CSB

WORKING PAPER

centreforsocialpolicy.eu November 2011 No 11 / 10

What if we voted on the weights of a multidimensional well-being index? An illustration with Flemish data

Koen Decancq, Luc Van Ootegem and Elsy Verhofstadt

Universiteit Antwerpen

CSB

University of Antwerp
Herman Deleeck Centre for Social Policy
Sint-Jacobstraat 2
BE – 2000 Antwerp
fax +32 (0)3 265 57 98

What if we voted on the weights of a multidimensional well-being index? An illustration with Flemish data

Koen Decancq[1], **Luc Van Ootegem**[2] and **Elsy Verhofstadt**[3]

Working Paper No. 11 / 10

November 2011

ABSTRACT

There is a widespread consensus that well-being is a multidimensional notion. To quantify multidimensional well-being, information on the relative weights of the different dimensions is essential. There is, however, considerable disagreement in the literature on the most appropriate weighting scheme to be used. Making use of a recent data set for Flanders, we compare various methods to select a weighting scheme. The results are indeed different such that, for instance, a policymaker would identify different groups of individuals as being worst-off depending on the scheme that is chosen. In order to compare and evaluate the weighting schemes, we simulate the support each scheme would get in a hypothetical voting procedure. Weighting schemes that obtain a higher support reflect better the priorities of the respondents themselves and suffer less from the problem of paternalism. Quite remarkably, the popular equal weighting scheme is found to be the least supported in our data set.

Keywords
Multidimensional well-being, weighting schemes, equal weights, voting.

JEL Classification
I31, C43, O1

---

[1] Herman Deleeck Centre for Social Policy, University of Antwerp, Belgium; Center for Economic Studies, KULeuven, Belgium; CORE, Université Catholique de Louvain, Belgium. e-mail: koen.decancq@ua.ac.be

[2] Faculty of Business Administration and Public Administration, University College Ghent, Belgium; Sherppa, Ghent University, Belgium; Higher Institute of Labour Studies (HIVA), KULeuven, Belgium. e-mail: luc.vanootegem@hogent.be

[3] Faculty of Business Administration and Public Administration, University College Ghent, Belgium; Sherppa, Ghent University, Belgium. e-mail: elsy.verhofstadt@hogent.be

We thank participants at the second “New Directions in Welfare Economics” conference in Paris (2011), as well as Luna Bellani, Lucio Esposito, Alessio Fusco, Tim Goedemé, Leen Meeusen, Erik Schokkaert, Nico Pestel, Nicky Rogge and Dirk Van de gaer for their valuable comments. All remaining shortcomings are ours.

1. Introduction

In recent theoretical contributions, a consensus emerged that well-being is a multidimensional notion. Stiglitz, Sen and Fitoussi (2009, p.14) recommend, in their report on the 'measurement of economic performance and social progress', that a multidimensional definition has to be used to define well-being. However, much less consensus has been achieved on *how* one should take this multidimensionality into account when empirically measuring well-being. This paper focuses on one particular aspect of the challenge of measuring multidimensional well-being: how to select the most appropriate weighting scheme.

A weighting scheme attributes a relative weight to each dimension of well-being. The selection of a weighting scheme has an extensive and an intensive part. First, for the extensive weighting part, one decides which dimensions are given a non-zero weight. In other words, a list of relevant dimensions that describe the well-being of an individual has to be selected.[4] We assume that this listing problem has been settled a *priori* to this research (for instance by philosophical enquiry, public debate or simply by data limitations). Table 1 overviews the list of dimensions used in this paper (their abbreviations will be useful for later reference). This list is inspired by the empirical study of Anand and Van Hees (2006) and by the aforementioned report by Stiglitz, Sen and Fitoussi (2009).

Table 1. Dimensions of well-being

Abbreviation	Dimension of well-being
GOA	Reaching goals in life
HEA	Health
EDU	Education
LIV	Material living standard
SOC	Social connections
ENV	Environment
VIS	Personal vision when making decisions
WOR	Work situation
REL	Personal relationships

The current paper is only concerned with the intensive weighting problem of choosing the relative weights which determine the relative importance and trade-offs between different dimensions of well-being. In a linear additive well-being index for example, a higher weight for health increases the amount of outcomes or realisations that are necessary in other dimensions (such as the material living standard) to compensate for a decrease in the health status. A weighting scheme embeds a particular view on the relative importance of the dimensions of well-being and hence on the Aristotelian question about 'the good life'. Given this fundamental importance of the weights, it should not come as a surprise that different

[4] See Hagerty et al. (2001), Costanza et al. (2007) and Rahman (2007) for an overview of different approaches to compound such a list of dimensions.

researchers have suggested different procedures to construct and to select the most appropriate weighting schemes.

In fact, a plethora of methods exists to select a weighting scheme. Decancq and Lugo (2013) provide an overview of the literature and classify the different approaches in three broad categories. Data-driven weighting schemes are simply and solely based on the outcomes of the individuals in the different dimensions. Normative weighting schemes, on the other hand, are only based on the prevailing opinions in the society on the relative importance of the dimensions of well-being (expressed through democratic deliberation, expert opinions, or by individual respondents themselves ...). Hybrid weights depend on both the outcomes and the opinions. The most popular weighting scheme is the equal weighting scheme in which the relevant dimensions are weighted equally. Obviously, this is an arbitrary choice.

Not only researchers have a difference of opinion about the most appropriate weighting scheme. One can indeed think of numerous real life examples that reflect the different opinions that people have on the relative importance of the dimensions of life. By using individual-specific weighting schemes, one can incorporate these different opinions in the measurement of multidimensional well-being and avoid that the analysis becomes overly paternalistic or perfectionist. Unfortunately, recent contributions in social choice theory have shown that allowing for heterogeneity in opinions comes at a high price: an approach based on individual-specific weights cannot satisfy the so-called dominance principle. According to the dominance principle, someone who is better-off in all dimensions of life should have a higher overall well-being than someone who is worse-off in all dimensions.[5] The impossibility to combine both requirements (the avoidance of paternalism and the compliance with the dominance principle) brings us on a cross-road. Either one avoids paternalism and sacrifices the dominance principle or one starts from the dominance principle and some kind of paternalism has to be allowed. Both, avoidance of paternalism and compliance with the dominance principle, are attractive normative principles when measuring human well-being. A priori, it is hard to say which is the most attractive route. We have explored the former route in earlier work (Fleurbaey et al. 2009 and Schokkaert et al. 2011). Here we take the latter one by assuming a common (and hence paternalistic) weighting scheme which applies uniformly to all individuals.

[5] This result follows from a more general result shown by Brun and Tungodden (2004) and Fleurbaey and Trannoy (2003). To see the impossibility, think of a situation with two individuals that are on two crossing iso-well-being curves (the different slopes are the result of having different weighting schemes), such that one individual has more in all dimensions and is thus considered better-off according to the dominance principle. By moving along their iso-well-being curves to the other side of the crossing, both individuals are now in a new situation that they themselves consider to be equally well-off, but it may occur that the initially worse-off individual has now more in all dimensions than the initially better-off individual, which goes against the dominance principle.

In this paper, we investigate how much support different common weighting schemes would get from the respondents themselves. In other words, we quantify the empirical severity of the problem of paternalism for various weighting schemes. To do so, we perform a thought experiment. We simulate the outcome of a voting procedure based on the opinions of the respondents on the relative importance of the dimensions. We use a Kemeny-Young or maximum likelihood voting procedure in which individuals give more support to a weighting scheme if it ranks more pairs of dimensions in accordance to their own opinions on the relative importance of the dimensions (Kemeny 1959; Young and Levenglick 1978; Young 1995). We believe that such a pairwise procedure is attractive in this context and furthermore it allows accommodating for the ordinal and interpersonally non-comparable nature of the information on the opinions.

The paper is structured as follows. In section two, we introduce the data set which is particularly designed for multidimensional well-being analysis. It contains simultaneously information on the outcomes in a wide set of dimensions of well-being, opinions on the importance of these dimensions and some auxiliary variables necessary to compute the weights. In section three, we present a conceptual framework for the construction of a well-being index with particular attention for the different methods to select a weighting scheme. We report the resulting weighting schemes for our data set and illustrate the implications of using different weighting schemes on well-being and on the identification of the worst-off. Section four introduces the hypothetical voting procedure and presents the simulated support for the different weighting systems. Section five concludes and hints at further research.

2. The data set: LEVO 2010

The empirical analysis in this paper is based on LEVO 2010, a survey conducted in Flanders (the northern part of Belgium).[6] The data set is collected through quota sampling and contains self-reported information for five different sub-samples of the population. The sub-samples are stratified according to their life situation in one of the following categories: students, part-time workers, full-time workers, retired people, unemployed and other non-employed people. Each sub-sample contains about 250 respondents. The different sub-samples are then statistically weighted to obtain a sample which is representative for the Flemish population according to the life situation, gender and age distribution. This does not provide a perfectly representative sample for the overall Flemish population so that we do not intend to make concrete or precise statements about well-being in Flanders as such. Our aim is rather to

[6] LEVO is the Dutch acronym for “LEvensomstandigheden in Vlaanderen Onderzocht” (research on living circumstances in Flanders). It is a yearly large-scale (N=1552) survey organised in the framework of master-papers and a research seminar at the University College Ghent. The field work is carried out by students of the University College Ghent. Organisation, supervision, controlling and cleaning is performed by the (Ghent affiliated) authors.

compute various weighting schemes and to compare them according to the support they get in a hypothetical voting procedure, indicating which weighting schemes suffers most from paternalism. For this purpose, the data in LEVO 2010 are particularly useful. The data allow for a broad description of human well-being and contain detailed information on three groups of variables that are necessary for our analysis: the *outcomes* of the respondents in the dimensions of life as listed in Table 1 as well as the *opinions* of these respondents on the importance of these dimensions and some important *auxiliary* variables (socio-demographic variables and personality traits) necessary to compute some of the weighting schemes.

The outcome variables measure the realizations on the nine dimensions of well-being on a 7-point Likert scale ranging from ‘complete disagreement’ to ‘complete agreement’. The second column of Table 2 lists the actual questions translated from Dutch. The 10^{th} percentile value, median and 90^{th} percentile value of the outcomes in each dimension is given in column three of Table 2. In general, the respondents report high scores on the outcomes. The biggest dispersion can be found in the dimensions education, health and living standards.

For the opinion variables, the respondent indicates for each of the dimensions, how important she considers this dimension to be on a 10-point-scale ranging from ‘completely unimportant’ to ‘very important’. Every dimension is addressed separately such that attributing equal importance to several dimensions is possible. In the survey, all respondents answer first the question on the opinions and then the question on the outcomes. Column four of Table 2 provides some summary statistics. There is more ordinal variation in the valuation of the importance of the dimensions than in the actual outcomes. The respondents consider health to be the most important dimension of well-being and education the least important one (note the large dispersion in the opinions on education however). This is a finding which is in line with the literature on health (see Rahman 2007) and on education (Dolan et al. 2008), but obviously it depends on the exact framing of the questions. For the income dimension the rank correlation coefficient between opinions and outcomes is not significant, for all other dimensions there is a significant but small positive correlation: ranging from 0.112 for the dimension ‘goals’ to 0.371 for the dimension ‘relationship’.

Table 2. Dimensions of well-being: summary statistics of outcomes and opinions on importance

	Statements on well-being	**Outcomes**			**Importance**		
		p10	**p50**	**p90**	**p10**	**p50**	**p90**
GOA	I am, given my age, satisfied with the goals I have reached	4	6	7	4	7	9
HEA	I consider myself in good health	3	6	7	7	9	10
EDU	I study (studied) or follow(ed) courses according to my wishes	2	5	7	2	6	9
LIV	My (household) income is sufficient to live well	3	6	7	5	7	9
SOC	I have a satisfying social life (friend, leisure, ...)	4	6	7	5	8	9
ENV	I live and spent my life in pleasant environments (house, work, environment)	4	6	7	5	8	9
VIS	I act according to my personal vision on life when making decisions	4	6	7	5	7	9
SIT	I am satisfied with my actual situation (work/study/retirement)	4	6	7	5	7	9
REL	I am satisfied with my relationship	4	6	7	5	8	10

Source: LEVO 2010

Various auxiliary variables are included in the survey such as socio-demographic characteristics (gender, nationality and age) and four personality traits which are constructed by a factor analysis based on eleven underlying personality questions. Particularly useful for the computation of the weighting schemes in the next section are the following two questions. First, the respondents are asked about their general satisfaction in life: "How is your satisfaction with life in general". The second question is a capability indicator resulting from the question "How do you consider your possibilities/opportunities in life in general?". The first question is answered on a 10-point scale (from 'very unsatisfied' to 'very satisfied'), the second question on a 7-point scale ranging from 'totally insufficient' to 'excellent'.

3. How to measure multidimensional well-being?

3.1. A conceptual framework to measure multidimensional well-being

Once the list of relevant dimensions is identified, the construction of an index of multidimensional well-being raises three questions (see also Decancq et al. 2009 and Decancq and Lugo 2013). First, one has to decide on an appropriate standardization of the dimensions. Second, one has to select a procedure to aggregate across dimensions and finally, a weighting scheme has to be selected. In this paper, the focus is on the latter issue, but the three questions are intimately linked and cannot be discussed nor solved in isolation. A measure of well-being depends on particular choices for each of these questions. We do not think that the choices made in this paper are the only possible ones, but we think that they present a reasonable trade-off between parsimony and flexibility for the problem at hand.

The outcomes are measured on a 7-point Likert scale. So, the standardization plays an important role in translating the verbal categories into cardinal values that can be sensibly aggregated. We choose to connect to each verbal category a value between 0 and 1 where 0 reflects the lowest outcome in that dimension and 1 the highest, assuming furthermore equal spacing between the verbal categories. In practice, this boils down to attributing the value 0 to ‘complete disagreement’, 1 to ‘complete agreement’ and the corresponding rational number to the intermediate verbal categories. Choosing for an alternative way to connect cardinal values to the verbal categories (for instance, based on the position of the respondent in the total sample) leads to different quantitative results in what follows, but sensitivity analysis has shown that the qualitative results remain.

In the second step, an appropriate aggregation procedure is needed to aggregate the standardized outcomes. The majority of the existing empirical approaches uses a linear aggregation, such that the well-being index becomes the weighted average of the standardized outcomes. Although this is a convenient procedure, its theoretical justification is rather weak (See Decancq and Lugo 2013 for some alternatives). Using this aggregation procedure, a perfect substitution is assumed between the dimensions of well-being. A decrease in one dimension can then be compensated with an increase of the same magnitude in any other dimension. Given the focus of the paper however, we adhere to the common practice of using a linear aggregation.

Once the weighted average has been selected as an aggregation procedure for the well-being indicator, the final choice is about the most appropriate weighting scheme. Though it is not necessary, one typically normalizes the weights so that they sum to one. We do the same here. Different approaches to select a weighting scheme have been proposed in the literature, ranging from strikingly simple to very complicated ones (Decancq and Lugo 2013).

3.2. Weighting schemes

In this section, we describe the different methodologies to set the weights and we apply them to the LEVO 2010 data set. It is useful to distinguish three broad categories of approaches: the data-driven approaches, the normative and hybrid ones.

Table 3. Resulting weighting schemes

	Data-driven			Normative		Hybrid			
	FREQ	**PRINC**	**MOFA**	**EQUA**	**UNDP**	**SAT(1)**	**CAP(1)**	**SAT(2)**	**CAP(2)**
GOA	0.1135	0.1290	0.0500	0.1111	0	0.2144	0.2030	0.2348	0.2224
HEA	0.1067	0.1078	0.0500	0.1111	0.3333	0.0713	0.1621	0	0.1307
EDU	0.1020	0.0869	0.0500	0.1111	0.3333	0	0	0	0
LIV	0.1085	**0.1180**	**0.0500**	**0.1111**	**0.3333**	**0**	**0.1033**	0	0.0899
SOC	0.1147	0.1152	0.0500	0.1111	0	0.1385	0	0.1008	0
ENV	0.1163	0.1261	0.0500	0.1111	0	0.1489	0.1243	0.1728	0.1350
VIS	0.1178	0.1148	0.0500	0.1111	0	0	0.2047	0	0.1783
WOR	0.1091	0.1266	0.0500	0.1111	0	0.2163	0.1069	0.2581	0.1405
REL	0.1114	0.0756	0.6000	0.1111	0	0.2105	0.0957	0.2335	0.1083

Source: LEVO 2010, own computations

Data-driven approaches use only information on the outcomes of the respondents. A first possibility is the so-called *frequency-based* weighting scheme (FREQ). In the literature on multiple deprivation measurement, it is often argued that there should be an inverse relation between the frequency of deprivation in a dimension and the weight of that dimension (Deutsch and Silber 2005). The more deprivation there is in a dimension, the more it is widespread in a society and the lower the weight the dimension receives. For our data set we obtain the number of deprived individuals by counting the ones scoring less or equal than 0.33 for a given dimension. This numerical value is obviously an arbitrary choice (our sensitivity analysis has shown that qualitative results remain for other choices). Afterwards these numbers are normalized so that the resulting weights sum to one Table 3 (second column) gives the resulting weighting schemes for the LEVO data set. We see that education, the dimension with the highest number of deprived individuals, gets the lowest weight.

A second data-driven approach uses statistical algorithms to determine the weights. The most commonly used approach uses the normalized loadings of the first *principal component* between the standardized outcomes as weights (PRINC). The weighting scheme is chosen to maximize the variance of the resulting well-being index. A recent example of this approach is by Nguefack-Tsague et al. (2011) who use principal components to justify statistically the equal weighting scheme selected by the Human Development Index (HDI). The HDI aggregates the country-level material living standard, education and health into one overall index of human development. The third column of Table 3 gives the principal component weights. The principal component approach attributes the lowest weight to personal relationships and the highest weight to the ability to reach goals in life.

The *most favourable weights* (MOFA) are a third possibility within the category of the data-driven approaches. This approach selects those weights that are most advantageous for computing the resulting well-being scores of the respondents. This is done by giving more weight to dimensions in which respondents score relatively well and less weight to dimensions in which they score less. In practice, these well-being maximizing weights are obtained by solving a linear programming problem. To avoid that all weight goes to one dimension, exogenous restrictions can be imposed (See Cherchye et al. 2007 for an overview). We select 0.05 as an exogenous lower bound. Generally, this approach leads to individual-specific weights. Here the focus is on common weighting schemes. Therefore, we use a variant based on Cherchye and Kuosmanen (2006): weights are selected that maximize the well-being score of a representative respondent with average standardized outcomes. As can be seen from Table 3, the most favourable weighting scheme attributes maximal weight to the dimension of personal relationships and the exogenous lower bound to all the other dimensions.

Completely opposite to data-driven approaches, normative approaches select the weighting scheme independent of the observed outcomes in the data set at hand. The weights reflect only the opinions on the 'good life' or the preferred trade-offs between the dimensions. The *equal weighting scheme* (EQUA) gives equal weights to all the dimensions. The usual arguments in favour of equal weighting schemes are its simplicity and the principle of "insufficient reason" to do otherwise. Chowdhury and Squire (2006, p. 762) consider equal weighting as an obviously convenient procedure, but add that it is universally considered to be a wrong one. As an alternative, we consider a particular weighting scheme which has been advocated in various publications of the *United Nations Development Program* (UNDP), for instance in their flagship indicator, the HDI, where the focus is exclusively on material living standards, health and education. The UNDP advocates this weighting scheme in the context of country-level outcomes, but one could also apply it to micro data (for instance as in the Multidimensional Poverty Index (MPI)) by further restricting the set of relevant dimensions of life. Column five and six of Table 3 present the selected normative (equal) weights.

Hybrid approaches make use of information on the outcomes of the respondents as well as their opinions on the 'good life'. To elicit the views of the respondents on the good life we use a *regression-based* approach. We make use of four different regressions. The first and second version, SAT(1) and CAP(1), use as dependent variable respectively the general satisfaction question and capability indicator as described in section two. The explanatory variables of the regression models consist of the outcomes of the respondents in the various dimensions of life. Socio-demographic variables can potentially affect value judgments on the importance of different dimensions. Since the aim is to obtain scores that are laundered from this effect (Schokkaert, 2007), we include in the third and fourth model, SAT(2) and CAP(2), control variables such as sex, age, age squared and nationality as well as four personality factors. The weights are derived from the regression coefficients by normalization such that they sum to one, provided they are non-negative and significant at the 5% level. Non-significant or negative weights are set to zero. The weighting schemes resulting from the regression based approach are presented in column seven to ten of Table 3. As can be seen, all regression based approaches attribute a zero-weight to education and give the highest weight to the ability to reach goals in life. From a satisfaction perspective, work and relationship are attributed a large weight. From a capability perspective, living according to one's vision gets a large weight. This confirms the findings of Van Ootegem and Verhofstadt (2011) that satisfaction and capabilities are influenced by different variables. Overall, the different approaches in this section lead to quite different weighting schemes.

3.3. Implications for the identification of the worst-off

To get an idea about the magnitude of the difference in the weights, we illustrate its potential impact on the targeting of social policy. Imagine a policy maker who wants to take the multidimensional nature of well-being seriously while targeting social policy towards the worst-off individuals in society. Using the conceptual framework and the different weights as described in this section, he can easily compute the well-being scores for all the individuals. Furthermore, assume that the policy maker identifies the worst-off individuals by using a cut-off of 60% of the median well-being score (as is the standard practice for the income dimension in EU-policy circles).

For the data at hand, this leads to 13.7% of the population which is identified as worst-off under one of the nine weighting schemes. Only 1.2% of the population is identified as worst-off irrespective of the weighting scheme that is used. For 12.5% the answer depends on the preferred weighting scheme. This result reflects the correlation structure between the well-being scores computed with alternative weighting schemes. The pairwise correlation ranges from 0.99 (between well-being computed with frequency and principal component weights) to a mere 0.51 (between well-being based on most favourable and UNDP-weights).

The crucial question now seems to be which weighting scheme the policy maker should use to identify the worst-off. Although there are theoretical reasons to prefer some weighting schemes above others (see Decancq and Lugo 2013), the policy maker risks receiving very different answers from different multidimensional experts. It appears to us that a natural answer in a democratic society has to make use of the preferences of the individuals themselves. This could be done, for instance, by organizing some kind of vote on the most appropriate weighting scheme. Performing such a vote is the thought-experiment that we will carry out in the next section.

4. Support of the weighting schemes in a hypothetical vote

It seems practically unfeasible to organize a real-life vote in which the respondents cast their vote on the different proposed weighting schemes. However, such a deliberation is not unthinkable in small focus groups (e.g. Van Ootegem and Spillemaeckers 2009) and has been advocated by Sen (1999). Here we simulate the outcome of a vote by exploiting some additional information we have in the data set on the opinions of the individuals on the importance of the dimensions of well-being. As in some other data sets[7], this information is ordinal in nature, without containing meaningful comparable information on the differences between the intensities of the expressed importance. We obtain for each respondent a ranking of the different dimensions of well-being from more to less important (based on her reported importance, as summarized in Table 2). The relative magnitudes of the weights in each weighting scheme imply a similar ranking of the dimensions. To simulate a vote, we need a method to measure the distance between the ranking implied by the opinions of the respondent and the ranking implied by the weighting schemes. In other words, we need a metric defined on the set of rankings.

In the voting literature, John Kemeny (1959) proposed the following natural metric: the distance between two rankings is the number of pairs of alternatives on which they differ (See also Young 1995). Because we are interested in the support that respondents give to the ranking implied by a certain weighting scheme, we use as metric the number of pairs of dimensions that are ranked in the same way by the respondent and by a weighting scheme, divided by the total number of pair wise comparisons. Individual support for a weighting scheme is therefore normalized between 0 and 1, being 0 if the respondent disagrees on all pairwise rankings and 1 if the respondent agrees with all pairwise rankings.[8]

An example can further clarify. Take a simplified case where only the material living standard, health and education are considered as dimensions of well-being. A respondent, who judges health to be more important than the living standard and the living standard more important than education, gives a support of 1 to the frequency weights (the weights of health, livings standard and education in Table 3 are respectively 0.1089; 0.1069 and 0.0985). The same respondent gives a support of 2/3 to the principal component weights (one can check that he disagrees with

[7] The Eurobarometer data set, for instance, contains similar ordinal information on the relative importance of the items of the relative deprivation index. Various studies have used this data set to derive a weighting scheme for the items of the material deprivation index computed on another data set, the EU-SILC (See Guio et al. 2009, Bossert et al. 2009 and Bellani 2011). In these applications, the weighting scheme is obtained from the percentage of individuals giving a high importance to a series of deprivation items. This (normative) approach differs from ours in its objective to derive a new weighting scheme rather than to evaluate other schemes. To obtain a weighting scheme, the authors rely on a particular cardinalization of the ordinal information on opinions and on arguably heavy assumptions on interpersonal comparability of the opinion information.

[8] Alternative metrics between rankings are possible, such as a rank correlation coefficient, Kendall's tau etc. We believe, however, that the pairwise approach as proposed by Kemeny is intuitive and attractive in this setting.

the ranking of the living standard over health, and agrees with the other rankings). Another respondent who considers the dimensions equally important gives a support of 1 to equal weights, but a support of 0 to frequency weights.

The overall support for a weighting scheme is computed as the average support that a weighting scheme gets in the population. This voting procedure is a variant of the so-called Kemeny-Young method or the maximum likelihood voting procedure. It has various attractive axiomatic properties (Young 1995). The first row of Table 4 gives the supports computed for the different weighting schemes. To get an idea of the statistical precision of the differences between the different supports, we perform a bootstrap procedure. We draw (1000 times) a new sample with replacement from our data set (assuming unaltered sample weights) and recalculate the weighting scheme and its corresponding supports. Row two and three of Table 4 present the 2.5th and 97.5th percentile value of these bootstrapped supports such that 95% of the bootstrapped results can be found between these limits.

A first glance at the results shows that a randomly selected respondent is expected to agree with only 24% to 40% of the pair wise rankings implied by the implemented weighting schemes. This finding highlights the magnitude of the problem of paternalism inherent in common weighting schemes.

Table 4. Support for the different weighting schemes

	Data-driven			**Normative**		**Hybrid**			
	FREQ	**PRINC**	**MOFA**	**EQUA**	**UNDP**	**SAT(1)**	**CAP(1)**	**SAT(2)**	**CAP(2)**
support	0.3829	0.3401	0.3060	0.2370	0.3186	0.3825	0.3973	0.3512	0.3826
p2.5	0.3665	0.3230	0.2629	0.2259	0.3111	0.3251	0.3533	0.3152	0.3361
p97.5	0.4071	0.3717	0.3162	0.2485	0.3268	0.4386	0.4419	0.4213	0.4301

Source: LEVO 2010, own computations.

We find that most support is given to the regression-based weighting schemes, and more specifically to the one using the capability indicator as dependent variable (without control variables). Quite remarkably, we find that in the data set at hand the equal weighting scheme gets the lowest support of all schemes. Contrary to earlier work of Chowdhury and Squire (2006) and Nguefack-Tsague et al. (2011) that makes use of expert opinions or principal component analysis, our voting procedure does not justify the use of equal weights, *au contraire*. Also the most favourable weights do not get much support. This may reflect the fundamental difference between giving a large weight to a dimension because of its intrinsic importance in the conception of a good life and valuing it instrumentally to reach a high well-being score.[9] Row two and three of Table 4 show that the confidence intervals for the support are wider for the regression-based approaches compared to the other ones. This is due to the fact that these weights are based on estimations rather than exact computations.

In our data set, we find that hybrid methods outperform the other methods (except the frequency based weights) in terms of support. This stands to reason given their underlying philosophy to take (implicitly) account of the respondents' opinions on the good life. The fact that also the frequency weighting scheme gets a high support, may reflect that there is less deprivation in dimensions which are deemed more important.

5. Conclusion

If individual well-being is conceived as a multidimensional notion, there is a need for measurement and weighting of the various dimensions that are considered to be relevant. This paper concentrates on the intensive weighting problem, assuming that the relevant dimensions are known and measured. We compare various weighting schemes making use of LEVO 2010, a novel data set for Flanders. We obtain quite different weights for the dimensions of well-being by using different procedures. We show that this finding has a potentially large policy impact in terms of the targeting of the worst-off in a multidimensional framework. To evaluate the different weighting schemes, we simulate the outcome of a vote where the respondents express their support for a weighting scheme depending on their agreement with the pairwise rankings of all the dimensions. This is a variant of the so-called Kemeny-Young method or the maximum likelihood voting procedure. We compare the average support that each weighting scheme would get in the population.

Overall, we find a quite low support (between 24% and 40%) for the different weighting systems, which indicates the empirical magnitude of the problem of paternalism of common weighting procedures for this data set. The lowest support has been reached for the equal weighting scheme. Based on this result we cannot justify the common approach of equal weighting. We suggest rather the use of hybrid approaches that incorporate at least some information on the opinions of the individuals themselves.

We can think of various extensions of our method. Firstly, we wonder whether the results can be confirmed by other (more representative) data sets for Flanders and Belgium or other countries. That clearly is an open question which deserves further investigation and requires increased

---

[9] Note that we imposed here common weights for all respondents. The maximal support for the most favorable weighting scheme with individual-specific weights is higher and reaches 0.3659. For this computation we make use of an exogenous lower bound that is chosen over a fine grid to maximize the resulting support (it is reached at 0.099, which brings the most favorable weights close to equal weights).

efforts on multidimensional data collection at the individual level. Secondly, an algorithm can be constructed to find the support maximizing ranking of the dimensions. Under some parametric assumptions on the desired inequality within the scheme, a support maximizing weighting scheme could be derived that is consistent with this ranking. Alternatively, some partial tests aiming at robust results could be implemented (see Foster et al. (2009) and Decancq and Ooghe (2010) for robust approaches with respect to the weights). Finally, it is an open question whether there are socio-demographic regularities in the determinants of support for certain weighting schemes. It would be interesting, from an academic and policy perspective, to gain more insight into who supports which weighting scheme.

References

Anand, P. and van Hees, M. (2006), 'Capabilities and achievements: An empirical study', *Journal of Socio-Economics* 35(2), 268-284.

Bellani, L. (2011), 'Multidimensional indices of deprivation: Should reference groups matter?' *mimeo*.

Bossert, W., Chakravarty, S. and D'Ambrosio, C. (2009), *Multidimensional poverty and material deprivation*, ECINEQ Working Paper 129.

Brun, B. C. and Tungodden, B. (2004), 'Non-welfaristic theories of justice: Is “the intersection approach” a solution to the indexing impasse?', *Social Choice and Welfare* 22, 49-60.

Cherchye, L. and Kuosmanen, T. (2006), ‘Benchmarking sustainable development: A synthetic meta-index approach’, in M. Clarke and M. McGillivray (eds.), *Understanding human well-being*. United Nations University Press, pp. 139-168.

Cherchye, L., Moesen, W., Rogge, N. and Puyenbroeck, T.V. (2007), 'An Introduction to "Benefit of the Doubt" Composite Indicators', *Social Indicators Research* 82(1), 111-145.

Chowdhury, S. and Squire, L. (2006), 'Setting weights for aggregate indices: An application to the commitment to development index and human development index', *Journal of Development Studies* 42(5), 761-771.

Costanza, R., Fisher, B., Ali, S., Beer, C., Bond, L., Boumans, R., Danigelis, N.L., Dickinson, J., Elliott, C., Farley, J., Gayer, D.E., Glenn, L.M., Hudspeth, T., Mahoney, D., McCahill, L., McIntosh, B., Reed, B., Rizvi, S.A.T., Rizzo, D.M., Simpatico, T. and Snapp, R. (2007), 'Quality of life: An approach integrating opportunities, human needs, and subjective well-being', *Ecological Economics* 61(2-3), 267-276.

Decancq, K., Decoster, A. and Schokkaert, E. (2009), 'The evolution of world inequality in well-being', *World Development* 30(1), 11-25.

Decancq, K. and Lugo, M.A. (2013), 'Weights in Multidimensional Indices of Well-Being: An Overview', *Econometric Reviews* (Forthcoming).

Decancq, K. and Ooghe, E. (2010), 'Has the world moved forward? A robust multidimensional evaluation', *Economics Letters* 107(2), 266-269.

Deutsch, J. and Silber, J. (2005), 'Measuring multidimensional poverty: An empirical comparison of various approaches', *Review of Income and Wealth* 51(1), 145-74.

Dolan, P., Peasgood, T. and White, M. (2008), 'Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being', *Journal of Economic Psychology* 29(1), 94-122.

Fleurbaey, M., Schokkaert, E. and Decancq, K. (2009), *What good is happiness?*, CORE discussion paper 2009/17.

Fleurbaey, M. and Trannoy, A. (2003), 'The impossibility of a Paretian egalitarian', *Social Choice and Welfare* 21(2), 243-263.

Foster, J., McGillivary, M. and Seth, S. (2009), *Rank robustness of composite indices*. OPHI Working Paper, no. 26.

Guio, A.-C., Fusco, A. and Marlier, E. (2009), *A European Union approach to material deprivation using EU-SILC and Eurobarometer data*. IRISS Working Papers 2009-19.

Hagerty, M.R., Cummins, R., Ferris, A.L., Land, K., Michalos, A.C., Peterson, M., Vogel, J. (2001), ‘Quality of Life Indexes for National policy; Review and agenda for research’, *Social Indicators Research* 55(1), 1-96.

Kemeny, J.G. (1959), 'Mathematics without numbers', *Daedalus* 88(4), 577-591.

Nguefack-Tsague, G., Klasen, S. and Zucchini, W. (2011), 'On Weighting the Components of the Human Development Index: A Statistical Justification', *Journal of Human Development and Capabilities* 12(2), 183-202.

Rahman, T. (2007), 'Measuring the well-being across countries', *Applied Economics Letters* 14(11), 779-783.

Sen, A.K. (1999), *Development as Freedom*. New Delhi: Oxford University Press.

Schokkaert, E. (2007), 'Capabilities and satisfaction with life', *Journal of Human Development* 8(3), 415-430.

Schokkaert, E., Van Ootegem, L. and Verhofstadt, E. (2011), *Preferences and subjective satisfaction: Measuring well-being on the job for policy evaluation*, CESifo Economic Studies, doi:10.1093/cesifo/ifr018.

Stiglitz, J.E., Sen, A. and Fitoussi, J.-P. (2009), *Report by the Commission on the measurement of economic performance and social progress*.

Commission on the Measurement of Economic Performance and Social Progress, mimeo.

Van Ootegem L. and Spillemaeckers S. (2009), 'With a focus on well-being', *Journal of Socio-Economics* 39, 384-390.

Van Ootegem, L. and Verhofstadt, E. (2011), 'Using capabilities as an alternative indicator for well-being', *Social Indicators Research*, doi: 10.1007/s11205-011-9799-4.

Young, P. (1995), 'Optimal voting rules', *The Journal of Economic Perspectives* 9(1), 51-64.

Young, P. and Levenglick, A. (1978), 'A Consistent Extension of Condorcet's Election Principle', *SIAM Journal on Applied Mathematics* 35, 285-300.