Version 1.0.0 | Working paper

# How weights change the winner

## A reproducible test of composite equality scores

**Working paper 1.0.0. Author review pending. All numerical profiles in this paper are artificial teaching examples.**

## Abstract

A composite index can reverse the order of two profiles without a single underlying observation changing. This paper demonstrates the mechanism using two explicitly artificial profiles measured on two dimensions. Profile A has values of 90 and 40; Profile B has values of 60 and 80. Under an arithmetic weighted average, their ordering reverses when the first dimension's weight crosses four-sevenths. The calculation is exact and reproducible. It is not evidence about any country or demographic group. We use the example to specify a practical reporting standard for World Equality Index: publish component values, define the admissible weights, show switching thresholds, distinguish model sensitivity from sampling uncertainty, and explain whether strengths may compensate for weaknesses. A dashboard of distinct dimensions remains the primary empirical object. An aggregate score, where defensible, is a declared interpretation of those dimensions.

## Why this matters for equality research

Rights, opportunity, outcomes, contribution, burden, and resilience answer different questions. A jurisdiction can perform well on one dimension and poorly on another. Adding those dimensions together creates an additional claim: that their tradeoffs are meaningful on a common scale and that the chosen exchange rates suit the stated purpose. A ranking can appear precise even when that interpretation is the most contested part of the calculation.

The OECD and European Commission Joint Research Centre handbook treats indicator selection, normalization, weighting, and aggregation as choices that require transparent justification. Its discussion of robustness recommends examining how uncertainty and alternative assumptions affect an index. This paper adopts that general methodological direction and supplies its own small analytic example and proposed WEI reporting rules. OECD, EU and EC-JRC handbook, 2008, especially Steps 6 and 7

There is no observed country dataset in this demonstration. The names Profile A and Profile B are deliberately artificial. The dimension labels Rights and Outcomes are illustrative placeholders, not measured legal codes or demographic outcomes. Their numerical values were chosen to create a transparent tradeoff that can be solved by hand. They must never be displayed as national scores, passed into an empirical explorer as real observations, or described as findings about social groups.

## The model

Assume two dimension scores lie on a fixed scale from 0 to 100, with larger values considered better under the illustrative scoring convention. Profile A has Rights = 90 and Outcomes = 40. Profile B has Rights = 60 and Outcomes = 80. Let w be the weight on Rights, and let 1 minus w be the weight on Outcomes. The admissible interval is 0 through 1.

The arithmetic score is:

$S(w) = w \times \text{Rights} + (1 - w) \times \text{Outcomes}$

Substituting each profile gives:

$S_A(w) = 90w + 40(1 - w) = 40 + 50w$

$S_B(w) = 60w + 80(1 - w) = 80 - 20w$

Subtracting the two scores yields:

$S_A(w) - S_B(w) = 70w - 40$

The two profiles tie when 70w equals 40, or w = 4/7, approximately 0.571429. Below that threshold B ranks higher. Above it A ranks higher. At the threshold they tie. The observations and arithmetic are fixed; the changed order is entirely a result of the permitted weight choice.

## Results of the demonstration

| Rights weight | Outcomes weight | Profile A score | Profile B score | Order |
|---|---|---|---|---|
| 0.00 | 1.00 | 40.00 | 80.00 | B above A |
| 0.25 | 0.75 | 52.50 | 75.00 | B above A |
| 0.50 | 0.50 | 65.00 | 70.00 | B above A |
| 4/7 | 3/7 | 68.571429 | 68.571429 | Tie |
| 0.75 | 0.25 | 77.50 | 65.00 | A above B |
| 0.80 | 0.20 | 80.00 | 64.00 | A above B |
| 1.00 | 0.00 | 90.00 | 60.00 | A above B |

Every value in this table is calculated from the two artificial profiles. No estimate was fetched from a social dataset. At equal weights B leads by five points. At a Rights weight of 0.80, A leads by sixteen points. Calling either result the objectively correct overall rank would conceal the substantive choice embodied in the weight.

The relevant finding is more informative than saying that weights matter. It identifies exactly where the answer changes. If a research brief had justified a Rights weight between 0.20 and 0.50, B would lead throughout that specified interval. If it had justified an interval from 0.65 to 0.90, A would lead throughout. If the admissible interval straddles four-sevenths, the

ordering is contingent within the model family. The defensibility of an interval must be assessed before selecting the preferred rank.

## Robustness has a defined domain

An ordering is robust only relative to a stated set of allowed assumptions. Reporting a ranking as robust after varying one weight from 0.49 to 0.51 says little about a plausible range from 0.10 to 0.90. The interval, normalization method, aggregation rule, and missing-data treatment belong next to the robustness result. The result cannot be detached from the experiment that produced it.

A useful special case is componentwise dominance. If one profile is at least as high on every dimension and strictly higher on a dimension receiving positive weight, an arithmetic score with nonnegative weights ranks it higher. No tradeoff reverses that ordering within those assumptions. Our example intentionally lacks dominance: A leads on one component and B leads on the other. The crossover is a property of that conflict, not a defect in arithmetic.

For more than two dimensions, a single slider no longer covers all combinations of weights. A future empirical implementation should define a feasible weight region and inspect rankings across it. A coarse grid can demonstrate examples, but it cannot automatically establish that an extreme rank is impossible between sampled points. Exact optimization or a documented approximation procedure should support claims about attainable rank bounds.

## What a slider does and does not mean

An interactive weight control can help a reader explore assumptions. The control should show the full weight vector and normalize it consistently, rather than changing hidden companion values. It should retain the raw components on screen and allow a scenario to be exported. A saved scenario is a statement of a reader's chosen calculation, not a replacement for the published reference method.

If a tool samples weights uniformly, the proportion of samples in which A leads is a property of that sampling scheme. It is not the probability that A is truly more equal, the probability of public agreement, or a statistical confidence level. In this two-dimensional example, uniform $w$ on the interval from 0 to 1 gives A the upper three-sevenths of the interval. That describes a geometric proportion under an arbitrary weight distribution. Another distribution changes the proportion while leaving all profile values untouched.

Nor are sensitivity ranges sampling intervals. A confidence interval addresses uncertainty in an estimated quantity under a statistical procedure. A score range over weights addresses the consequences of changing the scoring rule. Both can matter in empirical work, but combining them into an unlabeled uncertainty band would make the chart harder to interpret.

## Compensation and scale

The arithmetic formula permits compensation. A gain on one dimension can offset a loss on the other. Some research purposes may accept that tradeoff; others may require a minimum threshold for a right or an essential service. A rule with mandatory thresholds answers a different question from an unrestricted average. WEI should state the permitted tradeoff instead of treating a familiar formula as a neutral default.

The chosen scale also matters. Moving a dimension from a percentage to a count without rescaling can change its numerical influence dramatically. Even a nominally equal weight does not ensure equal practical influence when component distributions differ. Before assigning weights, a paper needs to define the construct, direction, units, transformation, and reference population for each component. A high score must have a precise interpretation in that context.

For equality measures, even direction may be contested. A smaller group difference in one observed outcome does not establish equal treatment, and unequal outcomes can arise under more than one institutional process. The choice to reward outcome parity should be documented as the construct being scored. A legal-rights measure must come from evidence about rights, rather than being silently inferred from an employment distribution.

## A proposed WEI publication standard

Each empirical composite should have an explicit purpose and a component inventory. The inventory should include source versions, geographical coverage, periods, denominators, missing values, transformations, and the substantive meaning of higher values. Readers should be able to inspect a component without accepting the aggregate interpretation.

The method should then identify a reference weight vector and justify its use. Alternative admissible weights and aggregation rules should be specified before reviewing which entities benefit. Results should report componentwise comparisons, reference scores, switching thresholds where calculable, and rank or score ranges under the declared alternatives. Cases that change order deserve visible disclosure, not a footnote describing the index as generally robust.

Missing components need a rule that preserves the stated comparison. Renormalizing weights for one entity while keeping the original weights for another can compare different constructs. A calculation should disclose when entities are excluded, when data are imputed, and whether a result depends on that choice. A missing value must not silently become a zero or the sample mean.

Finally, the release should include executable calculations and a record of editorial decisions. Corrections should identify whether underlying evidence, a coding error, or a normative scoring decision changed. A new weight policy is a method revision even if the dataset remains identical. Versioned downloads give a reader a stable object to cite and make disagreements about interpretation separable from disagreements about arithmetic.

## Limits and use

This demonstration proves a rank reversal within one deliberately simple arithmetic model. It does not evaluate existing equality indices, establish preferred weights, measure any real jurisdiction, or supply evidence for a political conclusion. Its reproducibility is exact because the inputs are defined examples, not because social measurement is free of uncertainty.

The accompanying script recreates all displayed scores and verifies the tie at four-sevenths. The practical implication is a publishing obligation: when the order changes under admissible assumptions, show the change and explain the assumptions. When components answer different questions, keep their meanings visible. A composite can summarize a declared perspective, while a transparent platform lets the reader see how that perspective shaped the result.