-called *compensatory relations* between job-related factors, which have an important impact on the results of job evaluations.

Second, by analyzing the weighting instructions in the ILO tool, I find that this weighting method is based on so-called *direct rating* of the relative importance of the job-related factors.

Third, I show that direct rating is based on an incorrect interpretation of the weights. Thus, users of the ILO tool will probably misinterpret the consequences of their weighting decisions. This, in turn, might give rise to a *biased weighting*, i.e. a weighting that the users would reject when they come to know the correct meaning of the weights.

Keywords: Job evaluation, additive value model, direct rating, biased weighting, compensatory weighting

Contents

1. Introduction

When implementing a comparable-worth policy,[1] *value discrimination* seems to partly explain the gender wage gap observed on many labor markets. A common explanation for value discrimination is that typical female jobs are under*valued* as compared to typical male jobs when accounting for job-related factors such as required skills, responsibility levels, effort and working conditions. Gender-neutral job evaluation has become a key method for confirming the presence of value discrimination and correcting for gender-biased pay setting. Many EU countries and other countries such as Australia, Canada, New Zeeland and the US use this method. Equal pay legislation also requires employers to implement gender-neutral job evaluations in many countries.

This extensive use of gender-neutral job evaluation makes it important to examine the validity of these tools. In this report I will focus on the validity of the weighting of the job-related factors, which seems to have a significant impact on the results of job evaluations.

More exactly, the purpose of the report is to evaluate the validity of a weighting method stated in the job evaluation tool (hereafter, the ILO tool), which is recommended by the International Labor Office and the European Commission.[2] Since the purpose of the ILO tool is to serve as a general and worldwide guideline for gender-neutral job evaluations it is, of course, important that the ILO tool fulfills basic validity requirements. This particularly concerns the weighting method since the designers of the ILO tool claim that the weighting of the job-related factors "has an extremely important impact on the value of jobs" (see Chicha 2008, p. 70).

The ILO tool is a so-called factor-point method, which is currently the most developed and used method.[3] In a factor-point method, the results of

[1] England (1999) defines comparable-worth policy as "strategy policies that ensure that jobs do not pay less because they are filled by women." Comparable-worth policy is adopted in many countries, e.g., Australia, Canada, EU countries, New Zeeland and the US.

[2] The job evaluation tool is named *Steps to Pay Equity*, *ISOS* and *Promoting equity – gender-neutral job evaluation for equal pay: a step-by-step guide* and is developed by the Canadian job evaluation expert M. Chicha (see Chicha 2008).

[3] For a description of the factor point method, see Armstrong et al (2003).

the evaluation of jobs are expressed or represented by means of a weighted sum of scores, i.e. *additive value models* are applied. This gives rise to the basic validity requirement that the weighting method should be based on a correct interpretation of the meaning of the weights in additive value models. This basic validity requirement constitutes the starting point for this evaluation of the weighting method stated in the ILO tool, which will be carried out through a discussion of these four questions. *First,* the basic question for the evaluation is: How should the meaning of the weights be interpreted in job evaluation when additive value models are applied? *Second*, is the weighting method in the ILO tool based on a correct interpretation of the weights? *Third*, to what kind of problems can this weighting method give rise? *Fourth*, what essential requirements should the weighting method in a job evaluation fulfill?

In order to clarify the purpose of the report, I want to point out that the evaluation presented in the report concerns a methodological and conceptual issue about the meaning of weights in job evaluation when additive value models are applied. This conceptual issue must be distinguished from the substantial normative issue concerning what weighting of job-related factors that is reasonable in a certain job evaluation situation. This kind of wage policy issue concerning a reasonable weighting can, of course, not be determined by means of an analysis of a weighting method. Instead, it is should be decided by the stakeholders taking part in a job evaluation. Nevertheless, achieving a gender-neutral weighting presupposes that the meaning of the weights is correctly interpreted.

The evaluation of the weighting method in the ILO tool will be based on multi-attribute value theory. Within multi-attribute value theory, there is an extensive number of theoretical as well as empirical studies about weighting when additive value models are applied.[4] However, in order to make the report available for readers with no or scant knowledge about multi-attribute value theory the evaluation will, as far as it is meaningful, be based on informal reasoning and simple examples.

The report has the following outline: In the next section I briefly explain the purpose of gender-neutral job evaluation and the adherent job evaluation procedure. The third section presents the evaluation of the weighting method in the ILO tool. The fourth section presents a summary

[4] Keeney and Raiffa (1993) is a classical reference on multi-attribute value theory. For a more informal discussion, see Belton and Stewart (2002).

of the evaluation. Based on multi-attribute value theory, the Appendix presents formal definitions of the weights in additive value models.

2. What is a gender-neutral job evaluation?

The key question for this evaluation is how the meaning of the weights should be interpreted when additive value models are applied in job evaluations. Correctly answering this question presupposes an understanding of the purpose of a job evaluation which, in turn, determines the interpretation of the result of a job evaluation.

The specific purpose of a gender-neutral job evaluation is to achieve a rational and gender-neutral pay grade classification of the jobs at a work place. It is essential to point out that it is certain features of the jobs, and not the performance of employees holding the jobs, that constitute the basis for the job evaluation. The job grade classification serves as a benchmark for assessing if and to what extent there are rational reasons for pay differentials between male and female employees holding jobs of equal value. If no rational reasons can be given, the employer should take measures to remove these kinds of unjustified pay differentials.

In order to achieve a gender-neutral job grade classification at a work place, a job evaluation procedure is carried out as follows. In the first stage of a job evaluation, decision makers (hereafter DMs)[5] have to, based on gender-neutral principles, choose and define certain job-related factors or so-called *compensable* factors, which are considered as relevant for a pay grade classification at the work place. An internationally accepted convention is that four so-called main criteria, *required skills, responsibility levels, effort* and *working conditions*,[6] serve as the basis for specific definitions of various compensable factors. Naturally, the specific definitions of the compensable factors vary between the different applications of job evaluation.

A basic assumption in job evaluations is that DMs' overall or aggregated evaluation of the jobs regarding the various factors can be represented by a

[5] A DM is a person or a group of persons that carries out job evaluations and is responsible for job evaluation results.

[6] These four main criteria (factors) are, for example, stated by the European Commission in *Code of Practice on the Implementation of Equal Pay for Work of Equal Value* (1996). In *Promoting Equity*, published in 2008 by the International Labour Office (ILO), the author states: "according to evaluation and compensation experts, these four factors are essential and sufficient for evaluating all the tasks performed in an organization, regardless of which economic sector the enterprise belongs to" (see Chicha, 2008, p. 27).

sum of weighted scores, i.e. additive value models are applied.[7] The model can be stated as follows:

$$V(A) = \sum w_i v_i(A),\ i = 1, 2, ..., m,$$

where:

$V(A) =$ the total score representing the overall value of a job A.

$v_i(A) =$ the score representing the value of job A regarding factor i.

$w_i \quad =$ the weight of the scoring scale $v_i(\cdot)$.

Thus, the result of a job evaluation is that each job is assigned a total score, which is constructed through a scoring and weighting procedure. The total scores are used by the DMs to decide how the jobs *should* be classified in various pay grades. For example, if two jobs receive the same total score, i.e. both jobs are of *equal value* (worth), this can be seen as an overall decision of DMs to classify the jobs in the same pay grade. This means that the total scores can be interpreted as expressing DMs' *overall decision* about how the jobs should be classified in various pay grades at the work place.[8] This interpretation does, in turn, mean that the scores and the weights will represent certain kinds of *partial decisions* in the job evaluation procedure, which are by means of the additive value model *aggregated* to an overall decision. Therefore, in order to correctly interpret the meaning of the weights, we have to answer this basic question: What kind of partial decisions will the weights actually represent in a job evaluation when additive value models are applied? This will be the key question for the evaluation of the ILO weighting method.

3. Evaluation of the weighting method in the ILO tool

The purpose of the ILO tool is to serve as a general guide for gender-neutral job evaluations. Requests for help – from states, unions, and other groups

[7] Naturally, it is not obvious that these assumption are valid. In multi-attribute theory, there is an extensive analysis of necessary and sufficient conditions that have to be fulfilled such that an additive value model exists that can represent the evaluations of multidimensional objects (see also Appendix A1).

[8] It should be pointed out that since a job evaluation is a multidimensional evaluation procedure giving rise to normative decisions about the pay setting, it is a fundamental mistake to interpret the result of a job evaluation as an empirical measure of jobs regarding some kind of descriptive feature. Such an interpretation will, among other things, give rise to an incorrect interpretation of the meaning of the weights in an additive value model.

that deal with gender and labor issues – drove its development. Its target groups consist of equal opportunity officers, HR managers and gender and financial (wage equity) specialists. This tool is based on reviews of job evaluation methods and other materials that were developed and used in various countries and on case studies and research in gender studies. The tool was tested and validated in ILO-supported training events. However, nothing specific is said about the validity of the weighting method. Thus, the purpose of this evaluation can be regarded as a validation of the weighting method recommended in the ILO tool, where the weighting is explained in the following way:

> The *weighting of evaluation factors involves determining their relative importance and assigning a numerical value to each of them.* It has an extremely important impact on the value of jobs. Even when extreme caution has been exercised during the preceding steps, inconsistencies and bias can nevertheless be introduced at this point. [Emphasis added], (see Chicha, 2008, p. 70).

As is evident from the quotation, the tool designer presupposes that the weights – the numerical values – shall represent the DMs' assessments of the relative importance of the compensable factors. The weighting method that is be recommended is a so-called *direct rating* method. This means that the DMs shall directly assess the relative importance of the factors without any support of more advanced weighting methods developed, for example, within multi-attribute value theory. This methodological decision appears somewhat surprising since the tool designer emphasizes that weighting has an extremely important impact on the outcome of a job evaluation. Further, the tool designer claims that there is an obvious risk that inconsistent and biased weights are introduced at this point of the evaluation procedure. However, the tool designer does not state any definition of a biased weight. Here, I suppose that the tool designer defines biased weights as weights that the DMs would, after a rational reflection, reject as not being reasonable for the pay setting at the work place.[9]

In order to achieve an unbiased and well-founded weighting, it is essential that the DMs understand what kind of partial decisions the weights will represent. The weighting method must therefore explain the meaning of the weights such that the DMs understand what kind of decisions to which they will commit themselves when they assign weights to the scoring scales. A

[9] This is also the definition of biased weighs used in multi-attribute value theory.

basic methodological issue is therefore if the direct rating method recommended in the ILO tool gives a correct explanation for the meaning of the weights. This question constitutes the starting point for the evaluation in the report where the four questions posed in the introduction will be discussed. *First*, what is the meaning of the weights in job evaluations when an additive value model is applied? *Second*, is the direct rating method based on a correct interpretation of the meaning of weights? *Third*, to what kind of problems can direct rating give rise? *Fourth*, what basic requirements should a weighting method in a job evaluation fulfill when additive value models are applied?

In order to answer these questions, I will use a simplified example of a job evaluation. The purpose is to demonstrate the meaning of weights in an additive value model in a transparent way. The drawback of simple examples is that many important and intricate issues concerning the weighting procedure cannot be discussed.[10] Nevertheless, I think that most of the central problems with direct rating can be elucidated.

In the example, I assume that DMs shall evaluate four jobs *A, B, C* and *D* at a work place based on the compensable factors *required education* and *required job experience*, respectively. The purpose of the evaluation is to classify the jobs in pay grades. In the first stage of the evaluation procedure, the DMs assessed that the requirement profiles of the jobs are as follows:

- Job *A* requires a *master degree* and *one year* of job experience.
- Job *B* requires a *bachelor degree* and *two years* of job experience.
- Job *C* requires a *high-school diploma* and *five years* of job experience.
- Job *D* requires a *bachelor degree* and *one year* of job experience.

Then, the overall value of each job can be represented by an additive value model as follows:

$$V(A) = w_1 V_1(\text{"Master degree"}) + w_2 V_2(\text{"One year of job experience"})$$

$$V(B) = w_1 V_1(\text{"Bachelor degree"}) + w_2 V_2(\text{"Three years of job experience"})$$

$$V(C) = w_1 V_1(\text{"High school diploma "}) + w_2 V_2(\text{"Five years of job experience"})$$

[10] Keenye (2002) contains an extensive discussion of common mistakes in weighting procedures.

$$V(D) = w_1 V_1(\text{"Bachelor degree "}) + w_2 V_2(\text{"One year of job experience"})$$

In the second stage, the DMs have to construct scoring scales and assign weights to the scoring scales. First, DMs construct scoring scales that represent or should at least represent the DMs' evaluation of the jobs regarding each factor.[11] The results are shown in Table 1.

Table 1. *Scoring of the requirement profiles*

Education	**Jobs**	**Score 1**	**Job Experience**	**Jobs**	**Score 2**
Master degree	A	3	Five years	C	3
Bachelor degree	B, D	2	Two years	B	2
High school diploma	C	1	One Year	A, D	1

Second, based on direct rating, we assume that the DMs assessed that the factors are of equal importance and therefore the DMs assigned equal weights to the scoring scales, i.e. $w_1 = w_2 = 1$. It is easy to see that the result of the evaluation is that jobs *A, B* and *C* receive the same total score, i.e. $V(A) = V(B) = V(C) = 4,$ and should therefore be classified in the same pay grade, whereas job *D* receives a lower total score. Based on this example, I will discuss and answer the four questions posed in the evaluation.

3.1 What is the meaning of the weights in job evaluation?

As discussed in section 2, the weights will represent a certain kind of partial decisions in the job evaluation procedure. A key question for a job evaluation is how these partial decisions should be interpreted. Using the above example, these partial decisions can be explained as follows. A consequence of assigning equal weights, i.e. $w_1 = w_2 = 1,$ to the scoring scales in Table 1 is for example that jobs *A* and *C* receive the same total score, i.e. $V(A) = V(C) = 4,$ since the additive model implies that (see page 14):

(1) $V_1(\text{"Master degree"}) + V_2(\text{"One year of job exp."}) =$
$V_1(\text{"High-school diploma"}) + V_2(\text{"Five years of job exp."}) = 4.$

[11] It is essential to note that scoring scales have to be in the form of interval scales, i.e. permissible transformation is defined as: $v_i^* = \alpha_i v_i + \beta_i,$ where $\alpha_i > 0$ and $i = 1,2.$ Naturally, this is a demanding requirement. But it is well-known that weaker scales such as ordinal scales mean that additive value models would give rise to meaningless representations of the overall value of the jobs (see e.g. French, 1993).

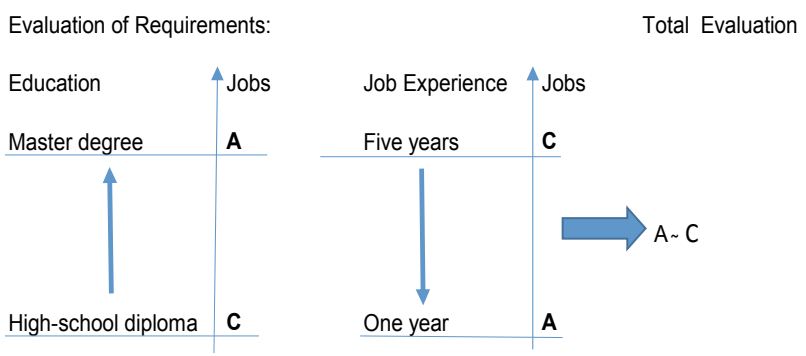
By a simple algebraic manipulation, expression (1) can be expressed as:

(2) $V_1(\text{"Master degree"}) - V_1(\text{"High-school diploma"}) =$
$V_2(\text{"Five years of job exp."}) - V_2(\text{"One year of job exp."})$.

Expression (2) makes it obvious that one meaning of equal weights in the example is to balance the *difference* between jobs *A* and *C* regarding the requirement of education against the *difference* between job *C* and job *A* regarding the requirement for job experience such that both jobs are of *overall equal value* and should therefore be classified in the same pay grade.

Another way of interpreting expression (2) is to say that the weights determine a certain *compensatory relation* between the factors such that the difference between the requirement for a master degree and a high-school diploma is *compensated for* by the difference between the requirement for five years and one year of job experience. This compensatory relation is also illustrated in Figure 1.

Figure 1: A compensatory relation between Education and Job Experience



Note: ~ means "is of equal value as"

The normative consequence of this weighting decision is that jobs *A* and *C* will be classified in the same pay grade. If the DMs were to find it unreasonable to place jobs *A* and *C* in the same pay grade, it means that the DMs have to reject this weighting decision. But such an insight requires that the DMs understand that the weights actually determine certain compensatory relations between the factors.

Assume that the DMs come to know the correct meaning of the weights. Because of this insight, the DMs might, after reflection, find that the

consequences of this weighting are not consistent with the pay setting principles adopted at the work place. The DMs might find it more reasonable to assign the weights $w_1 = 2$ and $w_2 = 1$, since this weighting implies the compensatory relation that a difference of four years of job experience now *only* compensates for the difference between a bachelor degree and a high-school diploma. This compensatory relation is easily derived as follows. Using the scores in Table 1, the weighting implies that:

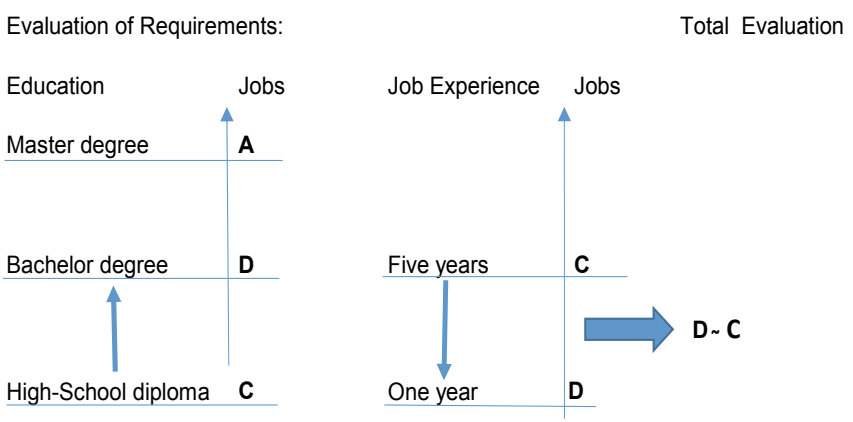
$$(3)\quad 2 \cdot V_1(\text{"Bachelor degree"}) + V_2(\text{"One year of job exp."}) = 2 \cdot V_1(\text{"High-school diploma"}) + V_2(\text{"Five years of job exp."}) = 5.$$

A simple algebraic manipulation implies that:

$$(4)\quad 2 \cdot \left[V_1(\text{"Bachelor degree"}) - V_1(\text{"High-school diploma}\right] = V_2(\text{"Five years of job exp."}) - V_2(\text{"One year of job exp."}).$$

Expression (4) represents the compensatory relation that the difference between a bachelor degree and a high-school diploma is compensated for by a difference of four years of job experience. This compensatory relation is also illustrated in Figure 2.

Figure 2: A compensatory relation between Education and Job Experience



Note: ~ means "is of equal value as"

The consequence of adjusting the weights is that job *A* receives a higher total score than job *C*. But jobs *D* and *C* now receive the same total score and will therefore be placed in the same pay grade. The difference between the two weighting decisions is that in the last weighting, the difference of four years of job experience has a relatively weaker impact on the overall value than in the first weighting. In other words, in the first weighting, the difference of four years of job experience compensates for the difference between a master degree and a high-school diploma, whereas in the second weighting, the same difference in job experience only compensates for the difference between a bachelor degree and a high-school diploma. The two weighting decisions can be seen as an application of two different pay setting principles.

The compensatory relations stated in expressions (2) and (4) can be stated in more general terms for job evaluations based on two factor such as:[12]

(5) $w_1\Delta V_1 = w_2\Delta V_2$

Expression (5) shows that the weights balance the two differences, i.e. the difference ΔV_1 is compensated for by the difference ΔV_2. The compensatory relation implies that if the differences are represented by two jobs *A* and *B* such that

$\Delta V_1 = V_1("A") - V_1("B") > 0$ and $\Delta V_2 = V_2("B") - V_2("A") > 0,$

means that both jobs are of equal overall value, i.e.

$V("A") = V("B")$. (see also Appendix A2-3).

The conclusion of the discussion is that the essential meaning of the weights in an additive value model is to represent the DMs' compensatory weighting decisions. This gives rise to the validity requirement that the weighting methods must support the DMs such that they come to know that a weighting decision will determine certain compensatory relations between

[12] A definition of the compensatory relations for an arbitrary number of factors is straightforward:
$w_i\Delta V_i = w_j\Delta V_j$, where $V_k = V_l$ for all factors beside factor *i* and *j*,
i.e the scale values are assumed to be equal for all factors beside factor *i* and *j*.

the factors. These compensatory relations will have an important impact on the result of a job evaluation and on the adherent pay grade classification of the jobs.[13]

The answer to the first question is: The meaning of the weights, when additive value models are applied, is to represent the DMs' compensatory weighting decisions.[14] In order to achieve unbiased and well-founded weighting, it is essential that the DMs understand this meaning of the weights.

3.2 Is direct rating based on a correct interpretation of weights?

The crucial question is whether the direct rating method in the ILO tool is based on this interpretation of the meaning of the weights. In the ILO tool, however, nothing is mentioned about the weights determining certain compensatory relations between the factors. It is said that the weights shall represent the relative importance of the factors, but without any further explanation of what is meant by relative importance (see the quotation in the introduction of section 3). This means that the direct rating method in the ILO tool will not support the DMs such that they can in a correct way assess to what extent their weighting decision is reasonable. Such an assessment presupposes that the DMs understand that their weighting decisions determine certain compensatory relations between the factors. The direct rating in the ILO tool therefore is based on an incorrect interpretation of the meaning of the weights. This should be regarded as a serious defect of a weighting method, in particular since the tool designer assumes that the weighting has an important impact on the outcome of a job evaluation and that there is an obvious risk for biased and inconsistent weighting at this point of the evaluation procedure.

Another fundamental problem for the direct rating method in the ILO tool is that DMs might not understand the essential point of a job evaluation and adherent weighting decisions. It is important to understand that job evaluation makes it possible through the weighting procedure to compare jobs with very different requirement profiles in such a way that a rational and gender-neutral pay grade classification might be achieved. To

[13] Besides the weighting decisions, the DMs' decision how to score each compensable factor does, of course, also have an impact on the pay grading of the jobs.

[14] The well-known labor economist Mark Killingsworth seems to make the same interpretation of the weights since he claims that "comparable worth amounts to nothing more radical than insisting that the economic theory of *compensating wage differentials* be taken seriously [emphasis added] (see Killingsworth 1987, p. 728).

understand this fundamental meaning of the weights, we can ask the basic question of why it is necessary to assign weights to the scoring scales in the first place. The answer is that the weights solve *value conflicts* that arise in the evaluation of the jobs regarding various compensable factors. If the DMs were not to take weighting decisions, it might be the case that many of the jobs are *not comparable*. If two jobs are not comparable, they are neither of equal value nor of unequal value. Such an outcome makes the pay grade classification very problematic. This problem can be illustrated by the above example. The comparison of jobs *A* and *C* regarding the requirement for education supports *in isolation* that job *A* should be classified in a higher pay grade than job *C*, and vice versa for the outcome of the comparison regarding the requirement for job experience. In the example, the DMs solve this value conflict by assigning equal weights to the scoring scales, but the value conflict is, of course, solved even when unequal weights are assigned to the scoring scales.[15] Solving the value conflict means that the DMs have to assign weights to the scoring scales such that a difference in one compensable factor is balanced against a difference in another compensable factor. The direct rating method does obviously not support this fundamental meaning of the weighting in job evaluations, which makes it possible to compare very different kinds of jobs. Note also that the so-called comparable worth policy is based on the assumption that very different kinds of jobs can be compared in such a way that rational and gender-neutral pay grade classifications can be achieved. But the validity of this assumption presupposes a correct interpretation of the meaning of the weights, which does not seem to be fulfilled in the direct rating method in the ILO tool.

The answer to this second question of the evaluation is:
The direct rating method stated in the ILO tool is based on a incorrect interpretation of the meaning of weights in additive value models. The method does not support the DMs such that they can take well-founded weighting decisions.

[15] It should be noted that assigning equal weights to the scoring scales is not equivalent to un-weighted scoring scales. Un-weighted scoring scales mean that DMs have not taken any weighting decisions that solve the value conflicts, whereas assigning equal weights to the scoring scales means that the DMs have taken a specific decision that solves the value conflict.

3.3. What kind of problems can direct rating give rise to?

I will comment on three problems that might arise when direct rating is used. *First*, direct rating will probably give rise to biased weighting since the DMs will probably not understand how to assess whether the outcome of the weighting is reasonable. Such an assessment presupposes that they understand that their weighting decision determines certain compensatory relations between the factors.

The origin of biased weighting can be elucidated by repeating the discussion in section 3.1 about the two different weighting decisions. In the first weighting, it is assumed the DMs used the direct rating method. The DMs decided that both factors are of equal importance and therefore assigned equal weights, i.e. $w_1 = 1$ and $w_2 = 1$, to the scores. Afterwards, we assume that the DMs come to know the correct meaning of the weights. After reflection, the DMs assess that the weighting is not reasonable. The DMs' reason is that differences in the requirement for job experience will have too strong a compensatory influence on the pay grade classification as compared to the differences in the requirement of education. Therefore, the DMs adjusts the weights to $w_1 = 2$ and $w_2 = 1$, which decreases the compensatory influence of differences in the requirement for job experience on the overall evaluation compared to the influence of differences in education (see figures 1 and 2). The first weighting based on direct rating is therefore biased since the DMs would be willing to adjust their weights based on an adequate compensatory weighting decision.

Assessing to what extent direct rating actually gives rise to biased weighting requires empirical studies of the DMs' weighting behavior. I will comment on this issue at the end of the last section. However, even if no biased weighting were to arise when direct rating is used, it does nevertheless mean that the DMs do not understand the meaning of the weights. This means that direct rating gives rise to a conceptual mistake concerning the meaning of weights in additive value models. In turn, this means that the DMs' weighting decision will probably be based on irrelevant reasons.

The *second* problem is that direct rating seems to be based on the tacit assumption that the weighting is independent of the scoring scales, which are used in the job evaluation situation. This assumption seems to be based on the idea that the weights are supposed to represent the relative importance of the factors, which is assumed not to be related to what scoring scales are used in the job evaluation situation. But this is a very problematic assumption as will be shown. It is easy to show that weights in an additive value model are scaling constants from a mathematical point of

view, which means that the weights have to be adjusted if some of the scoring scales are transformed. If, for example, we start from the scoring scales in Table 1 and the weights $w_1 = 2$ and $w_2 = 1$, and make a permissible transformation of the scores 1, 2 and 3 for the first factor to the scores 2, 4 and 6,[16] it is easy to see that we have to adjust the weights to $w_1 = 1$ and $w_2 = 1$. If the weights are not adjusted, this does, for example, mean that jobs *D* and *C* now receive the total scores $V(D) = 9$ and $V(C) = 7$, respectively. Thus, if the weights are not adjusted means that an arbitrary and permissible transformation of the scores will change the outcome of the evaluation, something which is obviously irrational.

This example can also be used to show that the assumption in direct rating that the *relative importance of the factors* can be represented by the weights in an additive value model is *not* valid. If we start from the weighting $w_1 = 2$ and $w_2 = 1$ implies that the first factor is more important than the second factor according to direct rating. But if we transform the scores for the first factor from 1, 2 and 3 to the scores 2, 4 and 6 we must, as explained above, adjust the weights to $w_1 = 1$ and $w_2 = 1$. But, according to the interpretation of weights in direct rating, this would imply that both factors are now of equal importance. This means that the relative importance of the factors has changed due to an arbitrary and permissible transformation of the first scoring scale, something which does not make sense. Or, in other words, the weights in an additive value model give rise to a meaningless representation of the notion relative importance of the factors (see Appendix A5).

The *third* problem, which is related to the second problem, is that when direct rating is used the DMs might not understand that the weights have to be adjusted if the definitions of any of the scoring scales are changed. A certain definition of a scoring scale for a factor here means that each score refers to a certain requirement level in the factor. This problem can be elucidated as follows. Let us assume that equal weights, assigned to the scoring scales in Table 1, happened to be a reasonable weighting decision. Assume now that in the next job evaluation situation at the same work place, the requirement for job experience has been drastically decreased as shown in Table 2.

[16] Formally stated, the chosen transformation is: $v^*_1 = 2v_1$.

Table 2. A simple job evaluation situation

Education	**Jobs**	**Score 1**	**Job Experience**	**Jobs**	**Score 2**
Master degree	A	3	One year	C	3
Bachelor degree	B, D	2	Six months	B	2
High school diploma	C	1	Nil year	A, D	1

The definition of the scoring scale for the factor requirement of job experience has therefore changed. The highest score (equal to 3) now refers to a requirement of only one year and the lowest score (equal to 1) refers to nil year of job experience. If the weights are not adjusted due to this change of the definition of the scoring scale, it means that a difference of only one year of job experience will now compensate for a difference between the requirement for a master degree and a high-school diploma, such that jobs *A* and *C* are classified in the same pay grade. Expressed in terms of scores, the compensatory relation is:

$$V_1(\text{"Master degree"}) - V_1(\text{"High-school degree"}) = V_2(\text{"One year of job exp."}) - V_2(\text{"Nil year of job exp."})$$

If the DMs were to correctly interpret the meaning of the weights, they would probably find that not adjusting the weights due to a change in the definition of the scoring scale would give rise to an unreasonable and biased weighting. The DMs would probably assign a relatively low weight to the scoring scale for requirement for job experience. But it is important to note that such an understanding of the need for an adjustment of the weights cannot be expected of DMs when their weighting decisions are only based on direct rating methods.

The answer to the third question is: The direct rating method might give rise to biased weights in the first place, because the DMs will not have a correct understanding of the meaning of the weights and, second, the DMs will not understand that the weighting is dependent on how the scoring scales are defined in the specific job evaluation situation. Further, the assumption in direct rating that the weights will represent the relative importance of the job-related factors is not valid.

3.4 What requirement should a weighting method fulfill when additive value models are applied?

As discussed in section 3.1, the basic requirement that a weighting method should fulfill is to support the DMs such that they can take compensatory weighting decisions in a systematic way (see also Appendix A2-3).

A number of various weighting methods which fulfill this basic requirement have been developed within multi-attribute value theory.[17] Without entering into technical details, these weighting methods are essentially based on two stages. Adjusted to a job evaluation the stages can briefly be described as follows.

Stage 1: Take a compensatory weighting decision

1. The DMs have to identify two jobs *A* and *B* such that:

   $\Delta V_1 = V_1(A) - V_1(B) > 0, \;\; \Delta V_2 = V_2(B) - V_2(A) > 0.$

2. The DMs compensatory weighting decision can be stated as:

   Are the jobs *A* and *B* of equal value, i.e. is it the case that: $V(A) = V(B)$?

3a) If the DMs' answer is *yes* to this questions, it means that the DMs have taken the compensatory weighting decision that the difference $\Delta V_1 = V_1(A) - V_1(B)$ compensates for the difference $\Delta V_2 = V_2(B) - V_2(A)$ such that $V(A) = V(B)$, i.e. the jobs should be placed in the same pay grade.

3b) If the DMs' answer is *no*, they have to repeat the process until they identify two jobs such that the answer is yes to the question in 3a.

[17] Examples of the weighting methods are, as named in the literature, trade off methods, swing-weighting methods, weighting based on regressions and weighting by means of pricing-out (see Belton and Stewart 2001 for an overview of various weighting methods).

Stage 2: *Assign weights to the scoring scales such that the compensatory relation is correctly represented*

When the DMs have answered yes to the question in 3a, they have to assign weights to the scoring scales such that the equality $w_1\left[V_1(A)-V_1(B)\right]=w_2\left[V_2(B)-V_2(A)\right]$ [18] is valid.

Thus, the meaning of weights is to balance the two differences such that the overall value of the two jobs will be equal. In this second stage, it is important the DMs understand how the scoring scales are defined. It should be noted that in order to uniquely determine the weights, it is sufficient that the DMs determine one compensatory relation between the factors as described in stage 3a.[19]

Evaluating whether any of these adequate weighting methods used in multi-attribute value theory would be appropriate for weighting in job evaluation requires both theoretical and empirical studies, something which is beyond the scope of this study. But it should be pointed out that using any of these methods does not mean that the weighting procedure becomes less demanding. It might be quite the opposite. It is well known within multi-attribute value theory that the weighting procedure in multidimensional decision problems is a cognitive demanding procedure. But for that reason, it seems important that DMs have access to adequate weighting methods.

The answer to the fourth question is: The weighing method should fulfill two requirements. *First*, the method should support the DMs such that they can take compensatory weighting decisions in a systematic way. *Second*, the method should support the DMs such that the assigned weights correctly represent the compensatory weighting decisions.

3.5 Further evidence for misinterpretation of the meaning of the weights in the ILO tool

By commenting on three quotations, I will in this final section present further evidence that the ILO tool gives a incorrect interpretation of the weights in additive value models. In the first quotation, it is said that:

[18] From a mathematical point of view, the weighting means that the two scoring scales are coordinated.

[19] In order to check for consistency in the weighting, it is common that the weighting methods support the DMs to base the compensatory weighting decisions on a number of different comparisons.

> For example: *In a company which develops software programs a high weight will be assigned to the analytical skills criterion*. In a day-care center the responsibility for people criterion will be of utmost importance, in a public works enterprise responsibility for equipment will be one of the key factors. [Emphasis added], (Chicha, 2008, p. 72).

But it is incorrect to claim that a large weight should be assigned to the scoring scale representing, for example, the degree of requirement for analytical skills for the only reason that analytical skills are in some sense important for a company which develops software programs. This is incurrect for at least three reasons. In the *first place*, the weights depend on how the scoring scales are defined in the specific job evaluation situation (see section 3.3). Thus, two companies where the analytical skills are in some sense of equal importance should nevertheless assign different weights to the scoring scales, due to the fact that the scoring scales are defined in different ways in the two companies. *Second*, the weights in an additive value model determine a certain compensatory relation between the compensable factors and therefore, there is no obvious relation between the assessed importance of analytical skills for the enterprise and the weights assigned to the scoring scales in the additive value model (see section 3.2). *Third*, weights in an additive value model cannot meaningfully represent statements about the relative importance of factors (see section 3.3). The same incorrect interpretation is evident in the following quotation:

> Consistency can be ensured by examining the weight assigned to each factor being assessed in light of the goals and values of the enterprise. An element which has great importance for the enterprise should not be given low weight and vice versa. (See Chicha 2001, p.70).

The tool designer makes the wrong assumption that there is an obvious relation between the importance of a factor (here also named elements) and the weight assigned to the scoring scale representing the evaluation of the factor.

In the final quotation, it is assumed that it is meaningful to compare weights between different job evaluation situations.

> In general, most experts agree on the following percentage [sic] ranges as approximate guidelines with regard to the relative importance of factors:
>
> 20% to 35% for qualifications

> 25% to 40% for responsibility
> 15% to 25% for effort
> 5% to 15% for working conditions.
> (See Chicha 2008, p. 71).

Agreeing about the weights does not mean that the experts agree on the essential point concerning what compensatory relations between the compensable factors that are reasonable. Naturally, the same weights can represent very different compensatory relations between the factors for the simple reason that the scoring scales can be defined in very different ways in two different job evaluation situations (see section 3.3).

I end the evaluation by speculating about the reasons for not recommending a more advanced weighting method in the ILO tool. One explanation might be that the tool designer does not have any knowledge of multi-attribute value theory. Another explanation might be that the tool designer has assessed that direct rating performs well, and that the outcome of using direct rating will be similar to the outcome when applying compensatory weighting methods. But no such assessments are mentioned in the ILO tool. As far as I know, the performance of direct rating used in job evaluation has not been studied.

However, the performance of direct rating has been studied in other kinds of multidimensional evaluation problems. The unanimous results of these studies are summarized by Belton and Stewart (2001 p. 289) as follows:

> The implication is that directly assessed importance ratios are almost certainly inappropriate for use in a value function model.

They continue with the strong recommendation:

> ... avoid questions which involve the less well-defined notion of "importance" in the abstract, since these may generate highly incorrect results if the intuitive notion of importance and the desired trade-off ratio do not coincide.

Thus, the conclusion seems to be that direct rating based on the notion of importance might give rise to highly incorrect or biased weights.

Therefore, I recommend that the weighting method in the ILO tool should be improved. In particular, it seems important in this case since the purpose of the ILO tool is to serve as a general and worldwide guideline for

gender-neutral job evaluations. Based on empirical as well as theoretical studies by means of multi-attribute value theory, it would be possible to develop adequate weighting methods that are appropriate for gender-neutral job evaluations.

4. Summary

The purpose of the report is to evaluate the validity of a weighting method stated in a gender-neutral job evaluation tool (the ILO tool), which is recommended by the International Labor Office (ILO). The starting point for the evaluation is the basic validity requirement that weighting methods have to be based on a correct interpretation of the meaning of weights in additive value models, which are used as "measures" of the value of jobs. The conclusion of the evaluation is that the weighting method in the ILO tool does not fulfill this basic validity requirement.

The conclusion is based on a discussion of the following questions. *First*, what is the meaning of the weights when additive value models are applied? I show that the meaning of the weights is to determine compensatory relations between job-related factors. The compensatory relations have an important impact on the result of a job evaluation. *Second*, is the direct rating method recommended in the ILO tool based on a correct interpretation of the meaning of the weights? The weights in direct rating are assumed to represent the relative importance of the factors. But there is no evidence in the ILO tool that decisions about the relative importance of the job-related factors should be interpreted as compensatory weighting decisions. Therefore, I conclude that the direct rating method is based on a misinterpretation of the meaning of weights in additive value models. *Third*, to what kind of problems can the direct rating method give rise? Since direct rating is based on a misinterpretation of the weights, the weighting decisions might give rise to biased weights, i.e. weights that would be regarded as unreasonable if the user of the ILO tool correctly interpreted the meaning of the weights. *Fourth*, what basic requirements should the weighting method fulfill? The weighting method should be constructed such that the users are supported to take compensatory weighting decisions in a systematic way. But the direct rating method in the ILO tool does not support the users to take such compensatory weighting decisions.

As far as I know, there are no empirical evaluations of whether and to what extent biased and not well-founded weighting decisions will arise when direct rating is used in job evaluations. However, conclusions of

studies about the performance of direct rating in other types of multi-dimensional decisions contexts strongly indicate that direct rating is not appropriate for weighting when additive value models are applied.

Finally, based on multi-attribute value theory, research efforts should be devoted to develop adequate weighting methods which are appropriate for gender-neutral job evaluation procedures.

References

Armstrong, M., Cummins, A., Hastings, S. and Wood, W. (2003), *Job Evaluation – A Guide To Achieving Equal Pay*, Kogan Page, London.

Belton, V. and Stewart, T.J. (2002), *Multiple criteria decision analysis: an integrated approach,* Kluwer Academic Publishers, Dordrecht.

Chica, M.T. (2008), *Promoting equity: gender-neutral job evaluation for equal pay: a step-by-step guide,* International Labour Office, Geneva.

Corominas A., Coves, A.M., Lusa, A. and Martinez, C. (2008), "ISOS: a job evaluation system to implement comparable worth", *Intangible Capital,* Vol. 4 No. 1, pp. 8-28.

England, P. (1999), "The case for comparable worth", *Quarterly Review of Economics and Finance*, Vol. 39, pp. 743-755.

Fischer, G.W. (1995), "Range sensitivity of attribute weights in multi-attribute value models", *Organisation, Behaviour and Human Decision Processes,* Vol. 62 No. 3, pp. 252-266.

French, S. (1993), *Decision Theory: An Introduction to the Mathematics of Rationality*, Ellis Horwood, London.

Harriman, A. and Holm, C. (2007), *Steps to Pay Equity*, Equal Opportunities Ombudsman, Stockholm.

Keeney, R. L. (2002), "Common Mistakes in Making Value Trade-Offs", Operations Research, Vol. 50, No. 6, pp. 935-945.

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

Killingsworth, M.R. (1987), "Heterogeneous preferences, compensating wage differentials, and comparable worth", *Quarterly Journal of Economics,* Vol. 102 No. 4, pp. 727-742.

Pöyhönen, M. and Hämäläinen, R.P. (2001), "On the convergence of multi-attribute weighting methods", *European Journal of Operational Research,* Vol. 129, pp. 569-585.

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

Von Nitzsch, R. and Weber, M. (1993), "The effect of attribute ranges on weights in multiattribute utility measurements", *Management Science*, Vol. 39, pp. 937-943.

Wakker, P. (1989), *Additive representations of preferences: a new foundation of decision analysis*, Kluwer Academic Publishers, Dordrecht.

Weber, M. and Borcherding, K. (1993), "Behavioural influences on weight judgments in multiattribute decision making", *European Journal of Operational Research*, Vol. 6, pp. 1-12.

Appendix: Additive value models and the meaning of weights in job evaluation

A1: Additive representation of the value of jobs

From a multi-attribute value theoretical point of view, a basic but not explicitly stated assumption in job evaluation is that the overall value of jobs can be represented by an additive value function. If we restrict the representation to two compensable factors, it can be stated as follows:[20]
For each pair of jobs $a, b \in A$, there is an additive value function such that:

$$a \succsim_v b \Leftrightarrow w_1 v_1(a) + w_2 v_2(a) \geq w_1 v_1(b) + w_2 v_2(b),$$

where A = the set of jobs to be evaluated.
$a \succsim_v b$ = Job *a* is at least of equal value as job *b*.
$v_i(\cdot)$ = a partial scoring scale representing the evaluation of jobs regarding factor *i*, where *i* = 1, 2.
w_i = weight (scaling parameter) assigned to the scoring scale $v_i(\cdot)$.

(An additive representation of the value of jobs for an arbitrary number of compensable factors is straightforward.)

A2: The algebraic meaning of weights in additive value models

The algebraic meaning of weights is to coordinate the scoring scales. The coordination can be defined as:

(1a) $\Delta V = w_1(-\Delta v_1) + w_2 \Delta v_2 = 0$, where $-\Delta v_1 < 0$ and $\Delta v_2 > 0$

or as

[20] From a formal point of view, the existence of an additive representation presupposes that the relation $\succsim_v$ satisfies a number of *necessary* and *sufficient* axioms. Further, scales in additive value models have to be in the form of an interval scale, i.e. permissible transformations are defined as: $v_i^*(\cdot) = \alpha_i v_i(\cdot) + \beta_i, \ \alpha_i > 0$. See e.g. French (1993).

(1b) $w_1\Delta v_1 = w_2\Delta v_2$, or as (1c) $\dfrac{\Delta v_1}{\Delta v_2} = \dfrac{w_2}{w_1}$

A3: The meaning and determination of weights in job evaluation

When the additive value model is applied in job evaluation contexts the expressions (1a-c) will determine certain compensatory relations between the compensable factors. The compensatory relation can be determined in two steps:

1) Define a difference in each factor, which can be represented by two (possibly hypothetical) jobs *A* and *B* such that $V(A) = V(B)$ where $\Delta v_1 = v_1(A) - v_1(B) > 0$ and $\Delta v_2 = v_2(B) - v_2(A) > 0$, i.e. the difference in the first factor is compensated for by the difference in the second factor such that $V(A) = V(B)$.

2) Assign weights to the scoring scales such that the equality

   (2a) $w_1\left[v_1(A) - v_1(B)\right] = w_2\left[v_2(B) - v_2(A)\right]$ is valid.

   The equality can be expressed as a trade-off ratio as:

   (2b) $\dfrac{v_1(A) - v_1(B)}{v_2(B) - v_2(A)} = \dfrac{w_2}{w_1}$

A4: Transformation of scoring scales and adjustment of weights

From expression (2b) it becomes obvious that the weights might have to be adjusted if any of the scoring scales are transformed, which is easily seen as follows: Given a tradeoff ratio as

$\dfrac{v_1(A) - v_1(B)}{v_2(B) - v_2(A)} = \dfrac{w_2}{w_1}$, if the scoring scales are transformed as:

$v_1^*(\cdot) = \alpha_1 v_1(\cdot) + \beta_1,\ \alpha_1 > 0$ and $v_2^*(\cdot) = \alpha_2 v_2(\cdot) + \beta_2,\ \alpha_2 > 0,$

implies that the trade-off ratio must be defined as:

$\frac{\alpha_1(v_1(A)-v_1(B))}{\alpha_2(v_2(B)-v_2(A))}=\frac{w_2^*}{w_1^*}$, i.e. the weights are adjusted as $w_1^*=\alpha_2 w_1$ and $w_2^*=\alpha_1 w_2$.

A5: A meaningless interpretation of weights in job evaluation

It is common in additive value models like in the ILO tool that the weights are supposed to represent the DMs' assessments of the relative importance of the factors. Thus, a ratio such as $(w_1/w_2)>1$ is interpreted as a numerical statement representing the assessment that factor *1* is more important than factor *2* in a certain job evaluation situation, i.e.

(3) $\frac{w_1}{w_2}>1 \Leftrightarrow$ factor 1 is more important than factor 2.

But it is easily shown that this numerical statement cannot be used as a meaningful statement about the relative importance of the factors.[21] Since the partial scoring scales are in the form of interval scales, i.e. permissible transformations are defined as:

$v_i^*=\alpha_i v_i+\beta_i,\ \alpha_i>0$ an additive representation of the overall value of job *a* and *b*, i.e.

$$a=_v b \Leftrightarrow w_1v_1(a)+w_2v_2(a)=w_1v_1(b)+w_2v_2(b)$$

can equivalently be stated as:

$$(4)\ a=_v b \Leftrightarrow v_1^*(a)+v_2^*(a)=v_1^*(b)+v_2^*(b),$$

where $v_i^*(\cdot)=w_iv_i(\cdot)$ or $v_i(\cdot)=\frac{1}{w_i}v_i^*(\cdot)$.

In expression (4), the ratio between the weights is equal to one. Thus, the truth value of the numerical statement (w_1/w_2) is *not invariant* for a permissible transformation of the partial scoring scales. In other words, the ratio (w_1/w_2) is not a meaningful numerical statement about the relative importance of the two factors.

[21] Roberts (1979, chap. 2) presents criteria for the meaningfulness of numerical statements.