Limitations of Simple Additive and Multiplicative Scoring Methods

Ben Klein
San Francisco State University
Computer Science Department

February 28, 2026

Abstract

Decision making systems have relied heavily on two simple methods for aggregating criteria: Simple Additive Scoring (SAS) and Simple Multiplicative Scoring. These methods are popular for their transparency and simplicity, however they assume simple relationships between factors that often don't reflect reality. Relying on these models can lead to cost inefficiencies in low-stakes cases, and injury or death in high-stakes clinical cases. This study evaluates the effectiveness of these aggregation methods in the domains of education, Geographical Information Systems (GIS), healthcare, and measuring human development.

1 Introduction

Simple Additive Scoring (SAS, also known as Simple Additive Weighting [1] or SAW) and Simple Multiplicative Scoring (SMS, also known as Weighted Product Model [2] or WPM) systems are commonly used throughout many domains to aggregate and score multiple factors. The SAS formula can be expressed as

$$SAS = \sum_{i=1}^{n} w_i a_i \tag{1}$$

where w_i represents a weight corresponding to an attribute a_i. Similarly, SMS can be represented as:

$$SMS = \prod_{i=1}^{n} a_i^{w_i} \tag{2}$$

For most cases, the sum of all weights should equal 1. This property is called idempotency and ensures that if all inputs are equal, the output will also equal the input.

While these formulas look similar, they represent a crucial difference in logic. SAS models replaceability, allowing a surplus of one attribute to make up for a deficiency in another. In contrast, SMS

models simultaneity, where a low value in one attribute can be a harsh penalty to the resulting score. In addition, any attribute equaling 0 in the SMS model, reduces the result to zero, a property known as an annihilator. This allows the model to represent mandatory requirements.

The SAS model is widely used throughout many different domains [3], primarily due to its simplicity and intuitive nature. As a common example, grade point averages (GPA) are calculated using SAS.

Despite its popularity, SAS is often an overly simplistic model. In the following sections, this study provides a critical review of SAS and SMS in decision making systems.

2 Scoring Methodology Survey

To understand the implications of SAS and SMS in different environments, this section examines real world usage of these aggregation functions.

The most common example is Grade Point Average (GPA). In most simple SAS scored GPA systems, all classes are averaged for a score between 0.0 and 4.0. These are typically mapped such that $A = 4.0$, $B = 3.0$, $C = 2.0$, etc. [4] This method aggregates unrelated subjects into one value. The assumption is that all subjects are equally valuable and a high score in one subject can make up for a low score in another.

In healthcare, the Sequential Organ Failure Assessment (SOFA) [5] is used to track patient morbidity in the Intensive Care Unit. The model evaluates six independent organ systems - respiratory, coagulation, hepatic, cardiovascular, central nervous, and renal - assigning each a score from 0 (normal function) to 4 (severe failure). While in some cases the mean SOFA score is used, and in others the total of all scores is used, in either case the score suffers from similar issues that other SAS models face. The following section examines these issues in more detail.

In the domain of Geographic Information Systems (GIS), SAS is used (called a Weighted Linear Combination) to create landslide susceptibility maps. Ayalew et al. (2004) [6] aggregated six different factors, assigning the highest weights to slope gradient (0.2651) and lithology (0.2569). The resulting score is then used to divide the map into five classes of landslide susceptibility.

For evaluating human development across countries, the United Nations Development Programme (UNDP) uses a metric called the Human Development Index (HDI). [7] The metric uses three basic dimensions: a long and healthy life, access to knowledge, and standard of living. Until 2010, the HDI was calculated using a SAS aggregation function. The metric was found to inflate uneven scoring across the dimensions meaning, for example, a high score in life expectancy but low in knowledge would be too high. In 2010, the UNDP switched to a SMS model:

$$HDI = (I_{Life}^{1/3} \cdot I_{Education}^{1/3} \cdot I_{Income}^{1/3}) \tag{3}$$

3 Critical Review of Simple Scoring

Building on the survey in Section 2, this section examines the limitations of simple additive and multiplicative scoring. The following analysis uses the previously examples in education, healthcare, GIS, and human development metrics to pinpoint exactly where simple models break down. Through specific examples of logic failures, this study shows why simple aggregations are often insufficient for modeling complex systems.

3.1 Critical Review of Simple Scoring in Education (GPA)

In the context of education (GPA), SAS is used to build a single value to convey academic success. The result is that a low score in core subject (e.g. Math) can be made up for with a high score in an elective (e.g. Drawing). This type of unweighted aggregation can discourage students from taking difficult classes, where the student could lower their GPA while learning complex topics.

Another problem raised with the GPA example is the lack of an annihilator function. GPA alone does not show whether a student is successful because a student can do poorly on one class and still proceed without going back to ensure they learned the critical material.

The problems with GPA have been known for a long time, however. Caulkins and Larkey [8] argued in 1996, that GPA was a flawed measure and led to perverse incentives for students in class selection. In 2019, Meadows et al. [9] argued that GPA is a poor measure of success, and that alternative goals like aiming for 'self-authorship' results in more successful students. Despite the problem being identified many times, GPA is still a standard, likely due to it's simplicity.

3.2 Critical Review of Simple Scoring in Healthcare (SOFA)

The Sequential Organ Failure Assessment (SOFA) uses SAS to combine six factors into one measure of morbidity. The model assumes a high score in one organ (indicating organ failure) is equivalent to low scores across a few organs. In reality though, medical professionals don't simply take the total score. In some cases, a change of 2 or more points is used to diagnose sepsis. In other cases, the maximum value is taken to predict a patient's risk of death. The reliance on alternate interpretations of the score implies that additive aggregation alone is not sufficient to explain organ failure, and could result in misinterpretations or mistakes. [5]

In 2024, Dusaj et al. [10] looked at measures of heterogeneity between patients who exhibited scores of 6, 9 and 12. Despite similar scores and similar mortality rates, clustering analysis showed distinct groups within these subjects. The patients varied in key areas like organ failure patterns, comorbidities, infection types, and duration of organ support. A SOFA score of 6 can be made up of 426 subscore combinations and a score of 12 has 1751. These combinations reflect meaningful information that is lost in simple aggregation.

Three distinct clusters were found in the cohort (Figure 1):

Cardiovascular Failure Cluster: Typically older patients, more likely to be female, with higher rates of septic shock and bloodstream infections.

Respiratory Failure Cluster: Patients with more comorbidities and a higher likelihood of pneumonia.

Mixed Organ Failure Cluster: Generally younger patients, often experiencing immunosuppression.

Despite having similar scores, these patients require different interventions. Simple scoring in SOFA scores allow for convenience over nuance, leading medical professionals to depend on intuition rather providing helpful guidelines.

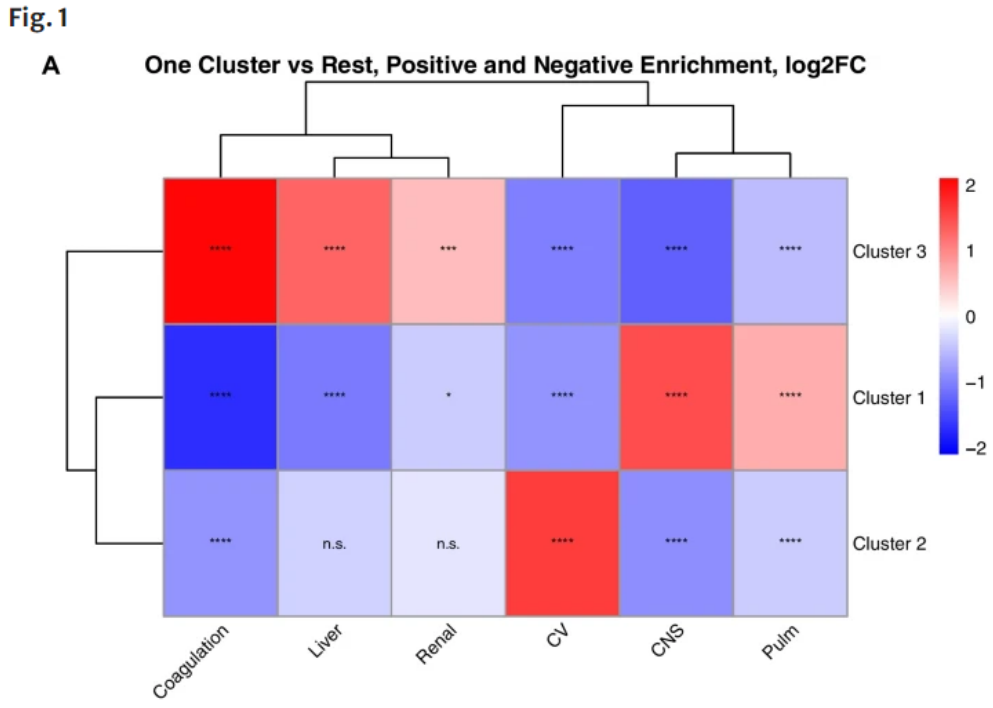

Figure 1: Cluster analysis of patients grouped by SOFA score (from Dusaj et al., 2024) [10]. Positive enrichment (red) shows the corresponding organ is failing more severely in this cluster than the other clusters. Negative enrichment (blue) shows less organ failure compared to the other clusters.

3.3 Critical Review of Simple Scoring in GIS (Landslide Susceptibility Maps)

Another example of a logic failure can be found in the SAS scoring for landslide susceptibility maps in the domain of Geographical Informational Systems (GIS). Landslides are created by non-linear combinations of factors, where some factors are mandatory.

In 2004, Ayalew et al. [6] details an aggregation of six factors to create a landslide susceptibility map, shown in Figure 2.

While in practice, these weights can provide a fairly accurate map, in edge cases, the logic fails.

Table 1 A pair-wise comparison matrix for calculating factor weights

Pair-wise comparison 9 point continuous rating scale								
Extremely	V. Strongly	Strongly	Moderately	Equally	Moderately	Strongly	V. Strongly	Extremely
Less important				Important	More Important			
1/9	1/7	1/5	1/3	1	3	5	7	9
	Aspect	Elevation	Lithology	Plan curvature	Profile curvature	Slope gradient	Factor weights	Consistency ratio (CR)
Aspect	1						0.0657	0.07
Elevation	3	1					0.1929	
Lithology	3	3	1				0.2569	
Plan curvature	1	1/5	1/3	1			0.0715	
Profile curvature	3	1	1	1	1		0.1478	
Slope gradient	3	1	1	5	3	1	0.2651	

Figure 2: A pair-wise comparison matrix seen in a 2004 paper by Ayalew et al. [6]. The slope gradient makes up about 26% of the resulting susceptibility score, meaning a flat surface could score up to about 74% landslide susceptibility rating given maximum values in other factors.

This aggregation allows for a flat surface to score up to a 74% landslide susceptibility score, which does not represent a realistic risk for a flat plane.

The other problem with this model is in double counting. For example, factors like elevation and slope are correlated, which causes the risk score to be inflated in areas where both elevation and slope are high.

3.4 Critical Review of Simple Scoring in Human Development Metrics (HDI)

In 2010, the United Nations Development Programme (UNDP) acknowledged one failure case in simple scoring. Before the 2010 changes, The Human Development Index (HDI) was calculated using the arithmetic mean (equivalent to SAS) of life expectency, education and income. [7] Figure 3 shows the comparison in HDI scores before and after 2010.

After 2010, the UNDP switched to a geometric mean aggregation function (equivalent to SMS) to penalize scores with uneven development. Before the change, a country could make up a disasterous life expectency with a high income. Another issue with the previous model, is that another year of life expectency has the same value whether the country has a life expectency of 40 years or 80 years. This fails to capture the preference that all countries should have at least a moderate score in all three key areas to be considered successful in human development.

4 Conclusions and Future Work

This review of simple scoring methods shows multiple common points of failure in logic. Simple Additive Scoring (SAS) is a model that destroys important nuance in complicated factors and treats mandatory requirements like electives. It produces poor results in GPA, GIS landslide susceptibility mapping, and HDI. Simple Multiplicative Scoring (SMS) on the other hand, can represent annihilator values, mandatory criteria and diminishing returns providing a more complex expression of factor importance. However, it works equally across all factors, meaning all scores are equally mandatory.

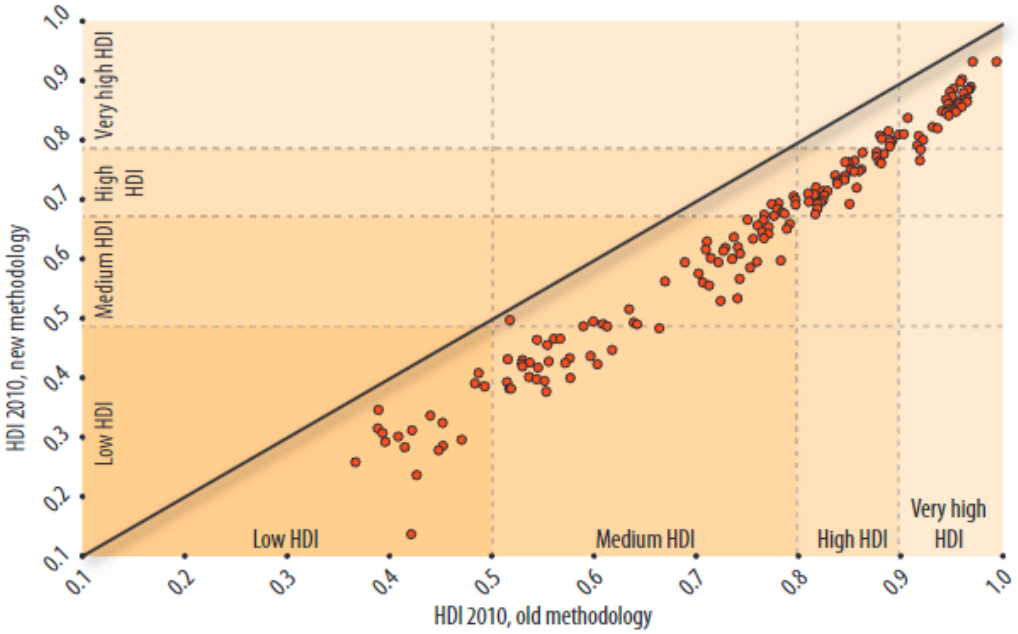

Figure 3: A comparison of the HDI factor aggregation methodolgies, before and after 2010. [7]

Future work in this area would benefit from a detailed evaluation of each scoring criteria. For example, a better model could allow an expert to decide the 'andness' and 'orness' of each factor, allowing for one factor to have a 70% degree of simultaneity and another as only 30%, while selecting which variables should function as annihilators. This kind of aggregation is a major goal of Jozo Dujmovic's graded logic and Logic Scoring of Preference (LSP). [11] More careful aggregation of factors can allow for more nuanced and precise decision making processes with better outcomes while still maintaining explainability.

References

[1] Ali Afshari, M. Mojahed, and Rosnah Yusuff. Simple additive weighting approach to personnel selection problem. *International Journal of Innovation, Management and Technology*, 1:511–515, 12 2010. doi: 10.7763/IJIMT.2010.V1.89.

[2] D K Behera, Dhiren Behera, and Soupayan Mitra. A comprehensive study of weighted product model for selecting the best laptop model available in the market. *Brazilian Journal of Operations & Production Management*, 17, 05 2020. doi: 10.14488/BJOPM.2020.017.

[3] Hamed Taherdoost. Analysis of simple additive weighting method (saw) as a multiattribute decision-making technique: A step-by-step guide. *Journal of Management Science & Engineering Research*, 6, 02 2023. doi: 10.30564/jmser.v6i1.5400.

[4] GPA Calculator - Residential Life, 2026. URL https://reslife.berkeley.edu/gpa-calculator/.

[5] Simon Lambden, Pierre Francois Laterre, Mitchell M. Levy, and Bruno Francois. The SOFA score—development, utility and challenges of accurate assessment in clinical trials. *Critical Care*, 23(1):374, November 2019. ISSN 1364-8535. doi: 10.1186/s13054-019-2663-7. URL https://doi.org/10.1186/s13054-019-2663-7.

[6] Lulseged Ayalew, Hiromitsu Yamagishi, and Norimitsu Ugawa. Landslide susceptibility mapping using gis-based weighted linear combination, the case in tsugawa area of agano river, niigata prefecture, japan. *Landslides*, 1:73–81, 01 2004. doi: 10.1007/s10346-003-0006-9.

[7] Jeni Klugman. Human development report 2010–20th anniversary edition. the real wealth of nations: Pathways to human development. *United Nations Development Programme (UNDP)*, 2010. See Technical Notes, pp. 216-217.

[8] Jonathan P Caulkins, Patrick D Larkey, and Jifa Wei. Adjusting gpa to reflect course difficulty. Technical report, Carnegie Mellon University, 1996.

[9] Lorelle A Meadows, Maura Hollister, Mary Raber, and Laura Kasson Fiss. Gpa as a product, not a measure, of success in honors, 2019.

[10] Neville Dusaj, Eleni Papoutsi, Katherine L. Hoffman, Ilias I. Siempos, and Edward James Schenck. Lost in a number: concealed heterogeneity within the sequential organ failure assessment (SOFA) score. *Critical Care*, 28(1):6, January 2024. ISSN 1364-8535. doi: 10.1186/s13054-023-04782-2. URL https://ccforum.biomedcentral.com/articles/10.1186/s13054-023-04782-2.

[11] Jozo Dujmović. *Graded Logic: Modeling Human Commonsense Reasoning in Decision-Making and AI*, volume 273 of *Intelligent Systems Reference Library*. Springer Nature Switzerland, Cham, 2025. ISBN 978-3-031-88557-0 978-3-031-88558-7. doi: 10.1007/978-3-031-88558-7. URL https://link.springer.com/10.1007/978-3-031-88558-7.