Child Well-Being in Rich Countries

Dijkstra, Theo K.

Child Well-Being in Rich Countries: UNICEF's Ranking Revisited, and New Symmetric Aggregating Operators Exemplified

Theo K. Dijkstra

Abstract In a report published in 2007 UNICEF measured six dimensions of child well-being for the majority of the economically advanced nations. No overall scores are given, but countries are listed in the order of their average rank on the dimensions, which are therefore implicitly assigned 'equal importance'. In this study we take 'equal importance' to mean that the final aggregation is symmetrical in the scores and the ranks, i.e. permuting them leaves the aggregate unchanged. We rank the countries by aggregating the numerical information using a variety of techniques, geared to the measurement scales we distinguish ('ordinal', 'interval', 'ratio'). The aggregators are symmetrical and mildly demanding, emphasizing good performance across the board. The rankings obtained deviate from the UNICEF ranking, but not over-dramatically. Our purpose is not only to study alternative approaches for the particular data at hand, but also to introduce and exemplify new and useful aggregation techniques: we propose ways to select weights for OWA-operators and weighted geometric means, and we suggest how to circumvent the choice of a power for the power means. In addition we extend the Borda method so that it values dominance as well.

T. K. Dijkstra
Economics & Econometrics, University of Groningen, Groningen, The Netherlands

T. K. Dijkstra (✉)
SNS Asset Management, Pettelaarpark 120, 5201 DZ 's-Hertogenbosch, The Netherlands
e-mail: theo.dijkstra@snsam.nl

1 Introduction

The UNICEF Innocenti Research Centre in Florence, Italy, published in 2007 a report on 'child poverty in perspective, an overview of child well-being in rich countries', see UNICEF (2007). In this path breaking work Innocenti brought together the best available data for a multi-dimensional overview of the state of childhood in a majority of economically advanced nations of the world. It distinguished six dimensions: 1. Material well-being, 2. Health and Savety, 3. Educational well-being, 4. Family and peer relationships, 5. Behaviors and risks, and 6. Subjective well-being. Each dimension combines components, that aggregate indicators; the study comprises 18 components and 40 indicators in total. We refer to the reports, UNICEF (2006, 2007), for an in-depth and comprehensive discussion of the collection and construction of the variables involved, including the unavoidable data limitations and intricate methodological issues. At each level of the analysis the data are transformed into z-scores, and simply averaged, up to the level of the dimensions.[1] The dimensions themselves are not averaged. Innocenti resisted the temptation to calculate an overall score, because (see UNICEF 2007, page 39):

> In part this is to maintain opacity and avoid leaning too hard on limited data; composite indicators...need to be as transparent as possible both to keep the process open to debate and to avoid elevating the data to heights of authority that their foundations cannot sustain. But in part, also, reducing the overview to a single score or number would undermine the emphasis on children's well-being as a multi-dimensional issue requiring a wide range of policy responses. Sometimes the whole can be less than the sum of the parts.

Nevertheless, UNICEF (Innocenti) presents a table with an overview of the performances of the countries involved, where the listing is not in alphabetical order, but is based on the average rank on the six dimensions. It also presents 'main findings' directly related to this ranking. These include the observation that European countries, in particular the Northern countries, dominate the top half of the overall league table and that there is no obvious relationship between levels of child well-being and GDP per capita. In addition, care is taken to divide the countries into three subgroups, with seven members each,

---

[1]So the weight that an indicator receives depends on the number of indicators constituting the component that it belongs to, as well as on the number of components that form the dimension containing the indicator. The transformation to z-scores in intermediate stages interferes with this. It is not an issue for this paper, we treat the scores on the dimensions as given.

suggestive of 'leaders', 'followers', and 'laggards'. We feel free therefore to try to rank the countries ourselves and compare our results with the UNICEF ranking as presented. The methods we use are based mainly on Dijkstra (2008), where Ordered Weighted Averaging (OWA-) operators are analyzed for situations where all criteria are deemed equally important. In particular, we focus on a sub-class that awards good performance across the board: since all criteria are important we want to see acceptable scores on all of them. This contrasts with the simple mean where a bad score can simply be compensated by a high score. Or, to put it in terms of policy, a simple average of the scores can be improved by raising the score on any one of the criteria, but our aggregators encourage to raise the lowest scores. The approach we use belongs to a sub-family of Choquet integrals where sets of criteria are assigned weights that exceed the sum of the weights for the separate criteria (so sometimes the whole can be more than the sum of the parts).

An outline of the remainder of the paper now follows. The next section, Section 2, takes the scores[2] on the dimensions in descending order and aggregates them using 'concave, reflection neutral OWA operators'. The ensuing ranking is compared with the ranking on the basis of the average score, taken to be representative for the Innocenti approach. The original scores as obtained from Innocenti are linear transforms of z-scores. Section 3 reverts them to z-scores again and transforms the numbers into normal percentages, i.e. we calculate the area under the normal density to the left of the z-score. In this way we create a kind of ratio scale. We aggregate the percentages using the direct analogue of the OWA-averages, namely the 'concave, reflection neutral, ordered weighted geometric means'. The representative ranking will be compared with the one based on the simple geometric mean. We also use the family of concave power means, with the harmonic mean as its representative. The harmonic mean was suggested by Anand and Sen (1995) as a technique that corrected in a moderate way for imbalances (in particular for imbalances related to gender in the Human Development Index). In Section 4 we return to the ranked data, as in Innocenti's approach. Innocenti used in effect the Borda method, where a country's rank is determined by the total number of countries it outperforms on each of the dimensions. We extend this by taking into account also the total number of countries it outperforms on both of any two dimensions, and similarly on any three dimensions, et cetera up to all six dimensions, in order to honor dominance and a good balanced performance. Countries are awarded points for the Borda numbers in a way that is consistent with the approaches in the previous sections (except the power means). In Section 5 we collect all rankings obtained. Section 6 concludes.

---

[2]We gratefully acknowledge Innocenti's willingness to allow us to use their data.

2 A More Demanding Aggregator: Assigning Higher Weights to Lower Scores

In this section we will aggregate the scores on the six dimensions, denoted by $x_1, x_2, ..., x_6$,[3] using

$$A(x) := \sum_{k=1}^{6} w_k \cdot x_{(k)} \tag{1}$$

where the w_k 's are weights and $x_{(k)}$ is the k-th largest score, so $x_{(1)} \geq x_{(2)} \geq ... \geq x_{(6)}$. These OWA operators were introduced by Yager (1988, 1996, 1999). With equal weights we get the simple average, $\overline{x}$. With all weight assigned to w_1 we take the best score to be representative, and we turn a blind eye to the other scores, no matter how low. This approach is surely too lenient. With all weight assigned to w_6 the scores are summarized in an unforgiving way by the worst score: there is no compensation whatsoever by any of the other scores, no matter how high. This approach is surely too demanding. A compromise that encourages good performance across the board, is obtained by assigning weights to all scores, with higher weights to lower scores. It is shown in Dijkstra (2008) that this is equivalent to concavity of A as a function of x. This implies in particular that when two countries with scores x and y respectively are valued equally, $A(x) = A(y)$, a third country with scores in between (: $\frac{1}{2} \cdot x + \frac{1}{2} \cdot y$) is valued higher. A related implication is that equal scores on all dimensions are valued more than a diverse set of scores with the same simple average: $A(\overline{x}, \overline{x}, ..., \overline{x}) \geq A(x_1, x_2, ..., x_6)$. In other words, below average scores are not simply compensated by equally large above average scores. This will be exemplified below.

Concavity appears to be a natural requirement, but aggregators satisfying it can still be overly demanding. The increase in the weights can be too large for comfort, after all, the minimum score is still one of the possibilities. We suggested in Dijkstra (2008) to add 'reflection neutrality' to ease things up. This concept can be explained as follows. Suppose we have scores 4, 8, and 9 on three equally important criteria. The scores are grades, with the Dutch interpretation: they range from 1 (worst) to 10 (best), with 5 (not passed, but close) and 6 (just passed). The simple mean of the given grades is 7. The differences between grades and mean are -3, $+1$ and $+2$ respectively. If we reverse their signs, and add the ensuing numbers to 7, we get scores 5, 6 and 10, with of course the same mean. We will call the latter set of scores the *mirror* scores, they are obtained by reflecting the old scores in the mean, so to speak. Which sequence is better, (4, 8, 9) or (5, 6, 10)? The latter sequence has both a better worst score and a better best score, but it also has a decidedly lower median score. Without the lowest scores we have (8, 9) versus (6, 10),

---

[3]The approaches we use can be applied to any number of dimensions or criteria of course, there is nothing peculiar about the number 6. But since this paper is not about mathematics, we prefer to stay close to the particular application at hand.

and without the highest scores, (4, 8) versus (5, 6). It appears that one could argue either way. We chose to impose equality for the aggregates of both scores and mirror scores, called *reflection neutrality*, in addition to concavity. More formally, we require that $A(x) = A(y)$ for every x and y that mirror each other, meaning that y_k, the observation on the k-th dimension, equals $\bar{x} - (x_k - \bar{x})$ whereas $x_k = \bar{x} + (x_k - \bar{x})$. It can be shown that the set of weights satisfying both concavity and reflection neutrality equals all possible weighted combinations of the rows of the following matrix E, say:

$$E := \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 & 2 & 2 \\ 0 & 0 & 0 & 2 & 2 & 2 \end{vmatrix} \div 6. \tag{2}$$

(Note that E is constructed in a very simple way, valid for all numbers of criteria: start with a row of ones, take away the first '1' and add it to the last '1' to get the second row, then take from the second row the first '1' and add it to the penultimate element to get the third row; repeat this until only one '1' or no '1' is left). The ensuing averages combine weighted averages covering the lower half of the scores and more. The minimum score is no longer an option. Now the most unforgiving aggregator, the least inclined to allow for compensation, is the average of the lowest three scores. Had we ignored reflection neutrality, we would have combined all averages of the form $\bar{x}_{(k:6)} := (x_k + ... + x_6) \div (7 - k)$. It is useful to note that when we average the rows of E we get the barycentre of the convex polytope generated by these rows: $[1, 2, 3, 5, 6, 7] \div 24$. It is the representative weight vector for concave reflection neutral weights. It yields almost the *rank weighted average* with weights $[1, 2, 3, 4, 5, 6] \div 21$. For uneven numbers of criteria the rank weighted average *is* representative. See Dijkstra (2008) for more general results.

We took 10.000 random combinations of the rows of E, for each combination we calculated overall scores for all of the countries involved and ranked them. (Here rank is defined as in sport competitions: one plus the number of countries with a better score.) Finally we noted how often each country occupied every one of the possible ranks,[4] and sorted the countries according to the average rank. This agrees with the ranking obtained using the representative weights. Table 1[5] collects the results.

Please observe that the scores are in descending order, so the identity of the dimensions is erased, in agreement with the assumption of symmetry or 'equal

---

[4]The 10.000 combinations virtually guarantee that the percentages obtained are well within 1 percentage of the values obtainable with an 'infinitely large' sample. This accuracy suffices for practical purposes.

[5]We deviate in one respect from the data as given by Innocenti. There was insufficient information to assign the USA a score on dimension 6, 'Subjective well-being'. Innocenti ranked the countries by their average ranking on the dimensions, using only 5 dimensions for the USA. This meant that the USA effectively got the average rank on this dimension. An analysis of the data indicated however that a score equal to Canada's score on dimension 6 with rank 15 had some plausibility. We work with this throughout this paper. Other choices are possible, it will not matter a great deal.

Table 1 Innocenti's scores, sorted

	(1)	(2)	(3)	(4)	(5)	(6)	$(\overline{x})$
1. Netherlands	118.3	112.8	112.4	112.0	108.2	101.6	1
2. Sweden	120.2	118.7	114.9	108.7	104.6	98.0	2
3.Norway	117.5	104.3	103.6	102.1	102.0	100.8	5
4. Finland	115.3	111.5	110.8	104.0	99.2	96.4	3
5. Denmark	112.2	110.5	106.7	102.2	100.9	97.8	4
6. Switzerland	110.0	108.3	106.2	102.0	101.2	95.5	7
7. Spain	116.6	106.5	106.2	103.8	99.7	91.2	6
8. Germany	104.0	102.0	101.5	100.0	99.3	97.2	11
9. Ireland	110.3	106.9	106.7	104.3	90.7	89.2	9
10. Belgium	116.2	106.5	105.2	99.0	91.6	91.0	8
11. Italy	115.3	106.3	102.7	99.5	96.6	83.8	10
12. France	105.0	102.0	99.5	99.4	92.0	90.8	14
13. Greece	115.5	103.7	100.8	94.8	91.2	90.3	12
14. Canada	114.0	105.4	97.2	94.3	93.5	89.3	13
15. Czech R.	103.9	102.9	101.0	100.1	91.0	87.2	16
16. Austria	108.0	104.6	97.5	96.2	86.3	85.5	17
17. Portugal	115.0	96.3	96.2	93.9	93.7	81.9	18
18. Poland	113.6	112.5	99.0	95.8	85.2	80.2	15
19. Hungary	105.8	95.9	95.8	92.8	91.1	82.4	19
20. USA	96.0	93.5	92.8	83.4	80.3	80.0	20
21. UK	98.0	91.1	89.1	82.3	77.6	72.5	21

importance'. Innocenti has normalized the z-scores in such a way that their average equals 100 and their standard deviation 10. We rounded the scores to the first decimal to ease the readability. As one can see, the ranking based on a more demanding way of aggregating differs from the one based on the simple average, indicated by the column headed by '$(\overline{x})$'. In particular, Germany, with a slightly lower average than Italy, is placed above Italy. Germany has scores hovering around 100, whereas Italy is once well above 100 but also well below 100. The bad score is not simply balanced out by the good score.

3 Means Appropriate for Ratio Scales

In the previous section we analyzed data that can be described, loosely and not without abuse of terminology, as measured on an interval scale. More precisely, we can multiply all numbers with the same positive constant and add an identical constant of any sign without changing the information: the mean of 100 and standard deviation of 10 are arbitrary. This entails that means like geometric or power means are not appropriate for the given data set. But they are appropriate, if we transform the scores into percentages by calculating the area below a density to the left of the original z-scores. Many densities are possible but we will use the normal or Gaussian density for illustrative purposes. So the score x_i on the i-th dimension is replaced by $\Phi((x_i - 100) \div 10)$ where $\Phi(z) := \int_{-\infty}^{z} \left(1/\sqrt{2\pi}\right) \cdot \exp\left(-\frac{1}{2} \cdot x^2\right) dx$ for real

values of z. We will not adjust the notation, in this section x_i now stands for the normal percentage score.

The obvious counterpart of the simple, arithmetic mean is the simple geometric mean, g, say, defined as

$$g(x) := \left(\Pi_{i=1}^{6} x_i\right)^{1/6}.$$ (3)

The corresponding OWA- operator is

$$G(x) := \Pi_{i=1}^{6} x_{(i)}^{w_i}$$ (4)

where the x's are sorted again in descending order. With increasing weights G has the property that when two vectors of scores x and y have identical valuations, $G(x) = G(y)$, then $G(z) \geq G(x) = G(y)$ for any z between x and y, that is $z_i := x_i^{\lambda} y_i^{1-\lambda}$ for any $\lambda \in [0, 1]$, the same for all i. So the more balanced the scores, i.e. the smaller the variation, the higher the valuation. As before we will also demand 'reflection neutrality'. This means here that when we take $y_i := g(x) \cdot (g(x)/x_i)$, so that $g(y) = g(x)$ and the y-scores can be said to mirror the x-scores, then $G(y) = G(x)$. Consequently, the weights of the previous section can be used again. Ten thousand different weighted combinations of the rows of the matrix E were generated, and for each combination we calculated the overall scores and ranked the countries. We ascertained how many times each possible rank was occupied and sorted the countries on the basis of their average rank. This happened to agree with the ranking based on the representative weights. Section 5 collects the results and also shows how countries rank using the simple geometric mean.

Anand and Sen (1995) advocated the use of the harmonic mean for valuations of scores where balance is important, or rather where *im*balance ought to be penalized, as in the case of the Human Development Index corrected for gender inequality. The harmonic mean is a member of the family of power means, $P_\alpha(x)$, say, defined by

$$P_\alpha(x) := \left(\sum_{i=1}^{6} x_i^{\alpha}/6\right)^{(1/\alpha)}$$ (5)

where α is a real number, different from zero. It is natural to define $P_0(x) := g(x)$, since $P_\alpha(x) \to g(x)$ when $\alpha \to 0$. The power mean is increasing in α, it ranges from the minimum score for $\alpha = -\infty$ to the maximum score for $\alpha = \infty$ with well-known stops in between: the harmonic mean for $\alpha = -1$, the geometric mean for $\alpha = 0$, the simple arithmetic mean for $\alpha = 1$. See Steele (2004) for a beautiful and complete analysis. An intuitive explanation is as follows. Exponentiation by α scales the observations, so raising both sides of equation (5) to the power α, yields that the scaled power mean is just the simple average of the scaled observations. A graphical analysis quickly reveals that when α belongs to $[0, 1)$ or to $(-\infty, 0)$, scaling stretches the lower range of the percentages and does the opposite to the higher range. So if we undo the scaling of the power mean we get a number that is closer to the lower scores

than the simple mean of the scores. The effect is stronger the further α is to the left of 1. For $\alpha > 1$ the opposite happens.

We obtain concavity of the power mean if and only if $\alpha \leq 1$, see Bullen (2003). In other words, if $\alpha \leq 1$ and the scores of two countries have the same power mean, a country with scores in between will be valued higher.

Anand & Sen suggested the harmonic mean because they felt that it imposes a moderate penalty for imbalance, but they did not offer compelling arguments against values other then $\alpha = -1$. We propose here to use the whole concavity range, $\alpha \leq 1$, and select values according to the density that maximizes the entropy among all densities with expected value equal to minus one. It is easily verified that the maximum entropy density is exponential, it equals $\frac{1}{2} \exp \left[\frac{1}{2} (\alpha - 1) \right]$ on $\alpha \leq 1$. We can generate values from this density using $1 + 2 \cdot \log (U)$ where U is uniformly distributed on $[0, 1]$. (If we had wanted the representative average to be the geometric mean, we would have used $1 + \log (U)$). So we work as before, generating 10.000 different α's, calculating for each α overall scores for the countries, ranking them, counting the number of times all possible ranks are attained, and finally sorting them according to the average rank.[6] In this case the average rank disagrees for a number of countries with the rank obtainable for the harmonic mean. We had expected this to happen also for the geometric means, but it did not. Section 5 has the results.

A final remark: there are many, many ways of aggregating scores on a 'ratio' scale. For an extensive overview of what the virtually limitless ingenuity through the ages has accomplished in this respect we refer to Bullen (2003). The geometric means and power means are perhaps among the more well-known and best understood in terms of axioms and generating functional equations, see also Aczél (1966, 2006).

4 Aggregation of Ranks, Rewarding 'Dominance'

When Innocenti constructed the overall league table, it effectively summed the ranks on all six dimensions. This is equivalent to the famous Borda method,[7] named after the French mathematician and naval hero de Borda, who introduced it to the French Academy of Sciences in 1770 (but documented applications by the Romans go back as far as the first century). Its pros and cons as a means of aggregating rankings are extensively discussed, if not hotly debated, starting with de Borda versus Condorcet. The matter is not quite

---

[6]We could define the expected value of the power mean of the scores of each country with respect to the exponential distribution just specified, as the power mean of country scores (the expected value will inherit the concavity). The ranking could then be based on these means. This has the advantage that the relative ranking of the countries at hand does not change when other countries are added.

[7]Listing the countries in ascending order of the sum of the ranks is equivalent to ranking them according to the Borda Count B_1, see below, in descending order.

settled till this day.[8] We venture here to add a variation, that attempts to award dominance on subsets of the dimensions. The basic idea is to extend the so-called Borda Count, B_1 say, calculated for each country by summing the number of countries it outperforms on each of the separate dimensions. We suggest here to *also* count the number of times it does better on both of any two dimensions, denoted by B_2, similarly for any three dimensions, denoted by B_3, et cetera, up to and including the full set of six dimensions, B_6. So for each country we obtain a vector $B := (B_1, B_2, B_3, B_4, B_5, B_6)$ that captures the extent to which the country dominates its competitors. The question now is how to assign credits to B, acknowledging the implicit double counting in its components. In order to do that in a relatively clean way, we resort to a transformation of B, denoted by $A := (A_1, A_2, A_3, A_4, A_5, A_6)$ where A_i counts the number of times the country outperforms the others on *exactly i* dimensions, not as before on *at least i* dimensions. It can be shown that

$$A_i = \sum_{j=i}^{6} \binom{j}{j-i} (-1)^{j-i} B_j, \tag{6}$$

$$B_i = \sum_{j=i}^{6} \binom{j}{j-i} A_j. \tag{7}$$

In particular, $B_1 = A_1 + 2 \cdot A_2 + 3 \cdot A_3 + 4 \cdot A_4 + 5 \cdot A_5 + 6 \cdot A_6$, as intuitively obvious. If we award outperformance on exactly i dimensions by C_i credits, a country receives in total $\sum_{i=1}^{6} C_i \cdot A_i$ points; we naturally put $C_0 := 0$. In terms of Borda Counts we have, as one may verify:

$$\sum_{i=1}^{6} C_i \cdot A_i = \sum_{i=1}^{6} \nabla^i C_i \cdot B_i \tag{8}$$

where ∇ is the backward difference operator, so that $\nabla^1 C_i = C_i - C_{i-1}, \nabla^2 C_i = \nabla^1 \left(\nabla^1 C_i \right) = C_i - 2 \cdot C_{i-1} + C_{i-2}$ et cetera, in short $\nabla^i C_i = \sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} C_{i-j}$. Therefore the coefficients of the Borda Counts measure how fast the 'credits function' grows. Clearly, C_i will increase or at any rate not decrease with i. If we normalize by setting $C_6 = 6$ we can reproduce Borda's method by taking $C_i = i$, as is clear from the expression for B_1. But more can be done if we want to award dominance: it seems reasonable to take $C_2 - C_1 \geq C_1 - C_0 > 0$, so $\nabla^2 C_2 \geq 0$, and more generally $\nabla^2 C_i \geq 0$.[9]

---

[8]See e.g. Brams and Fishburn (2002) and Pattanaik (2002). Donald Saari has published extensively about the Borda method, see e.g. Saari (2001a, b), emphasizing its ability to neutralize the (hidden) symmetries responsible for many of the classical voting paradoxes.

[9]It is shown in Dijkstra (2008) that this is equivalent to $C_{i+j+k} - C_{i+j} \geq C_{i+k} - C_i$ for nonnegative integers i, j, k. So k additional dimensions on which the country outperforms yields more credits when it outperforms on more dimensions to start with.

However, as in Dijkstra (2008), we will restrict C somewhat, to soften its demanding nature, by imposing reflection neutrality. This means simply that we will generate credits functions using the *cumulative weights* of the weights of Section 2. These cumulative weights, appropriately normalized, are E^* say, with

$$E^* := \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 1 & 2 & 3 & 4 & 6 \\ 0 & 0 & 1 & 2 & 4 & 6 \\ 0 & 0 & 0 & 2 & 4 & 6 \end{bmatrix}. \tag{9}$$

Arbitrary credits functions can be obtained by taking arbitrarily weighted averages of the rows of E^*. The representative credits function, where the rows are equally weighted equals $[1, 3, 6, 11, 17, 24] \div 4$. This is to be contrasted with Borda's $[4, 8, 12, 16, 20, 24] \div 4$. The latter is linear, the former is 'quadratic'.

As will be expected by now we generated 10.000 combinations of the rows of E^*. For each combination, credits function, we calculated the points for the countries, and ranked them. The average rank agreed once again with the rank based on the representative credits function. Table 2 gives the 'A-values' for the countries listed according to the consensus ranking as well as the Borda rank (as used by Innocenti).

The Borda Count B_1, for the Netherlands equals $0 \times 1 + 1 \times 2 + 0 \times 3 + 4 \times 4 + 7 \times 5 + 8 \times 6 = 101$ et cetera. The more balanced the outperformance the higher the rank: roughly, the top countries have smaller counts on the left side and larger counts on the right side, the opposite being true for the lower

Table 2 The number of countries outperformed on *exactly* 1,2,...,6 dimensions

	A_1	A_2	A_3	A_4	A_5	A_6	Borda rank
1. Netherlands	0	1	0	4	7	8	1
2. Sweden	0	0	0	7	10	3	2
3. Finland	1	1	4	6	5	3	3
4. Denmark	3	1	3	5	5	3	5
5. Spain	0	1	5	6	5	2	4
6. Norway	2	3	3	3	7	2	7
7. Switzerland	1	2	5	4	7	1	5
8. Italy	3	1	8	4	3	1	8
9. Belgium	1	6	4	6	2	1	9
10. Ireland	0	2	8	8	1	0	9
11. Germany	4	3	5	3	2	2	11
12. Greece	0	6	8	1	3	0	12
13. Canada	5	4	6	3	2	0	13
14. Czech R.	2	5	6	2	1	1	16
15. France	5	4	5	4	0	1	14
16. Poland	1	11	5	3	0	0	14
17. Austria	7	6	4	2	1	0	17
18. Portugal	8	4	3	3	1	0	18
19. Hungary	7	6	4	2	0	0	19
20. USA	9	3	0	0	1	0	20
21. UK	4	5	0	0	0	0	21

Table 3 Innocenti's ranks, sorted

	(1)	(2)	(3)	(4)	(5)	(6)	Borda rank
1. Netherlands	1	2	3	3	6	10	1
2. Sweden	1	1	1	5	7	15	2
3. Finland	3	3	4	6	11	17	3
4. Denmark	4	4	8	9	12	12	5
5. Spain	2	5	5	8	12	16	4
6. Norway	2	8	8	9	10	13	7
7. Switzerland	4	5	6	9	10	14	5
8. Italy	1	6	9	10	14	20	8
9. Belgium	1	5	7	12	17	19	9
10. Ireland	4	5	7	7	19	19	9
11. Germany	9	10	10	11	13	13	11
12. Greece	3	7	11	15	17	18	12
13. Canada	2	6	14	15	17	18	13
14. Czech R.	8	10	11	11	18	19	16
15. France	7	9	12	14	15	19	14
16. Poland	2	3	14	16	20	21	14
17. Austria	4	8	15	16	19	20	17
18. Portugal	2	14	15	16	16	21	18
19. Hungary	6	13	13	17	18	20	19
20. USA	12	15	17	20	20	21	20
21. UK	13	18	18	21	21	21	21

ranked countries. Table 3 yields the original Innocenti information, where we sorted the ranks in 'descending' order:

The Borda Count B_1, for the Netherlands can now be calculated as $(21 - 1) + (21 - 2) + (21 - 3) + (21 - 3) + (21 - 6) + (21 - 10) = 20 + 19 + 18 + 18 + 15 + 11 = 101$ as before. We feel that with equally important dimensions the 'A-table' adds valuable information and helps to determine the position in the group. But one should not overlook the pecularities of working with rankings and Borda-type counts. We use ordinal information only, which can be an advantage if differences or ratios on the original scales are not necessarily meaningful, but it can also be wasteful. And as opposed to the other methods we discussed, one cannot assign a stand-alone value to a country. Borda-counts are by neccessity always relative to the group of countries one studies. Adding another country could affect the relative ranking of the original members. This is also the case with our procedure for ranking the countries, using 10.000 different sets of weights, but not necessarily so: we can use the weights to get a representative score for a country, cf. footnote 6.

5 An Overview

Table 4 presents all the rankings as determined by the various techniques we employed. The countries are listed in alphabetical order to avoid the appearance of bias, and to ease the readability. The last column gives the full range of the ranks that can be obtained with the *general* OWA-method, where

Table 4 (Representative) rankings

	B_1	A	$(\overline{x})^a$	OWA/a	$(g)^b$	OWA/g	$(\alpha = -1)$	$(\alpha \leq 1)$	Range
Austria	17	17	17	16	16	17	16	16	14–19
Belgium	9	9	8	10	9	11	9	9	6–11
Canada	13	13	13	14	14	15	13	14	10–17
Czech R.	16	14	16	15	15	13	14	15	12–18
Denmark	6	4	4	5	5	4	4	4	3–8
Finland	3	3	3	4	3	5	5	5	3–6
France	14	15	14	12	12	12	10	11	11–17
Germany	11	11	11	8	8	8	7	8	7–16
Greece	12	12	12	13	13	14	12	12	8–16
Hungary	19	19	19	19	19	19	18	19	17–19
Ireland	10	10	9	9	10	9	11	10	5–14
Italy	8	8	10	11	11	10	15	13	7–15
Netherlands	1	1	1	1	1	1	1	1	1–2
Norway	7	6	5	3	4	3	3	3	3–10
Poland	15	16	15	18	18	16	19	18	5–19
Portugal	18	18	18	17	17	18	17	17	12–18
Spain	4	5	6	7	7	7	8	7	4–8
Sweden	2	2	2	2	2	2	2	2	1–2
Switzerland	5	7	7	6	6	6	6	6	5–10
UK	21	21	21	21	21	21	21	21	20–21
USA	20	20	20	20	20	20	20	20	20–21

[a]This corresponds to $\alpha = 1$.
[b]This corresponds to $\alpha = 0$.

we average the sorted z-scores with *arbitrary* weights,[10] not restricted to be concave and reflection neutral. The range of possible ranks varies in principle with the method chosen (as exemplified by France whose rank based on the harmonic mean lies outside of the given range). It is clear that depending on how demanding (emphasizing the worst score) or how lenient (looking mainly at the best score) one wants to be the ranks may vary a lot. A case in point is Poland, whose scores range from 5–19, although high ranks are rather infrequent (it occupies any of the ranks 5–10 in no more than 9% of the cases). The headings above the other columns containing numerals are probably self-explanatory, but to be a little more specific: B_1, $(\overline{x})$, (g) and $(\alpha = -1)$ indicate the rankings obtained based on the standard Borda method (the Innocenti ranking), the simple mean of the z-scores, the simple geometric mean and the harmonic mean of the percentage scores respectively; the others are *representative* rankings for the extended Borda method (based on A), and the 'concave reflection neutral ordered weighted averages', the additive variant OWA/a (Section 2) and the multiplicative variant OWA/g (Section 3) respectively.

---

[10]We used brute force here, by generating 40.000 different sets of weights, ranking the countries for each set and noting the range. The Netherlands is ranked first and second in 52% and 48% of the cases resp., Sweden complements these figures. France occupies the ranks 11, 12, 13, 14, 15, 16 and 17 with probabilities 5, 11, 15, 22, 20, 18 and 9 resp.

It is obvious that when countries do not show an approximate equally good or equally bad performance on the dimensions, then their ranks depend on the method chosen and the information used, whether ordinal,'interval-'or 'ratio-'scaled. Striking examples are provided by Italy and France if we take all columns to be comparably relevant and adequate. If we discard the techniques that use the normal percentages, which are arguably somewhat artificial, then the picture is relatively stable. In particular, the partitioning into three groups of equal size appears to be relatively robust, eventhough the within-group variation is nonnegligible.

6 Conclusion

In this paper we analyzed aspects of the valuable data set Innocenti has constructed for UNICEF to 'assess the lives and well-being of children and adolescents in the economically advanced nations'. With respect to content we had nothing to add, we just looked at the final scores per dimension, and tried to construct reasonable overall rankings. Our starting point was the assumption of equivalence of the six dimensions. This is implicit in the way Innocenti put the overall league table together, and is consistent with the equal weighting of indicators and components in the construction of the dimensions. Innocenti's position concerning the weighting issue is subtle and complicated (UNICEF 2007, p.5):

> Equal weighting is the standard approach used in the absence of any compelling reason to apply different weightings and is not intended to imply that all elements used are considered of equal significance.

This may very well reflect the (early) stage the research is in about the highly complicated, high-dimensional concept of child well-being. So it may indeed be quite sensible and rational to postpone the design of appropriate aggregators when one is still predominantly concerned with the mapping and measurement of the various aspects of the concept. Although of quite a different nature and arguably infinitely less important, there is nevertheless in some respects a resemblance with the ranking of decathlon athletes who compete in ten different if not incomparable disciplines. It took more then 70 years to establish something of a consensus about how to decide who does best overall. Part of the difficulty, and the source of the resemblance with child well-being, is of course that a concept like 'best all-round' athlete is a somewhat free-floating concept, not tied in a comprehensible way to some independent variable that is measurable. The scales on which the decathlon scores are measured are not constructed so as to optimally predict another variable. Similarly, the design of the scales and the choice of the aggregator for child well-being are not geared to the prediction of other variables, like 'happiness as an adult'. Rather, an attempt is made to summarize high-dimensional data by means of a one-dimensional construct. The aggregators we suggested all try to do justice to the idea that *all* dimensions are important, and that good performance across

the board is what one should aim for. Statisticians refer to these and similar situations as 'unsupervised learning', as opposed to 'supervised learning', where the success of a construction can be measured in terms of its ability to predict other variables. A quote from Hastie et al. (2001), p.439 seems appropriate here:

> It is difficult to ascertain the validity of inferences drawn from the output of most unsupervised algorithms. One must resort to heuristic arguments not only for motivating the algorithms, as is often the case in supervised learning as well, but also for judgements as to the quality of the results. This uncomfortable situation has led to heavy proliferation of proposed methods, since effectiveness is a matter of opinion and cannot be verified directly.

One way forward could be to try to embed child well-being into a theory that weaves a web between it or its dimensions and other causally or predictively related concepts. In other words, it could be helpful, albeit quite challenging, to resort to full-blown latent variable modelling. Child well-being would then be one of the latent variables, approximately measurable by its conditional expectation given its direct indicators, or by a suitable linear compound such as a canonical variable. Part of the challenge is due to the paucity of the data from a statistical point of view, since the number of countries that can be measured is rather modest. So the study may well require a longitudinal approach. But in principle at least the weights of an index for child well-being can be chosen to be mutually consistent with the weights of other relevant constructs,[11] enhancing in the process their acceptability as well as their usefulness. Of course, *usefulness* is the keyword here as is clearly recognized by UNICEF and Innocenti: the goal is ultimately not to rank countries, but to devise ways and means that help policy makers create the conditions in which child well-being can be fostered all around the globe.

References

Anand, S., & Sen, A. (1995). *Gender inequality in human development: Theories and measurement.* Occasional Paper 19. United Nations Development Programme, Human Development Report Office, New York (GDI, GEM).
Aczél, J. (1966, 2006). *Lectures on functional equations and their applications* (1996 ed., 2006 ed.). London: Academic, Dover.

---

[11]See Dijkstra and Hensler (2008) for an overview of the relevant literature, and suggestions how to construct indices that are linked together in (non)-linear relationships by means of 'best fitting proper indices'.

Brams, S. J., & Fishburn, P. C. (2002). Voting procedures. In K. J. Arrow, A. K. Sen, & K. Suzumura (Eds.), *Handbook of social choice and welfare* (Vol. 1, pp. 173–236). Amsterdam: Elsevier.

Bullen, P. S. (2003). *Handbook of means and their inequalities.* Dordrecht: Kluwer Academic.

Dijkstra, T. K. (2008). *'Equally important' criteria and demanding, balanced aggregation.* Working Paper, University of Groningen, Economics & Econometrics. http://www.rug.nl/staff/t.k.dijkstra/BalancedAggregation.pdf

Dijkstra, T. K., & Hensler, J. (2008). *Prescriptions for dimension reduction, with interacting factors.* Working Paper, University of Groningen, Economics & Econometrics. http://www.rug.nl/staff/t.k.dijkstra/Prescriptions.pdf

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning.* New York: Springer.

Pattanaik, P. K. (2002). Positional rules for collective decision-making. In K. J. Arrow, A. K. Sen & K. Suzumura (Eds.), *Handbook of social choice and welfare* (Vol. 1, pp. 361–394). Amsterdam: Elsevier.

Saari, D. G. (2001a). *Chaotic elections!: A mathematician looks at voting*. Providence: American Mathematical Society.

Saari, D. G. (2001b). *Decisions and elections. Explaining the unexpected.* Cambridge: Cambridge University Press.

Steele, J. M. (2004). *The Cauchy-Schwarz master class.* Cambridge: Cambridge University Press.

UNICEF (2006). *Comparing child well-being in OECD countries: Concepts and methods.* Innocenti Working Paper IWP-2006-03, by Bradshaw, J, Hoelscher, P. and D. Richardson.

UNICEF (2007). *Child poverty in perspective: An overview of child well-being in rich countries.* Innocenti Report Card 7.

Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man and Cybernetics, 18*, 183–190.

Yager, R. R. (1996). Quantifier guided aggregation using OWA operators. *International Journal of Intelligent Systems, 11*, 49–73.

Yager, R. R. (1999). Criteria aggregations functions using fuzzy measures and the Choquet integral. *International Journal of Fuzzy Systems, 1*, 96–112.