Redefining the decathlon scoring method

Wim Westera
Open University of the Netherlands
wim.westera@ou.nl

This paper investigates the multi-event scoring method that is currently being used in decathlon (athletics). It presents a summary of the paper in the March/April 2006 issue of New Studies in Athletics[1].

Decathlon scoring tables

Allroundness is the core idea of decathlon. Decathlon challenges the versatility of the competitors, which have to combine excellent physical power, explosiveness, technical (psycho-motor) skills and endurance. In order to determine total scores the separate performances in various jumping, throwing and jumping events are converted into points to allow simple addition. For this conversion the International Association of Athletics Federations (IAAF) has developed official scoring tables for each of the events. The scoring tables are the outcome of many modifications over the years to remove manifest flaws. For spectators, reporters and even decathletes the scoring method is quite mysterious. They cannot but accept the scoring outcomes indiscriminately as a fact of life. The current scoring tables are being used without modifications since 1984 and it turns out that today quite some unbalance has arisen.

Where do decathletes achieve their points?

In order to answer this question we need empirical data of decathlon events. It is essential that we consider an exemplary group. Obviously, amateur athletes, joggers and jokers may occasionally participate in a decathlon, but they are likely to show disproportional failures at certain disciplines due to poor training, lack of technical skills or insufficient versatility. For this reason we chose to use the all time top 100 decathlon ranking of the IAAF (www.iaaf.org). Figure 1 shows the distribution of average scores over the separate disciplines.

[1] Westera, W. (2006). Decathlon, towards a balanced and sustainable performance assessment method. *New Studies in Athletics*, March/April, pp. 37-48.

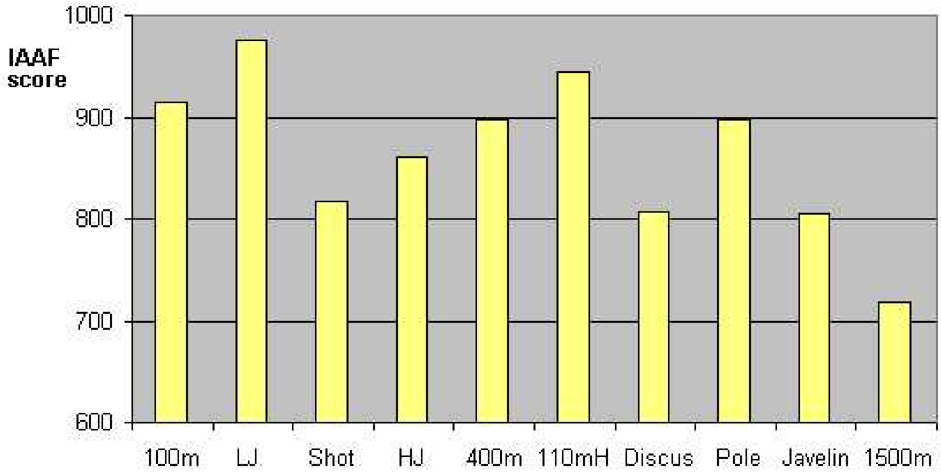

Figure 1. Average scores of the all time top 100 decathlons (IAAF, August 2005)

The athletes disproportionately seem to profit from the long jump, the 110 m hurdles and the 100 m, while - in contrast – the 1500 m, javelin, discus throw and shot put are highly unfavourable. Apparently, decathlon athletes seem to specialise in sprinting, which indeed may be regarded a common denominator of the long jump, the 110 m hurdles and the 100 m. Throwing capabilities and endurance, however, seem to be far less profitable and may even interfere with sprint performance. Clearly, the different scores in each event conflicts with the premise of allroundness and indicate the need call for a revision of the current scoring method.

The current scoring method

The current scoring tables have been set up in the 1980s after extensive discussions, negotiations and compromises, while taking into account an abundant amount of empirical evidence. Basically, the current scoring method for each discipline is covered by a mathematical expression of the type:

$$S(P) = A.(P-B)^C \qquad (1)$$

P is the performance (i.e. the achieved distance in the long jump).
S is the score (the number of assigned decathlon points).
A, B en C are event-dependent parameters that define the nature of the scoring table.
For running events (P- B) should be replaced with (B-P) because of the descending nature of performance with time. Note that the performance assessment method comprises two stages: first the performance P is measured (in units of time or distance), next the performances are converted to a score S in order to allow addition. Clearly, it is this second stage of assessment that is problematic.

Figure 2 shows the slightly progressive scoring curve for the long jump.

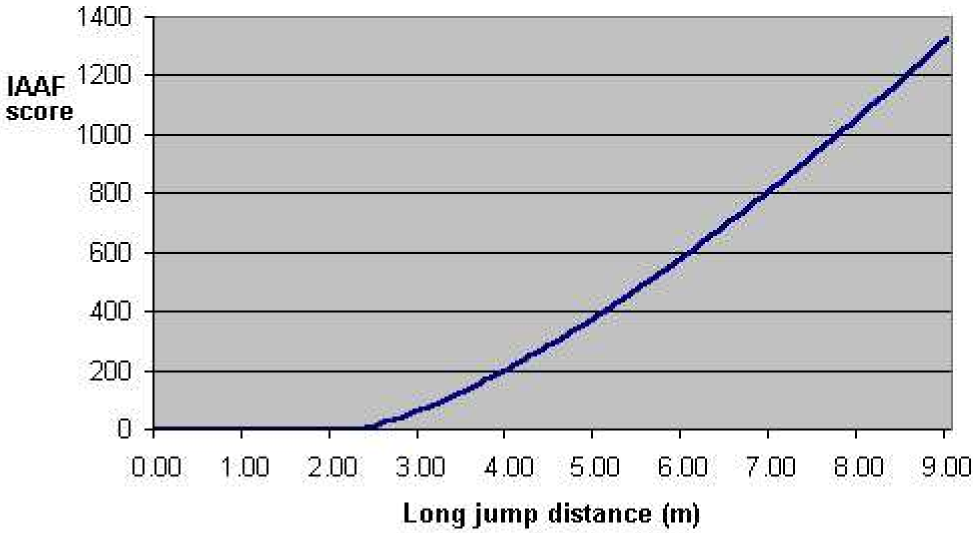

Figure 2. Current scoring curve for the men's long jump.

It uses the following values: A=0.14354, B=220 cm, C=1.40, while P is expressed in cm. Scoring curves according to equation (1) are slightly progressive, which nature is mainly determined by the power C. The underlying idea of this nonlinearity is that an improvement at low performances is much easier than an improvement at high performances. The overall scaling of the curve is determined by a parameter A. The parameter B defines a threshold value, below which no score is assigned. In case of the long jump no points are obtained when the long jump is below 220 cm. It should be noted that A, B and C are different for each discipline, for instance, for javelin (A= 10.14, B= 7.0 m, C=1.08) and for 100 m (A= 25.4347, B= 18.0 s, C= 1,81). Clearly, the current tables are pragmatic in kind and based on tradition rather than solid explanation. Consequently, some arbitrariness is involved (220 cm and not 240 cm!). Altogether, the current decathlon scoring method thus comprises a set of 10 power laws that is specified by 30 calibration parameters: A, B and C for each of the 10 events.

Premises for fair rating

For the construction of an improved scoring method we have defined the following requirements:

- allow a fair comparison between events,
- be uniform over all events (this follows from the starting points of the decathlon),
- use objective standards (distance and time measurements),
- be grounded in empirical evidence in decathlon (practical significance),
- be based on sound principles and logic (consistent, transparent and substantiated),
- be stable over time and thus possess self-stabilising characteristics,
- allow smooth transitions from the existing model (acceptability).

Devising a reference model

The ideal decathlon scoring model should indiscriminately use the same reference model for each of the events. This is a consequence of the decathlon's premise of allroundness. It follows that we would need to define 10 linear transformations to map each of the scoring curves (cf. figure 2) on to some normalised reference curve to be defined.

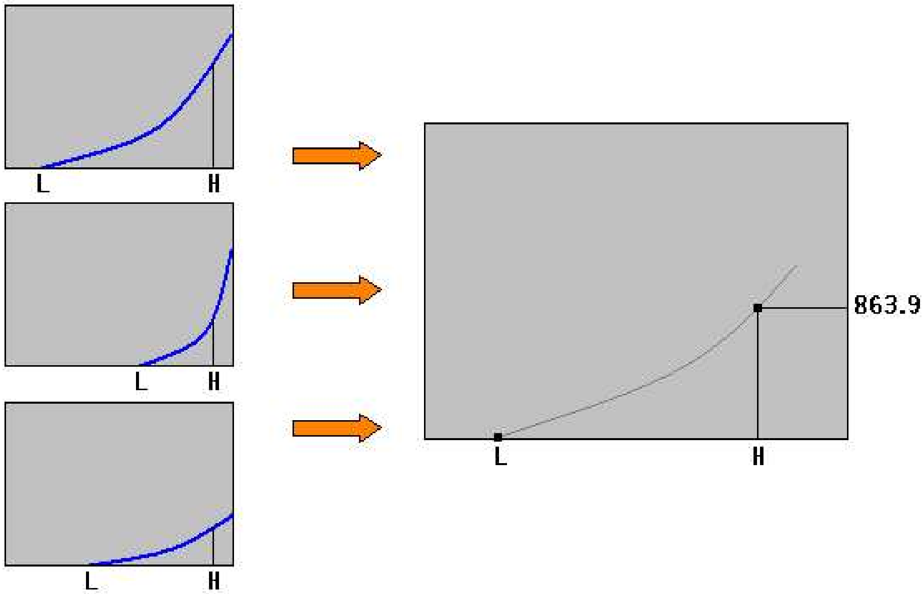

Figure 3. Transformations to a normalised reference curve

Such transformations comprise both horizontal scales (performances) and vertical scales (scores). As a first step the horizontal scales (performances) should be aligned. To achieve this (at least) two calibration points are needed, preferably at the higher end (H) and at the lower end (L) of the scale. For the high-end calibration (H) we refer to the empirical results of figure 1: indeed the data of the all time top 100 decathlons may be assumed representative for the high-end performance. The premise of allroundness implies that the average performances of this sample should be rated equally. This means that the performances listed in table 1 have the same reference performance (H):

100m	Long J	Shot P	High	400m	110mH	Discus	Pole	Javelin	1500m
10.76s	7.66m	15.47m	2.06m	48.22s	14.23s	46.92m	4.95m	64.46m	4:34.12m:s

Table 1: Average performances of all time top 100 decathlons (IAAF, August 2005)

For the low-end calibration (L) we have used the current IAAF threshold values B, while averaging over the (diverse) relative positions of the thresholds. Through these calibrations two points of the reference scoring curve have been fixed: at the low end (the threshold performance) the score would be zero; at the high end (the top 100 average) the reference score turns out to be 863.9 points (as the averaged top 100 score is 8639 points).
As a final step we have to define the progression of the reference curve. For this purpose we have elaborated 3 alternative models, which we will briefly explain below:

Model 1. Power law

In accordance with the current scoring method we may assume a power law curvature. Naturally, the power C (cf. equation 1) determines the progressive form of the scoring curve, so it follows that C>1. A simple estimate of the power C can be obtained by conforming to the 10 IAAF power parameters C that are used in the current method. When we equate the reference power C with the average of the current powers we obtain C = 1.479.

Model 2. Parabolic

Theoretically, the progressive form of the scoring curve may be associated with the role of the kinetic energy that is developed by the athlete. Along this line of thought the resulting scoring curve should be parabolic in kind, because kinetic energy is expressed as (distance/time) squared and performance is always expressed in units of distance or units of (reciprocal) time. Clearly, the parabolic model yields C=2. The same conclusion prevails when we assume that the extra score dS(P) that follows a performance improvement dP is proportional with the performance P.

Model 3: Exponential

Starting from statistics we arrive at an exponential curvature. The underlying assumption is that the distribution of performances might be approximated by the negative exponential distribution. It can be shown that this assumption is equivalent with the sensible premise that a performance increment dP causes a frequency (occurrence) change df(P) that is linearly proportional with the frequency f(P) (with coefficient λ). The exponential distribution is often associated with the survival of species in biology or similar processes that account for failures and drop outs. This process of survival has many things in common with decathlon. Consider for example high jump and pole vault, were the requested performance of athletes is stepwise incremented, until eventually all competitors have dropped out. Theoretically all decathlon events can be mapped on to this approach and thus match a regular survival pattern. In order to establish the progression of exponential curve we have set the pragmatic requirement that the exponential curve has an intermediate position between the power curve and the parabolic curve. By minimising the total squared differences between the curves, we find λ=1,602.

Outcomes

All three suggested models meet the requirements for justified rating that we have expressed before. Relevant data and formulas for these suggested models are summarised in table 2. here P is the performance, S is the score, P_0 and P_1 are reference values, A, C and λ are constants.

I. Power law $S(P)=A\,.((P-P_0)/(P_1-P_0))^C$ with A = 863.9 en C= 1.479		
II. Parabolic $S(P)=A\,.((P-P_0)/(P_1-P_0))^C$ with A = 863.9 en C= 2.000		
III. Exponential $S(P)=A\,.\,(e^{\lambda P_N}-1)/\,(e^{\lambda}-1)$ with A = 863.9 en λ=1.602		
Event	**P_0**	**P_1**
100 m	$(31.64\ s)^{-1}$	$(10.76\ s)^{-1}$
Long jump	2,60 m	7,66 m
Shot put	5.26 m	15.47 m
High jump	0.70 m	2.06 m
400 m	$(141.81\ s)^{-1}$	$(48.22\ s)^{-1}$
110 m H	$(41.85\ s)^{-1}$	$(14.23\ s)^{-1}$

Discus throw	15.95 m	46.92 m
Pole vault	1.68 m	4.94 m
Javelin throw	21.92 m	64.46 m
1500 m	$(13\text{m } 26.16 \text{ s})^{-1}$	$(4\text{m } 34.12\text{s})^{-1}$

Table 2. Summary of three alternative scoring models

Recalculation of the all time top 100 ranking according to the proposed models shows some interesting changes, although in general the changes are limited. Table 3 shows the current all time top ten as well as the ranking outcomes of the three alternative models. In the new rankings the original IAAF-ranking is indicated between parentheses.

Rank	IAAF-model		Power model		Parabolic model		Exponential model	
1	Šebrle	9026	Dvorák (2)	9232	Dvorák (2)	9468	Dvorák (2)	9777
2	Dvorák	8994	Šebrle (1)	9128	Šebrle (1)	9318	Šebrle (1)	9590
3	Dvorák	8902	Šebrle (5)	9101	Šebrle (5)	9280	Šebrle (5)	9577
4	Dvorák	8900	Dvorák (4)	9030	Šebrle (8)	9182	Smith (59)	9554
5	Šebrle	8893	Šebrle (8)	9029	Dvorák (4)	9182	Freimuth (21)	9480
6	O'Brien	8891	Dvorák (3)	9016	Freimuth (21)	9170	Šebrle (8)	9474
7	Thompson	8847	Freimuth (21)	9014	Dvorák (3)	9164	Dvorák (4)	9471
8	Šebrle	8842	Smith (59)	8988	Smith (59)	9156	Dvorák (3)	9446
9	Dvorák	8837	O'Brien (6)	8976	O'Brien (6)	9110	O'Brien (6)	9397
10	Hingsen	8832	Dvorák (9)	8953	Dvorák (9)	9081	Clay (13)	9377
	Runners-up							
13	Clay	8820						
21	Freimuth	8792						
59	Smith	8626						

Table 3 Comparison of decathlon scoring methods

Remarkably, all three models indicate a new world leader and record holder: Thomáš Dvorák's performance of 04-07-1999 in Praha outstrips Roman Šebrle's world record of 25-05-2001 in Götzis, which is unanimously ascribed rank 2. The new world record would read 9232 points using the power law, 9469 with the parabolic or 9777 for the exponential curve. Note that these scores largely exceed the current world record (9026), especially in the case of parabolic and exponential scoring due to the relative high progression of the curves at world level performances. As in the current model the absolute values of the scores are irrelevant as such, because of their arbitrariness (choosing different high end calibration points (H) would yield different scales). The average change in the top 100 ranking turns out to be 10 positions for each of the models. Biggest leap is observed for the number 59 athlete in the current ranking (Mike Smith), who may be assumed to have been greatly underrated and put at a disadvantage by the current system because of relatively poor sprinting (100 m in 11.23 s; 110 m H in 14.77). The parabolic and power method allocate Mike Smith rank 8; the exponential yields even rank 4. Likewise number 21 of the IAAF ranking (Uwe Freimuth: 11.03 s, 14.66 s) is allowed to enter the all time top 10: rank 6 (parabolic), rank 7 (power) or rank 5 (exponential). Indeed the alternative models seems to counteract the sprint bias of the current model.

In this paper we have shown that the current decathlon scoring method suffers from severe bias and produces unfair outcomes. We have demonstrated that it is possible to devise alternative scoring methods that are uniform, balanced and substantiated and that avoid the negative effects of the current method.