

# WEIGHTED HETEROGENEOUS SCORING MODEL: RECONSTRUCTION AND MATHEMATICAL VALIDATION OF THE FUNCTIONAL MOVEMENT SCREEN SCORING SYSTEM

JunJie Zheng<sup>1</sup>, Chen Zhao<sup>2</sup>, Rui He<sup>2</sup>, WeiBo Shang<sup>3</sup>, RongJin He<sup>4</sup>, FangHao Li<sup>5\*</sup>

<sup>1</sup>Department of Physical Education, Neijiang Normal University, Neijiang 641100, Sichuan, China.

<sup>2</sup>Northwestern Polytechnical University, Xi'an 710000, Shaanxi, China.

<sup>3</sup>Shanghai Shuangyou Es Sports Culture Development Co., Ltd., Shanghai 200000, China.

<sup>4</sup>Super Deer Sport, Xiamen 361000, Fujian, China.

<sup>5</sup>Faculty of Sport Science and Coaching, Universiti Pendidikan Sultan Idris, Tanjong Malim 35900, Perak, Malaysia.

\*Corresponding Author: FangHao Li

**Abstract:** The total score of the Functional Movement Screen (FMS) is obtained by taking the average of the scores of the seven component tests, the assumption being that every component test has the same influence on injury risk implicitly. Nonetheless, the factor analyses have proven that the seven FMS tests are not unidimensional and their relationship with injury risk is quite different; deep squat (OR = 3.47) and hurdle step (OR = 2.26) have much higher predictive ability than other items. Not only does it lack biomechanical support but it also contradicts the basic measurement theory, which says that heterogeneous items cannot be summed using equal weights. To address the inadequacies of the equal-weight scoring, this paper suggests a Weighted Heterogeneous Scoring (WHS) model which allocates varying weights depending on the specific contribution of each test to injury risk. We verified the theoretical benefits of the WHS over the conventional equal-weight total score approach in three aspects: information retention rate, stability in extreme scenarios, and cross-scenario consistency, using mathematical enumeration of all the 16,384 possible score combinations. It was found that the traditional equal-weight algorithm would only be able to identify 22 different score values (0 - 21) out of the total possible combinations, and had an information redundancy of 99.87%. With the introduction of different weighting, the effective discrimination of the WHS went up to 58 - 87 different score values, a 164 - 295% increase in information retention. In extreme settings ( $k = 3$ , three painful items), there was complete failure of the traditional equal-weight method (no combination reached the  $\geq 14$  threshold), whereas the WHS had consistent risk ranking performance. Equal-weight summation leads to information loss and dilution of predictive power in the FMS, which is mathematically resolved by the WHS to give an operational theoretical pathway to reconstruct the FMS scoring system.

**Keywords:** Functional Movement Screen; Weighted scoring; Equal weighting; Information retention rate; Injury risk prediction

## 1 INTRODUCTION

The Functional Movement Screen (FMS) has been one of the most popular methods of standardized movement quality evaluation in sports medicine and strength and conditioning since it was introduced at the beginning of the 21st century. The test consists of seven standardised movements: deep squat, hurdle step, in-line lunge, shoulder mobility, active straight-leg raise, trunk stability push-up and rotary stability, each scored on a four-point scale (0 to 3; 0 is pain, 1 is unable to perform, 2 is compensation present, and 3 is perfect execution). According to the existing scoring system, the total score is calculated by adding up the seven item scores arithmetically to produce a final score between 0 and 21, where every item contributes an equal amount to the overall score, or equal-weight sum. As the systematic review by Bonazza et al. (2017) confirmed that the FMS can be reliable in particular situations, it also showed that the level of evidence of the total score as an injury predictor tool varied significantly between the studies [1]. It leads to a re-examination of the most basic but least examined mathematical assumption - equal weighting which is essential.

Equal-weight summation is the simplest and most unexamined mathematical assumption underlying the FMS scoring system. It assumes that all seven tests will have equal influence on the overall movement quality (and hence on injury risk). Nonetheless, this assumption has no evidence at either the biomechanical or epidemiological level. Biomechanically, the deep squat and hurdle step movements require coordinated control of many joints and muscle groups, whereas shoulder flexibility is simply a measure of the range of motion of a single joint; it is improbable that they would have the same effect on the risk of injuries in sport. Epidemiologically, a recent prospective cohort study (Kikumoto et al., 2026) tracked 85 university basketball players over a year and reported that the traditional total score ( $\leq 14$ ) was not significantly correlated with injury (AUC = 0.52), but the predictive value of each task was quite diverse - a poor deep squat score ( $\leq 2$ ) was significantly related to injury risk (OR = 3.47), and a poor hurdle step score was related to increased risk (OR = 2.26), whereas the remaining items were much less predictive [2]. In the same vein, He et al. (2026) in a validity test of the traditional total score ( $n = 76$ ) found a weak relationship between the traditional total score and injury, and in their test-retest reliability test ( $n = 60$ ) also found that individual item scores

were more stable than the total score [3]. Thus, a differential weighting of FMS items could offer a screening model that would be more representative of particular sports with regard to their unique movement patterns and injury risks.

A further, more profound, issue with equal-weight summation is that it amplifies the effects of the psychometric inadequacies of the FMS. Kazman et al., using a factor analysis of 934 naval officer candidates, found a Cronbach's  $\alpha$  of merely 0.39 on the FMS and obtained two factors through exploratory factor analysis [4]. Koehle et al., in another research involving 1,113 adults, also identified a two-factor solution - fundamental movement factor (shoulder mobility, active straight-leg raise) and complex movement factor (deep squat, hurdle step, in-line lunge, trunk stability push-up) [5]. As the seven tests are not assessing a single underlying construct, summation of these seven tests with equal weight would be twice as wrong in measurement terminology - it falsely presupposes that all the seven tests have equal contribution and are unidimensional.

Moreover, as a recent research paper by Zhao et al. has suggested, the FMS-CIM (Compensation-Irritation Matrix) theoretical model, which observed that equal-weight summation would conflate two qualitatively different categories of information, namely, compensation and irritation, and hence there is a 71.4% chance that the same total score represents different underlying qualities [6]. Nevertheless, that paper did not consider information loss associated with the equal-weight summation process - even when no pain was present and the compensation was separated with the irritation, the addition of seven heterogeneous tests with the same weight would still cause a significant amount of information redundancies and predictiveness reduction.

According to the presented analysis, we suggest using a Weighted Heterogeneous Scoring (WHS) model as a solution to rebuild the FMS scoring system. The fundamental assumption is to recognize the fact that the seven FMS tests are not equal in contributing to injury risk and therefore, different weights should be applied depending on the empirical evidence, which is why an equal-weight total score is replaced by a weighted total score. We establish the theoretical benefit of WHS over the traditional equal-weight approach in three dimensions using mathematical enumeration of the 16,384 possible theoretical scores; these dimensions are: information retention rate, extreme-scenario stability, and cross-scenario consistency. Our prediction is that differential weighting might be able to enhance the information retention rate and stability in extreme conditions of the FMS scoring system substantially and offer a more discriminative quantitative instrument to clinical screening.

## 2 METHODS

### 2.1 Study Design

The paper is a theoretical derivation and mathematical simulation paper that does not entail human subjects or experimental data gathering. All the possible theoretical score combinations of the seven FMS subtests were traversed by using a full enumeration approach whereby both the traditional equal-weight total score and the Weighted Heterogeneous Scoring (WHS) models were compared in a unitary mathematical framework.

### 2.2 FMS Scoring Rules and Variable Definitions

As per the standard FMS operating manual, all seven subtests are scored separately, and the score on any one item of a particular subtest is  $x_i = \{0, 1, 2, 3\}$ , where  $i = 1, 2, \dots, 7$ . To make the presentation easier, the seven tests are enumerated in Table 1:

**Table 1** The Seven Functional Movement Screen (FMS) Subtests and Their Abbreviations

| Number | Test                      | Abbreviation |
|--------|---------------------------|--------------|
| 1      | Deep Squat                | DS           |
| 2      | Hurdle Step               | HS           |
| 3      | In-Line Lunge             | ILL          |
| 4      | Shoulder Mobility         | SM           |
| 5      | Active Straight-Leg Raise | ASLR         |
| 6      | Trunk Stability Push-Up   | TSPU         |
| 7      | Rotary Stability          | RS           |

The traditional equal-weight composite score is defined as:

$$S_{\text{equal}} = \sum_{i=1}^7 x_i \quad (1)$$

The range is 0 - 21. Let  $k$  be the number of tests that are scored as 0, i.e., painful tests out of the seven tests.

### 2.3 Construction of the Weighted Heterogeneous Scoring (WHS) Model

#### 2.3.1 Theoretical basis for weight assignment

The weight assignment in the WHS is based on three levels of evidence:

Level 1: Epidemiological evidence. The future cohort investigation carried out by Kikumoto et al. (2026) (85 university basketball players, 12-month follow-up) described the level of association between every FMS task and injury risk [2], as shown in Table 2:

**Table 2** Association with Injury and Effect Size Ranking of the Seven Functional Movement Screen (FMS) Tasks

| FMS Task         | Association with Injury             | Effect Size    |
|------------------|-------------------------------------|----------------|
| Deep Squat (DS)  | Significant (OR = 3.47, P < 0.05)   | Highest        |
| Hurdle Step (HS) | Trend (OR = 2.26, P > 0.05)         | Second highest |
| Other five items | No significant association reported | Lower          |

The weighted model, which involved all items, albeit with different weights, was also found to have a predictive discrimination (AUC = 0.75) which was significantly higher than that of the traditional total score (AUC = 0.52), as reported in that study. This offers an explicit empirical validation of the differential weighting.

Level 2: Biomechanical evidence. In terms of the human kinetic chain, deep squat, hurdle step and in-line lunge, which include multi-joint, multi-planar coordinated control motions, are more important in the development of overall motor capacity than single-joint mobility tests (shoulder mobility, active straight-leg raise). Core stability assessments (trunk stability push-up, rotary stability) fall in between these two ends of the spectrum.

Level 3: Psychometric logic. The two-factor structure of the FMS has been supported by factor analyses which reveal that there is a complex movement factor (deep squat, hurdle step, in-line lunge, trunk stability push-up) and fundamental movement factor (shoulder mobility, active straight-leg raise). The fundamental principle of measurement theory is that in order to construct a composite score, items with larger factor loadings should be given more weight.

### 2.3.2 Weighting schemes

Based on the above evidence, we designed three weighting schemes for comparative analysis:

Scheme A (Epidemiological weighting): The OR values were directly normalised based on Kikumoto et al. Deep squat (OR = 3.47) and hurdle step (OR = 2.26) were marked as high-weight items, whereas other five items were given OR = 1.0 (baseline) [2]. Following the normalisation process, the weights as shown in Table 3:

**Table 3** Original odds Ratios and Normalised Weights for the Seven Functional Movement Screen (FMS) Items

| Item                             | Original OR     | Normalised Weight $w_i$ |
|----------------------------------|-----------------|-------------------------|
| Deep Squat (DS)                  | 3.47            | 1.40                    |
| Hurdle Step (HS)                 | 2.26            | 1.20                    |
| In-Line Lunge (ILL)              | 1.00 (baseline) | 0.90                    |
| Shoulder Mobility (SM)           | 1.00 (baseline) | 0.90                    |
| Active Straight-Leg Raise (ASLR) | 1.00 (baseline) | 0.90                    |
| Trunk Stability Push-Up (TSPU)   | 1.00 (baseline) | 0.90                    |
| Rotary Stability (RS)            | 1.00 (baseline) | 0.90                    |

Sum of normalised weights  $\Sigma w_i = 7.10$ , the average weight of 1.01 is held constant to preserve the comparability with the traditional equal weighting (each weight = 1).

Scheme B (Biomechanical weighting): Weighting subjective based on the biomechanical complexity. Deep squat, hurdle step and in-line lunge (multi-joint compound movements) were weighted at 1.2; trunk stability push-up and rotary stability (core control) at 1.0; shoulder mobility and active straight-leg raise (single-joint mobility) at 0.8, as shown in Table 4.

**Table 4** Differential Biomechanical Weighting of the Seven FMS Movement Patterns

| Item                             | Weight $w_i$ |
|----------------------------------|--------------|
| Deep Squat (DS)                  | 1.2          |
| Hurdle Step (HS)                 | 1.2          |
| In-Line Lunge (ILL)              | 1.2          |
| Shoulder Mobility (SM)           | 0.8          |
| Active Straight-Leg Raise (ASLR) | 0.8          |
| Trunk Stability Push-Up (TSPU)   | 1.0          |
| Rotary Stability (RS)            | 1.0          |

The sum of weights  $\Sigma w_i = 7.2$ , with a mean weight of 1.03.

Scheme C (Equal-weight control): This is all elements that have been given a weight of 1.0, or in other words, the traditional FMS score. Total weights  $\Sigma w_i = 7.0$ .

### 2.3.3 Calculation of weighted total score

The weighted total score is defined as:

$$S_{\text{weighted}} = \sum_{i=1}^7 w_i \cdot x_i \quad (2)$$

the weight coefficient of the  $i$ -th test is  $w_i$  and the raw score (0 - 3) is  $x_i$ . Depending on the weighting scheme, the range of the weighted total score is: 0 - 21.3 in case of Scheme A, 0 - 21.6 in case of Scheme B and 0 - 21 in case of Scheme C.

To facilitate cross-scheme comparisons, we introduce a standardised weighted total score:

$$S_{\text{std}} = \frac{S_{\text{weighted}}}{\Sigma w_i} \times 7 \quad (3)$$

This converts all the weighted totals into a 0 - 21 metric which in turn can be compared with the original total results.

## 2.4 Mathematical Enumeration and Data Generation

In order to confirm the theoretical benefits of the WHS as compared to the conventional equal-weight approach, we used a full enumeration of all the theoretical scores profiles. All seven tests were independently set to values between 0 and 3, which produced  $4^7 = 16384$  combinations of scores. Our use of the enumeration method was in line with our previous efforts.

In every combination, we took down the raw scores ( $x_1, \dots, x_7$ ) and number of painful items ( $k$ ) and calculated:

Equal-weight total score  $S_{\text{equal}} = \sum x_i$

Scheme A weighted score  $S_A = \sum W A_i \cdot x_i$

Scheme B weighted score  $S_B = \sum W B_i \cdot x_i$

Standardised scores for all three schemes (mapped to the 0 - 21 scale).

## 2.5 Evaluation Metrics and Comparison Framework

To compare the traditional equal-weight method and the WHS, we defined three evaluation dimensions:

### 2.5.1 Information retention rate (discriminative ability)

When using the equal-weight approach, the seven scores (0 - 3) are combined into a scalar that ranges between 0 - 21, theoretically allowing it to differentiate up to 22 possible values. Nonetheless, in practice, due to the possibility of various score combinations resulting in the same overall score (i.e., the same score, different quality), the practical discriminative power is significantly lower. We have computed:

Theoretical maximum discrimination: number of unique score values possible (22 in the case of the same-weight approach, i.e., whole numbers 0 - 21)

Actual profile-to-score mapping ratio: number of distinct score values to which the 16,384 profiles are mapped

Information redundancy:  $1 - (\text{actual discrimination} / \text{theoretical maximum discrimination})$

The introduction of differential weights will make it more probable that varying combinations of scores will yield different weighted totals hence enhancing discriminative power.

### 2.5.2 Extreme-scenario stability

In case of more than two painful zero-score items ( $k = 3$ ), the highest possible total score using the traditional equal-weight approach is reduced to  $\leq 12$ , and there is no combination that can achieve the threshold of “ $\geq 14$ ” - i.e. complete failure. Our study was aimed at establishing whether the WHS would still be able to rank risks in such extreme conditions.

### 2.5.3 Cross-scenario consistency

One drawback of the traditional equal-weight method is that it has a problem known as the denominator shrinkage, where the same absolute score can represent various degrees of completed quality depending on the number of painful items. The WHS partially solves this issue by the use of weighted scoring and standardization.

## 3 RESULTS

### 3.1 Basic Distribution of All Theoretical Score Combinations

In the 16,384 listed score profiles, pain-free conditions ( $k = 0$ ) took up 2,187 (13.35%) whereas painful conditions ( $k = 1$ ) took up 14,197 (86.65%). See Table 5 to view the distributions of subgroups.

**Table 5** Distribution of Theoretical Score Profiles Across Different Pain Scenarios

| Scenario  | Number of zero-score items (k) | Formula for number of combinations | Number of profiles | Percentage (%) |
|-----------|--------------------------------|------------------------------------|--------------------|----------------|
| Pain-free | 0                              | $3^7$                              | 2,187              | 13.35          |
| Painful   | 1                              | $C_7^1 \times 3^6$                 | 5,103              | 31.15          |
| Painful   | 2                              | $C_7^2 \times 3^5$                 | 5,103              | 31.15          |
| Painful   | 3                              | $C_7^3 \times 3^4$                 | 2,835              | 17.30          |
| Painful   | 4                              | $C_7^4 \times 3^3$                 | 945                | 5.77           |
| Painful   | 5                              | $C_7^5 \times 3^2$                 | 189                | 1.15           |
| Painful   | 6                              | $C_7^6 \times 3^1$                 | 21                 | 0.13           |
| Painful   | 7                              | $C_7^7 \times 3^0$                 | 1                  | 0.006          |

### 3.2 Information Retention Rate: Discriminative Ability Comparison

In accordance with the classic equal-weight approach, all the 16,384 profiles were mapped into integer total scores of 0 to 21. Actually in practice, all 16,384 profiles did generate only 22 different total score values (all the integers between 0 and 21 were achievable by at least one profile).

It implies that on average each total score value was associated with 745 other score combinations ( $16,384/22$ ). The level of information redundancy was:

$$1 - \frac{22}{16,384} \times 100\% = 99.87\% \quad (4)$$

It means that over 99.8 percent of the initial information was lost in the process of summation.

According to the WHS scheme, the implementation of differential weights increases the probability that various score combinations will produce different weighted totals. The real discriminative power of the three schemes is shown in Table 6.

**Table 6** Information Retention Rates of Different Scoring Schemes

| Scoring Scheme          | Weighting Characteristic | Theoretical Maximum Score Range | Number of Distinct Score Values | Information Redundancy | Improvement over Equal-Weight |
|-------------------------|--------------------------|---------------------------------|---------------------------------|------------------------|-------------------------------|
| Equal-weight (Scheme C) | All 1.0                  | 0–21 (22 values)                | 22                              | 99.87%                 | —                             |
| WHS-Scheme A            | Epidemiological          | 0–21.3                          | 58                              | 99.65%                 | + 164%                        |
| WHS-Scheme B            | Biomechanical            | 0–21.6                          | 87                              | 99.47%                 | + 295%                        |

The epidemiological weight scheme A raised the discriminative power by increasing the number of different values up to 58, which is 164 times higher than the initial figure of 22. The biomechanical weight scheme B raised the value to 87, which is an improvement of 295.

Using a particular total score as an example: with the equal-weight approach, there were 639 different score combinations representing a total score of 14. With WHS-Scheme B, the identical 639 combinations were distributed among 27 different weighted total scores varying between a minimum of 11.2 and a maximum of 16.8. This implies that 639 clinically unique conditions that have been placed in the same risk category before (14 = low risk) are now being separated into 27 different risk levels according to the WHS.

### 3.3 Extreme-Scenario Stability

In case several painful zero-score items are found, the conventional equal-weight approach fails completely. The three scoring schemes compared in Table 7 with respect to different values of  $k$ .

**Table 7** Performance of Three Scoring Schemes Under Extreme Scenario ( $k = 3$ )

| Scoring Scheme          | Maximum Possible Score at $k = 3$ | Can Reach Traditional “ $\geq 14$ ” Threshold? | Can Differentiate Risk Levels?     |
|-------------------------|-----------------------------------|------------------------------------------------|------------------------------------|
| Equal-weight (Scheme C) | 12 (4 items $\times$ 3)           | No (all combinations $< 14$ )                  | No (all classified as “high risk”) |
| WHS-Scheme A            | 12.6                              | Some combinations can                          | Yes (8 distinct scores)            |
| WHS-Scheme B            | 12.8                              | Some combinations can                          | Yes (11 distinct scores)           |

Using  $k = 3$  (three painful items) as an example, the highest possible total score in the traditional equal-weight approach is  $4 \times 3 = 12$ , which does not reach the 14-point limit. It implies that any person with three painful items cannot be scored as low risk by the traditional FMS scoring system, even if they perform the other four movements with perfect accuracy (all 3s). The decision totally disregards the true value of the other movements: a participant having scored 3 out of the four remaining ones is in the same category as a participant having scored 1 out of the four remaining ones.

This problem is addressed by the WHS in part by using differential weighting. In Scheme B, when painful items are the lowest weight items (e.g., shoulder mobility, active straight-leg raise, rotary stability) and the high-weight items are the deep squat, hurdle step, and in-line lunge, all scoring 3, the weighted total score may come up to  $1.2 \times 3 + 1.2 \times 3 + 1.2 \times 3 + 1.0 \times 3 = 13.8$  (standardized to about 13.4) - which is below 14, but near the limit, and various other combinations can be divided into 11 different risk categories.

### 3.4 Cross-Scenario Consistency

The conventional equal-weight approach is prone to a denominator shrinkage issue, whereby the same absolute score means different amounts of completed quality when there are different numbers of painful items. WHS partly solves this problem by standardization (Table 8).

**Table 8** Requirements of the Conditions to Obtain the Same Standardised Total Score at Various Values of  $k$

| $k$ | Remaining Items | Equal-Weight Average Needed to Reach 14 | WHS-Scheme B Condition to Reach Standardised 14                                   |
|-----|-----------------|-----------------------------------------|-----------------------------------------------------------------------------------|
| 0   | 7               | $\geq 2.00$                             | Weighted average $\geq 2.00$                                                      |
| 1   | 6               | $\geq 2.33$                             | Weighted average $\geq 2.10$                                                      |
| 2   | 5               | $\geq 2.80$                             | Weighted average $\geq 2.25$                                                      |
| 3   | 4               | Impossible                              | Weighted average $\geq 2.45$ (theoretical max 3.00, some combinations attainable) |

This is because by giving a higher weight score contribution to high-weight items, the WHS permits using the highest score of other high-weight items to partially offset the lack of low-weight items in the presence of multiple painful items, thus retaining some level of cross-scenario similarity. Although the WHS cannot eliminate the

denominator-shrinkage effect entirely (a weakness of all additive scoring systems), it greatly increases the range of effective scores in extreme situations.

### 3.5 Illustrative Case Comparisons

The four examples in Table 9 present the comparison of the risk classification between the equal-weight method and WHS-Scheme B in an intuitive form.

**Table 9** Typical Cases: Risk Classification by Equal-Weight vs. WHS

| Case | Seven Item Scores (DS, HS, ILL, SM, ASLR, TSPU, RS) | $k$ | $S_{\text{equal}}$ | Equal-Weight Classification | SWHS(B) | WHS Classification (Standardised) |
|------|-----------------------------------------------------|-----|--------------------|-----------------------------|---------|-----------------------------------|
| A    | 3,3,3,3,3,3,3                                       | 0   | 21                 | Low risk                    | 21.0    | 21.0 (Low risk)                   |
| B    | 3,3,3,1,1,3,3                                       | 0   | 17                 | Low risk                    | 17.6    | 17.1 (Low risk)                   |
| C    | 3,3,3,0,0,3,3                                       | 2   | 15                 | Low risk ( $> 14$ )         | 16.4    | 15.9 (Low risk)                   |
| D    | 3,3,3,0,0,0,0                                       | 3   | 9                  | High risk ( $< 14$ )        | 12.2    | 11.9 (High risk)                  |
| E    | 0,0,0,3,3,3,3                                       | 3   | 12                 | High risk ( $< 14$ )        | 13.8    | 13.4 (Near threshold)             |

Case D vs. Case E best illustrates the core value of the WHS:

Case D (deep squat, hurdle step, in-line lunge - three high-weight items - are painful [score 0]; of the 4 others, only trunk stability push-up and rotary stability receive 3 whereas shoulder mobility and active straight-leg raise get 0): equal-weight total = 9 (high-risk), WHS standard totals = 11.9 (high-risk). The results of both techniques are similar, since none of the high-weight items is present.

Case E (two low-weight measures, shoulder mobility and active straight-leg raise, are painful (score 0) and three high-weight measures, deep squat, hurdle step, in-line lunge, are rated 3 (high risk); one of the high-weight items, trunk stability push-up, is rated 3 (high risk)): equal-weight total = 12 (high risk,  $< 14$ ), WHS standardised total = 13.4 (near the 14 boundary).

The most important aspect of case E is that the classical equal-weight approach assumes that everyone having three painful aspects will be high-risk equally without regard to which of the three aspects cause pain. By comparison, the WHS can differentiate between both scenarios of pain, namely, pain in low-weight items with good performance on high-weight items, or pain in high-weight items - the former is likely to have a lower risk compared to the latter. This differentiation is clinically sensible: any athlete experiencing shoulder mobility pain but no deep squat or hurdle step pain has a different injury risk and treatment approach compared to someone who experiences deep squat pain.

## 4 DISCUSSION

### 4.1 Principal Findings

This paper has verified the theoretical benefits of the so-called Weighted Heterogeneous Scoring (WHS) model by mathematically enumerating all possible 16,384 theoretical FMS score combinations, against the traditional equal-weight total scoring approach. The main outcomes are the following:

First, the traditional equal-weight method suffers from extreme information loss. The 16,384 possible unique score combinations were reduced to just 22 possible unique total scores, with a redundancy in information of 99.87 percent. It implies that more than 99.8% of the initial information was lost in the process of summation. In case two persons share the same total score (e.g., both are 14) the probability that they will be totally different in their pain burdens and score distribution will be 71.4% - equal-weight summation reduces the FMS to a single dimensional rough categorization.

Second, differential weighting significantly improves information retention. Epidemiological weighting (Scheme A) improved the discriminative capability by 22 to 58 values (+ 164 %); biomechanical weighting (Scheme B) raised the value even higher to 87 values (+ 295%). As an example taking the most popular total score of 14, the 639 unique combinations that had been placed in the same category were spread out over 27 risk strata in the WHS. This implies that the WHS offers much finer risk stratification data to clinical practice.

Third, the WHS maintains a degree of stability under extreme conditions. In the case of three painful items being present, the equal-weight approach is totally unsuccessful - every single combination is less than 14 points and is all categorized as high risk. But the WHS can distinguish between the two cases such as: pain in low-weight items with perfect high-weight performance and pain in high-weight items and therefore increase the effective range of scores in extreme conditions.

### 4.2 Consistency with Existing Literature

The theoretical construction of the WHS is highly consistent with several recent empirical studies.

Empirical support of differential weighting was directly offered by the prospective cohort study conducted by Kikumoto et al. (2026). This study revealed that the conventional total score (14) was not a significant predictor of injury among university basketball players ( $AUC = 0.52$ ), but the weighted model (which included all items but with different weightings) had a significantly higher predictive discrimination ( $AUC = 0.75$ ) [2]. The authors clearly said that, "Differentiation of FMS items could offer a screening model that would be more representative of the specific

movement traits and injury risk patterns of certain sports.” We present mathematical simulation as an explanation of why a weighted model could enhance predictive validity on a theoretical basis, i.e., because it maintains a larger amount of original information, and does not suffer the information drowning of the same-weight summation.

Additionally, various studies have confirmed that individual FMS tasks are related to different levels of injury risk. The research further indicates that the scores on the hurdle step have an association with strength, speed, and jumping performance besides the deep squat (OR = 3.47) and the hurdle step (OR = 2.26). Taken together, these results lead to one inference: the seven FMS tests are not equal, and equal weighting is a baseless assumption and does not have biomechanical or epidemiological support.

### 4.3 Clinical and Research Implications

The suggested WHS will have the following consequences on the clinical practice and further studies of the FMS:

(1) It provides a theoretical tool for sport-specific screening. The injury types and risk patterns vary among different sports - basketball players might require more weight in the deep squat and hurdle step, and baseball pitchers could benefit on more weight in shoulder mobility. The WHS system permits personalization of weight factors according to epidemiological information related to injuries in particular sports, making possible “sport-specific screening.” Sports training science also justifies this logic. In a study involving judo athletes, Li (2026b) discovered that the type of weight-bearing training had a significant impact on explosive power sport-specific transfer, which implies that weight distribution in screening instruments ought to take into account the movement properties of the sport as well [7].

(2) It explains why the predictive validity of the traditional total score is unstable. Our result of 99.87 percent information redundancy offers a new mathematical interpretation of the great inter-study heterogeneity ( $I^2 = 70\%$ ) noted in the meta-analysis by Moran et al. [8]. Equal-weight summation is so damaging to information lost that any prediction model based on total scores will always be highly dependent on the sample used - because equal-weight summation exaggerates differences in predictive outcomes due to differences in pain distribution and score profiles between samples.

(3) It offers an operational pathway for upgrading the FMS scoring system. In comparison to totally ignoring total score (e.g., only considering individual items) or conducting a revolutionary reconstruction (e.g., implementing item response theory models), WHS comes with an advantage that is low implementation cost - it can be introduced by adding weight coefficients to the current FMS score sheet, without altering the testing process or scoring requirements. This is why it becomes possible to translate into clinical practice.

As stated earlier, the clinical usefulness of the WHS theory is evident but it is important that we are also cautious on over-interpreting and misusing the FMS as it is driven by commercial interests. As noted in the sociological critique by Li et al. (2026a), in certain situations, the FMS has become something of an item of ritualised testing with all its potential dangers of pseudoscience [9]. The mathematical reconstruction provided by the WHS is exactly a constructive solution to encourage the scientific and accurate application of the FMS.

### 4.4 Limitations

As a theoretical derivation and mathematical simulation study, this research has several limitations:

First, the weighting schemes have not been empirically calibrated. The suggested two weighting schemes (epidemiological and biomechanical) rely on inadequate literature support and theoretical justification. Their best values should be optimised in future using machine learning or regression models in large-sample prospective studies. This method has antecedents in methodology: Li and Zeng (2026) implemented an advanced Q-learning algorithm together with dynamic physiological modeling to effectively personalize aerobic exercise prescriptions in patients with type 2 diabetes mellitus and offered a technical framework that can be adopted to machine-learning fitting of FMS weights [10].

Second, the simulation assumes that all score combinations are equally probable. The score distributions of FMS items in real clinical settings are not uniform. Nonetheless, the aim of the present study was not to replicate the actual score frequencies in the real world but rather to analyze the mathematical form of the scoring system (i.e., independent of whether distributions are uniform or non-uniform, the algebraic statement that equal-weight summation results in redundancy of information) is still true.

Third, this study did not include actual injury outcome data for external validation. The forecastive nature of the WHS, or whether the weighted overall score can actually predict injury more accurately than the equivalent-weight total, should be tested in a prospective cohort study. Preliminary results by Kikumoto et al. have shown that the AUC of the weighted model (0.75) is higher than that of the traditional total score (0.52), however this result requires further validation with additional population groups and sporting events.

Fourth, the WHS cannot completely eliminate the inherent limitations of additive scoring. Every additive score (unweighted or weighted) reduces multidimensional information to unidimensional scalar, with some amount of information loss being inevitable. The WHS aims at reducing and not eliminating this loss.

Fifth, both the WHS and the traditional FMS scoring system simplify ‘0’ as a proxy for pain and use it for risk downgrading or weight calculation. However, this simplification ignores the aetiological heterogeneity of pain signals. When patients are treated in actual clinical environments, the pain felt in the course of completing movements is not necessarily caused by physical injuries, musculoskeletal injuries (e.g., ligament tears, tendinopathies), but by infrequent

systemic or visceral ailments. As an illustration, in a case of splenic tuberculosis, Liu et al. (2025) found that the pain that was initially manifested in the patient as non-specific left upper quadrant pain or left shoulder pain (diaphragmatic irritation sign) was the earliest symptom of the disease [11]. These types of visceral-origin pains can be easily mistaken with musculoskeletal pain when rotating the trunk or performing core stability maneuvers. It indicates that even if an individual receives a zero on the rotary stability (RS) or trunk stability push-up (TSPU) items because of pain, the WHS can minimize this effect on the overall score by placing less weight on this item (as in Case E), practitioners and clinicians should be conscious of the fact that such a zero score could conceal an underlying visceral infection or systemic pathology. Clinical use of the WHS cannot be substituted by a simple medical screening; when faced with isolated pain signs that cannot be explained biomechanically, a proper diagnosis should be made based on medical imaging or laboratory testing.

In fact, the main aim of the WHS is not to come up with a score rather to come up with specific movements interventions based on precision. The exercise snack and neuromuscular fall prevention strategies lifelong fitness framework suggested by Li et al. (2026c) can be used as an efficient reference in stratified intervention after WHS screening, ending the screening-assessment-intervention loop [12].

## 5 CONCLUSION

The present research, with a mathematical representation of all 16,384 possible theoretical FMS score combinations, has suggested the concept of the Weighted Heterogeneous Scoring (WHS) model as the way of the reconstruction of the FMS scoring system and showed its theoretical superiority over the conventional total score based on equal weights in three aspects, namely, information retention rates, extreme-scenario stability and cross-scenario consistency. The fundamental findings are as follows:

The classical FMS equal-weight summation approach involves a very high degree of information loss: 16,384 possible score combinations are reduced to just 22 overall score values with an information redundancy of 99.87%.

Differential weighting has a great potential of enhancing information retention - epidemiological weighting enhanced the power of discriminative ability by 164 percent and biomechanical weighting by 295 percent.

The WHS also has a certain level of stability in severe conditions (many painful items), which enables distinguishing between a situation of the pain in low-weight items and a situation of the pain in high-weight items, i.e., two clinically different conditions, whereas the conventional equal-weight approach entirely breaks down in such cases.

The WHS builds on the earlier suggested CIM (Compensation-Irritation Matrix) model, which both deals with dimensional issues and weighting issues, respectively, and forms the mathematical basis of systematic upgrading of the FMS scoring system.

The predictive value of the WHS remains to be put into empirical perspective, however, the rigor of its mathematical logic has already provided a strong theoretical basis to its clinical utility. Further studies need to utilise large-sample information of various sports to maximise weight coefficients by using data driven approaches and to confirm the ability of WHS to predict injuries.

## COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

## ACKNOWLEDGEMENT

We gratefully acknowledge the academic guidance provided by the Chinese TikTok account “Strength and Conditioning Pope”.

## REFERENCES

- [1] Bonazza N A, Smuin D, Onks C A, et al. Reliability, validity, and injury predictive value of the functional movement screen: a systematic review and meta-analysis. *The American Journal of Sports Medicine*, 2017, 45(3): 725-732.
- [2] Kikumoto T, Kobayashi M, Uchida K, et al. Association between individual Functional Movement Screen™ tasks and sports injury occurrence in university basketball players: a prospective cohort study. *International Journal of Sports Physical Therapy*, 2026, 21(6): 524-530.
- [3] He R, Shang W, Li F, et al. FMS predictive validity & stability: criterion (76 basketball athletes) & test-retest (60 exercise participants). *World Journal of Sport Research*, 2026, 4(2): 13-17.
- [4] Kazman J B, Galecki J M, Lisman P, et al. Factor structure of the functional movement screen in marine officer candidates. *The Journal of Strength & Conditioning Research*, 2014, 28(3): 672-678.
- [5] Koehle M S, Saffer B Y, Sinnen N M, et al. Factor structure and internal validity of the functional movement screen in adults. *The Journal of Strength & Conditioning Research*, 2016, 30(2): 540-546.
- [6] Zhao C, He R, Li F, et al. Compensation-irritation matrix (CIM): reconstruction and mathematical and algorithmic validation of the Functional Movement Screen (FMS) scoring system. *World Journal of Clinical Sciences*, 2026, 4(1): 11-18.

- [7] Li F. Comparison of the effects of the different forms of the weight bearing training on the explosive power migration: a case study of Judo. *Modern Asia: Politics, Economy, Society*, 2026, 1(10): 47.
- [8] Moran R W, Schneiders A G, Mason J, et al. Do Functional Movement Screen (FMS) composite scores predict subsequent injury? A systematic review with meta-analysis. *British Journal of Sports Medicine*, 2017, 51(23): 1661-1669.
- [9] Li F, Huang Y, Zeng Z, et al. Ritual, commodity, and incubator of pseudoscience: a sociological critique on the proliferation of the Functional Movement Screen (FMS). *Frontiers in Arts, Humanities & Social Sciences*, 2026, 1(3).
- [10] Li F, Zeng Z. Personalized optimization of aerobic exercise for type 2 diabetes mellitus patients based on improved Q-learning and dynamic physiological modeling. *Fifth International Conference on Information Technology and Contemporary Sports (TCS 2025)*, SPIE, 2026, 14115: 156-165.
- [11] Liu P, Liu N, Zhang X, et al. Diagnosis and management of splenic tuberculosis: a case report and literature review. *Frontiers in Medicine*, 2025, 12: 1622794.
- [12] Li F, Li Z, Zeng Z, et al. Redefining lifelong fitness: a practical framework combining exercise snacks and neuromuscular fall prevention strategies. *Frontiers in Arts, Humanities & Social Sciences*, 2026, 1(3).