Lolli, L, Johnson, A, Monaco, M, Di Salvo, V and Gregson, W

Relative Skeletal Maturity and Performance Test Outcomes in Elite Youth Middle Eastern Soccer Players

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Lolli, L, Johnson, A, Monaco, M, Di Salvo, V and Gregson, W (2022) Relative Skeletal Maturity and Performance Test Outcomes in Elite Youth Middle Eastern Soccer Players. Medicine and Science in Sports and Exercise, 54 (8). pp. 1326-1334. ISSN 0195-9131

Relative Skeletal Maturity and Performance Test Outcomes in Elite Youth Middle Eastern Soccer Players

Heading title: Skeletal maturity and physical performance

Authors: Lorenzo Lolli[1,2], Amanda Johnson[3], Mauricio Monaco[4], Valter Di Salvo[1,5], Warren Gregson[1,2]

1 Aspire Academy, Football Performance & Science Department, Doha, Qatar
2 Football Exchange, Research Institute of Sport Sciences, Liverpool John Moores University, Liverpool, UK
3 Faculty of Health, Psychology & Social Care, Health Sciences Department, Manchester Metropolitan University, Manchester, UK
4 National Sports Medicine Program, Aspetar Orthopaedic and Sports Medicine Hospital, Doha, Qatar
5 Department of Movement, Human and Health Sciences, University of Rome "Foro Italico", Rome, Italy

Corresponding Author:
Lorenzo Lolli, Football Performance and Science Department, Aspire Academy, PO Box 22287, Doha, State of Qatar.
e-mail: Lorenzo.Lolli@aspire.qa

Abstract

Purpose: To explore the influence of differences in relative skeletal maturity on performance test outcomes in elite youth soccer players from the Middle East. **Methods:** We integrated skeletal age and performance assessments using mixed-longitudinal data available for 199 outfield players (chronological age range: 11.7 to 17.8 yr) enrolled as academy student-athletes (annual screening range: 1 to 5 visits). Skeletal age was determined as per the Tanner-Whitehouse II (TW-II) protocol. Relative maturity was calculated as the difference (Δ) between TW-II skeletal age *minus* chronological age. Performance test outcomes of interest were 10-m sprinting, 40-m sprinting, countermovement jump (CMJ) height and maximal aerobic speed (MAS). Separate random-effects generalized additive models quantified differences in performance test outcomes by relative skeletal maturity. Estimated differences were deemed practically relevant based on the location of the confidence interval (95%CI) against minimal detectable change values for each performance test outcome. **Results:** For 40-m sprinting, differences of +0.51 s (95%CI, +0.35 to +0.67 s) and +0.62 s (95%CI, +0.45 to +0.78 s) were practically relevant for relative maturity status of Δ = -1.5 yr versus Δ = +0.5 and Δ = +1 yr, respectively. For CMJ height, a difference of -8 cm (95%CI, -10 to -5 cm) was practically relevant for Δ = -1.5 yr versus Δ = +1 yr relative maturity status comparison. Effects for 10-m sprinting and MAS were unclear. **Conclusion:** Integration of skeletal age and performance assessments indicated conventional maturity status classification criteria were inconsistent to inform player development processes in our sample. Between-player differences in test performance may depend on a substantial delay in skeletal maturation (Δ ≤ -1.5 yr) and the performance outcome measure.

Keywords: skeletal age, performance, soccer, CMJ, sprint, youth, maturity

1. Introduction

Competing at an elite level in soccer requires players to be proficient in a number of physical performance attributes (1). This includes high levels of aerobic fitness, the ability to sprint, anaerobic power, strength and flexibility (1). In elite youth soccer, physical performance assessment therefore represents an important element relevant to talent identification and development processes generally evaluated at the age-specific category level (2). Previous research in general and athletic populations highlighted the non-linear increases in physical and performance capacities throughout adolescence (3), yet anthropometric and physical performance measurements may be prone to differences in biological maturation (4, 5). The general notion of biological maturation refers to the process of progressive changes that lead from an undifferentiated or immature state to a highly organised, specialised, mature or adult state (6).

In the sports and exercise science, there is a general appreciation regarding the importance of tracking measures of biological maturation (5), given the relative contribution of, for example, sexual maturation in explaining 21% to 50% of the variance in 30-m dash, vertical jump, and Yo-Yo Intermittent Endurance Test Level 1 performance in youth soccer players (4) . In practice, the assessment of biological maturation generally involves the examination of discrete indicators during the course of development such as skeletal age and sexual characteristics (6). The determination of skeletal age represents a criterion method to assess biological maturation and has received particular attention in youth soccer research (5). The assessment process involves visual or automated rating of left-hand and wrist roentgenograms, with the assignment of skeletal ages determined by the developmental stages for each epiphyseal centre of interest (6). Studies in youth soccer have gathered measurements based on different protocols and criteria, with the most commonly used yet distinct protocols including Greulich-Pyle, Tanner–

Whitehouse, and Fels methods (4, 7-18). The Greulich-Pyle is an example of an atlas technique assigning the skeletal age to the roentgenogram as the chronological age consistent with the pictorial standard from the reference population (6). The general Tanner–Whitehouse (TW) method, and subsequent iterations (19, 20), determines the skeletal age of the subject based on a cumulative score derived from a series of indicators relating to the appearance of each specific bone of the hand and wrist (6). The principal revisions of this method are based on the assessment of the radius-ulna-short (RUS) bones, with full maturation (RUS score = 1000 au) corresponding to a skeletal age of 18.2 yr in TW-II and 16.5 yr in TW-III (6). The Fels hand–wrist method is a more recent iteration, similar to the Tanner–Whitehouse method, combining estimates of the age of appearance from 98 indicators with the addition of metric ratios of lengths of radius, ulna, metacarpals and phalanges also informing the overall skeletal maturity scale (6).

Measures of skeletal age are used to inform classification of players based on relative maturity status (5, 21, 22). Specifically, researchers derive measures of relative maturity, calculated as the difference (Δ) between skeletal age *minus* chronological age, as an indicator relevant to inform grouping and treatment pathways (5, 21-23). In sports performance research, irrespective of the selected protocol, the relative maturity indicator (Δ) is generally used to classify player as *late* (*delayed*) if skeletal age *minus* chronological age difference $\Delta < -1$ yr; *average* (*on time*) if skeletal age *minus* chronological age difference lies within $\Delta \pm 1$ yr; *early* (*advanced*) if skeletal age *minus* chronological age difference $\Delta > +1$ yr; *mature* if skeletal age meets full maturation criteria (5). The ±1 year band criterion is generally deemed to approximate typical standard deviations for skeletal age within children of a similar age (5). In sport, the definition of these relative maturity bands was illustrated, for the first time, from the re-examination of Todd atlas-based skeletal ages in a small-scale sample of 55 baseball players (chronological age range: 11 to 13 yr) competing in the 1957 World Series (24). While the extrapolation and application of these relative maturity bands is grounded on anecdotal experience, Krogman concluded that advanced relative skeletal maturity ($\Delta > +1$ yr) impacted decisions for selection of young players in baseball. (24). Researchers in sports and exercise sciences deem the ± 1 yr band consistent with typical variability (SD) for skeletal age within age-specific categories (5), yet the conceptual definition of the resulting classifications remains arbitrary and prone to bias for a number of reasons. Firstly, Krogman extrapolated maturity bands and generalized them to a sample of youth American baseball players with skeletal age determined using Todd standards now deemed obsolete for modern populations (6). Secondly, converting continuous measurements into categorical variables by grouping measures in two or more categories is a common practice in medical and sports research (25). The adoption of this approach, however, causes loss of statistical power and introduces residual confounding with players prone to misclassifications (25). From a practical standpoint, in the context of youth soccer studies, categorising youth athletes can result in a loss of discriminatory value within a given clinical or performance-related measure selected as a benchmark. Accordingly, it seems more reasonable that formal examination of differences in outcomes of interest should involve regression modelling strategies integrating relative maturity and response variables treated as continuous measurements (25).

The measurement purpose dictates methods and procedures for skeletal age assessment (26), and standards may require adjustments or formal validation when applied to non-reference samples (27). Accordingly, a principled justification of the protocol for skeletal age determination appears fundamental and relevant to informing maturity status classifications in a given population of interest (26). Malina et al., (14) showed relative maturity status classifications were inconsistent between TW-III and Fels methods in a sample of 40 elite youth soccer players. Notably, the TW-III protocol misclassified subjects aged 15 years or more as mature compared to Fels ratings (14), likely reflecting the fundamental differences in skeletal age ranges between TW-III and Fels measurement scales (15). The failure of published studies in this field to justify the protocol selection for skeletal age assessment renders findings potentially ungeneralizable, suggesting the application of the current criteria for relative maturity status classifications may be unreliable. For example, Carling and colleagues (7) and Gouvea and colleagues (10) explored differences in measures of anthropometry and performance with relative maturity status of youth players determined as per the Greulich-Pyle atlas. Investigations by Coelho-e-Silva and colleagues (8), Figueireido and colleagues (9), Texeira and colleagues (16), and Valente-dos-Santos and colleagues (18) used the Fels method, whereas, more recently, Itoh and Hirose (12) adopted the TW-III method for skeletal age determination. Limited guidance on protocol selection for skeletal age assessment reflects the lack of investigations on the properties of each assessment method. From a clinical standpoint, the methodological assessment of an adult height prediction method for bias and random error represents a formal validation of the reference skeletal age protocol application to a population of interest (28). In line with these observations, a recent method comparison study indicated that TW-II can be considered the protocol of choice for adult height prediction purposes in youth soccer players from the Middle East (13). Despite its application for assessing skeletal maturity and adult height prediction in Middle Eastern players, the role of skeletal maturation as a potential mediator of differences in test performance outcomes remains unexplored in youth Arab athletes. With this in mind, the methodological inconsistencies of previous investigations and recent evidence from Middle Eastern soccer players (13) informed considerations on study design and procedures involving the integration of TW-II skeletal age and test performance assessments.

To address the current evidence base in this field, we therefore assessed the appropriateness of current maturity status classification criteria by examining the influence of differences in relative skeletal maturity on performance test outcome measures in elite youth soccer players from the Middle East.

2. Methods

2.1 Participants

The study sample included skeletal age and performance assessments data available for a sample of n = 199 male, outfield soccer players enrolled as academy student-athletes (chronological age range: 11.7 to 17.8 yr; standing height range: 135 to 190.3 cm, body mass range: 28.9 to 78.7 kg) over nine competitive seasons from the historical population (N = 876). The data collection was part of the annual medical screening and a longitudinal growth and maturation project (protocol number: E202008009) involving also regular performance/fitness screenings. Signed parental consent was obtained before each academy season to use data for research purposes. This retrospective study was approved by the Aspire Zone Foundation Institutional Review Board, Doha, State of Qatar.

2.2 Design and procedures

The present investigation adopted a retrospective, mixed-longitudinal study design (29) involving student-athletes measured once and others more than once (annual screening range: 1 to 5 visits). A mixed-longitudinal design represents a plausible option for studies on growth and development to isolate the contributions of age, cohort and time-of-measurement effects to developmental data, thereby limiting the confounding of cohort-related differences typical of cross-sectional designs (29). Hand x-rays, standing height, body mass and performance test outcome measurements collected in student-athletes as part of the annual screening were retrieved from the Academy medical records, anonymised, analysed and used to determine skeletal age at the time of the scan. Standing height was measured using a wall-mounted stadiometer to the nearest 0.1 cm according to the stretch stature protocol (Holtain Limited, Crosswell, UK), and body weight measurements were obtained using digital scales.

Physical performance assessments took place on distinct occasions and, approximately, every three months during the course of a competitive seasons. Players performed 2 maximal 40-m sprints during which 10-m split times were recorded using electronic timing gates and measured to the nearest 0.01s (Swift Performance Equipment, Lismore, Australia). Players commenced each sprint when ready from a standing start with their front foot half a meter behind the first timing gate and were instructed to sprint as fast as possible over the full 40-m distance. Trials were separated by at least 60s of recovery with the best performances used as the final result.

Countermovement jump (CMJ,) height was derived using a force plate (Kistler 9286AA, Kistler Instrument Corp., Winterthur, Switzerland). Players were instructed to keep their hands on their hips with the depth of the counter movement self-selected. Each trial was validated by visual inspection to ensure each landing was without significant leg flexion. At least three valid CMJ's were performed separated by 25-s of passive recovery, with the best performance recorded.

A continuous incremental field running test was used to determined maximal aerobic speed (MAS), with the assessment beginning at an initial running speed of 8.5 km·h^{-1} followed speed increases of 0.5 km·h^{-1} each minute until volitional exhaustion. A player's MAS (km·h^{-1}) was recorded as the average velocity of the last stage completed. The MAS was calculated according to the equation: MAS = S + (t/60 × 0.5), where S is the last completed speed in km/h and t is the time in seconds, if the stage was not completed. Using recent test-retest data from a sub-sample of $n = 62$ elite youth soccer players (chronological age range: 12.2 to 18.3 yr) from the available population ($N = 876$), the estimated minimal detectable change values for 10-m sprinting, 40-m sprinting, CMJ height, and MAS were ± 0.12 s (95% confidence interval, 0.10 to 0.14 s), ± 0.28 s (95% confidence interval, 0.24 to 0.34 s), ± 4.4 cm (95% confidence interval, 3.7 to 5.4 cm), and ± 1.4 $\text{km}\cdot\text{h}^{-1}$ (95% confidence interval, 1.2 to 1.7 $\text{km}\cdot\text{h}^{-1}$), respectively.

Assessment of skeletal age involved standard radiographs (Digital Diagnost, Philips, USA) of the radius, ulna, carpals, metacarpals and phalanges (5). Modern technology now allows to minimising the exposure to radiation to as little as 0.0001 millisievert (mSv), which is commensurate to less than natural background radiation walking around a city centre, or any radiation associated with a 2-hr flight (5). Roentgenograms were evaluated as per the manual Tanner-Whitehouse RUS protocol by the same rater (AJ) with twenty years of experience. Test-retest assessment of the manual rating method suggested reasonable intra-rater reliability for this protocol, with ratings being practically equivalent to automated imaging assessments (13). Data relevant to tracking skeletal maturation and growth in this population informed the conversion of summary RUS scores to TW-II skeletal ages (range: 10 to 18.2 yr) (13).

2.3 Statistical analysis

Separate random-effects generalized additive models with restricted maximum likelihood (30) estimated effects for performance test outcomes by skeletal age and relative skeletal maturity (Δ) at the time of the hand-wrist x-ray scan as the explanatory variable, respectively. Models included the performance test outcome measure as the response variable, with the smooth term for the explanatory variable set at 3,5,7, and 9 basis functions plus a subject-specific random effect penalized by a ridge penalty (30). Optimal smooth model selection was determined via information theory (30). Post-estimation model diagnostics was conducted based on visual inspection of each model residuals (31). Effects were reported as estimated marginal means (32) presented with 95% confidence interval (CI) describing the likely range of values compatible with the true population parameter. A 95% prediction interval (PI) was estimated to quantify the range of values within which 95% of future similar observations may lie for descriptive analyses only (33, 34). Existing literature in this field informed comparisons for analyses with relative skeletal maturity as the explanatory variable (5). In the absence of an established anchor defining a practically relevant increase or reduction for each of our physical test performance outcome measures, we considered the estimated minimal detectable change values to inform interpretations in the present study (35). Specifically, in the present study, the notion of practical relevance refers to whether the size of a change or difference between two testing occasions or comparisons of interest is distinguishable from the random within-subject variability of the measurement (35). Estimates for each relative skeletal maturity comparison were declared practically relevant based on the location of the 95%CI for the mean effects and interpreted against pre-defined minimal detectable change values for each performance outcome measure. Random-effects variance decomposition was conducted to explore the proportion of variance explained by skeletal age and relative skeletal maturity in each model (30). Statistical analyses were conducted using R (version 3.6.3, R Foundation for Statistical Computing).

Table 1 about here

Figure 1 about here

3. Results

Descriptive data for maturity and performance outcome measures were illustrated in Figure 1 and Table 1. Random-effects variance decomposition suggested TW-II skeletal age accounted for 21.2%, 16.4%, 10.5%, and 10.2% of the between-subject variability in 10-m sprinting, 40-m sprinting, CMJ, and VAM performance, respectively. Difference in test performance outcomes by relative maturity were presented in Figure 2. For 10-m sprinting, effects were, in general, not practically relevant (Figure 2). The mean difference in test performance for Δ = -1.5 versus Δ = +1 in relative maturity status was +0.16 s (95%CI, +0.11 to +0.21 s). For 40-m sprinting, practically relevant effects of +0.51 s (95%CI, +0.35 to +0.67 s) and +0.62 s (95%CI, +0.45 to +0.78 s) were associated with a relative maturity status of Δ = -1.5 yr versus Δ = +0.5 and Δ = +1 yr, respectively (Figure 2). Practically relevant differences of +0.39 s (95%CI, +0.27 to +0.54 s) for Δ = -1 yr versus Δ = 0.5 yr, +0.39 s (95%CI, +0.24 to +0.54 s) for Δ = -1.5 yr versus Δ = 0 yr, and 0.51 s (95%CI, 0.28 to 0.57 s) for Δ = -2 yr versus Δ = 0 yr relative maturity status comparisons, respectively (Figure 2). For CMJ, a practically relevant effect of -8 cm (95%CI, -10 to -5 cm) was observed for Δ = -1.5 yr versus Δ = +1 yr relative maturity status comparison (Figure 2).Practically relevant differences of -7 cm (95%CI, -11 to -4 cm) for Δ = -2 yr versus Δ = 0 yr, -7 cm (95%CI, -9 to -4 cm) for Δ = -1.5 yr versus Δ = +0.5 yr , and -5 cm (95%CI, -8 to -3 cm) for Δ = -1.5 yr versus Δ = 0 yr relative maturity status comparisons, were observed respectively (Figure 2). Irrespective of differences in relative maturity status, effects for MAS were not practically relevant. Analysis of the random-effects variance components indicated relative skeletal maturity accounted for 8.6%, 8.4%, 5.8%, and 1.1% of the between-subject variability in 10-m sprinting, 40-m sprinting, CMJ, and VAM performance, respectively.

Figure 2 about here

4. Discussion

In sports, assessing skeletal maturity can be useful for grouping athletes and gathering preliminary information of the remaining growth potential to guide athlete development processes. With the objective to address the contradictory evidence base in this field and informing our study framework based on evidence from this population (13), we investigated, for the first time, the influence of differences in relative skeletal maturity on performance test outcomes in elite youth Middle Eastern soccer players. When integrating skeletal age and performance assessments, our main findings suggested conventional criteria used to define early, on-time, and advanced maturity categories in youth soccer studies lacked empirical support for grouping in the present study population. Between-player differences in test performance may depend on a substantial delay in relative skeletal maturity ($\Delta \leq -1.5$ yr) and the physical performance outcome being assessed.

A number of practical factors pose challenges in gathering longitudinal, paired measurements of skeletal age and test performance in sports academy settings which likely explains a general lack of investigations in this field. Furthermore, test performance comparisons between relative maturity status groups are also limited to studies involving samples from Western countries and using different skeletal age protocols (7-11, 16-18). While our study lends indirect support to general considerations in the youth soccer literature, evidence in this field remains contradictory. In particular, researchers in this field (7-11, 16-18) treated the continuous relative skeletal maturity variable as categorical for a priori classifications, a practice which is discouraged on statistical grounds (25). Notably, categorization rests on the implausible assumption of regression discontinuity as interval boundaries are crossed (25). This also might have contributed to yielding results unnecessarily prone to sampling imprecision given the low number of subjects in outer categories for some previous studies (7-11, 16-18). Our explorations indicated that differences in 40-m sprinting and CMJ performance were consistent only for Δ = -1.5 yr versus Δ = +1 yr in relative skeletal maturity comparisons, with unclear effects for the 10-m sprinting and MAS variables (Fig. 2A,D). Specifically, the mean effect for relative skeletal maturity of Δ = -1.5 yr versus Δ = +1 yr in 10-m sprinting was +0.16 s (95%CI, +0.11 to +0.21 s). The degree of the difference we observed would not exclude the presence of a potential effect (36), yet not exceeding clearly our pre-defined target difference value deemed of practical relevance for this variable (Δ = ± 0.12 s). In this context, Carling and colleagues assessed skeletal age using the Greulich-Pyle method in French youth soccer players (n=158) and concluded early maturing players (Δ > +1 yr) performed better, with similar findings for CMJ height (7). Using the Fels method to assess 159 youth players from five clubs in the midlands of Portugal, Figueiredo and colleagues showed early (Δ > +1 yr) and on-time (Δ = ±1 yr) maturing players differed in CMJ height compared to late (Δ < -1 yr) maturing players (9). Subsequent explorations from this same sample revealed youth players that moved to an elite playing standard performed better in physical tests and were skeletally older than regional counterparts and dropouts (37). Gouvea and colleagues (37) determined classifications based on the Greulich-Pyle method, with inconsistent effects of relative skeletal maturity on anthropometric indicators, functional capabilities and technical skills in a sample of youth soccer players from Brazil (n=60). More recently, using TW-III to assess skeletal maturity, Itoh and Hirose (12) concluded that late (Δ < -1 yr) maturing players had worse test performances than on-time (Δ = ±1 yr) and early (Δ > +1 yr) maturing players from Asia (n=49). Likewise, the general lack of clear effects we observed for the MAS variable is another aspect of our findings deserving consideration (Fig. 2A). When comparing maturity status categories, Carling and colleagues (7) and Texeira and colleagues (16) found trivial differences in maximal and peak oxygen uptake ($mL \cdot kg^{-1} \cdot min^{-1}$), respectively. In contrast, Gouvea et al.,

(10) showed higher intermittent endurance values for on-time (Δ = ±1 yr) and late (Δ < -1 yr) *versus* early (Δ > +1 yr) maturing players, whereas Figueiredo and colleagues (9) found late (Δ < -1 yr) maturing boys had greater endurance capacity than on time (Δ = ±1 yr) and early (Δ > +1 yr) maturing boys. The direction and the degree of the effects we observed for the performance outcomes we investigated might reflect the nature of these measures and their underlying sensitivity to the influence of differences in skeletal maturity. In practice, our study indicated that grouping of players based on conventional relative maturity categories, with a particular reference to average (on time; Δ = ± 1 yr) and early (advanced; Δ > +1 yr) classifications, lacks empirical support (Figure 2).

Our investigation advances current knowledge on the influence of differences in relative skeletal maturity on performance test outcomes in elite youth soccer. Nevertheless, the heterogeneity and inconsistency of methods and procedures from previous investigations precluded formal comparisons with our findings. First, researchers adopted skeletal age assessment protocols without formal justification and knowledge of their applicability to the respective study sample (7-11, 16-18). In line with recommendations in the field (5, 28), we therefore informed our study design based on outcomes from comparisons of different protocols in our study population which indicated TW-II as the method of choice for assessing skeletal maturity and tracking growth (13). The arbitrary selection of skeletal age protocols may have contributed to introducing biases in the effects of relative skeletal maturity on test performance outcomes reported in previous studies. In practice, using different protocols to appraise the same construct would suggest that relative skeletal maturity status, generally calculated as difference between skeletal age *minus* chronological age, may lack a conceptual basis for player grouping beyond a specific study context. For example, Figueiredo and colleagues assessed the skeletal age of Portuguese players using Fels (9), whereas Itoh and

Hirose used TW-III (12). Notably, Malina and colleagues showed a substantial degree of misclassification in Fels versus TW-III relative maturity status classifications (14). The mean relative skeletal maturity status for the Fels method was greater than the mean difference with the TW-III method for players in 12 to 14 yr of age range, with more 15-year-old boys classified as skeletally mature as per TW-III versus Fels criteria (14). Accordingly, any consideration on potential over- or under-representation of late maturing players may lack clinical and practical context in the absence of established consensus on protocol selection. The fact that previous sports performance studies found youth athletes as relatively advanced in their relative skeletal maturity (Δ > +1 yr) (7-11, 16-18) is, however, clinically normal and plausible (38) in a well-nourished setting with limited constraint on development. In contrast, an exaggerated advancement (Δ > +2 yr) or delay (Δ < - 2 yr) in relative skeletal maturity may only occur as a result of an underlying endocrine pathology (21). Our study findings (Figure 2) seemed aligned with these clinical considerations regarding how a substantial delay in relative skeletal maturity may influence test performance (Figure 2). However, considerations of any potential nexus are contingent on more appropriate clinical designs to conduct formal explorations in sports populations.

We also highlight that, in previous studies, sports performance researchers interpreted differences in test performance between relative skeletal maturity categories based on their statistical significance rather than practical relevance (35). In lay terms, published studies investigating the influence of differences in skeletal maturity on performance test outcome measures failed to provide consistent guidelines to inform strategies for optimal youth player development and performance enhancement (7-11, 16-18). A clear definition of target effects deemed of practical relevance is paramount for rationalized interpretations of changes and differences in performance test outcomes within athletic development programmes (35).

Different methods are available for researchers to establish practically relevant effects, with decisions on criterion selection depending on the context and purpose of the measurement (35). We adopted a pragmatic approach with values of interest established on error-based statistic whose magnitude was similar to previous reports in our study population (39). Any conclusive inference on test performance differences by relative skeletal maturity would have been unwarranted if previous studies followed similar conceptual procedures. Carling and colleagues (7) concluded 40-m sprinting and CMJ height differed between late, on-time, and early maturity categories, but meaningful effects would have only been observed between late ($\Delta < -1$ yr) and early ($\Delta > +1$ yr) maturing players if interpreted as per our study methods. Likewise, the CMJ height differences reported by Figueiredo and colleagues (9) would be potentially trivial if more rationalised methods supported the interpretations of the estimated effects (35).

Our line of evidence highlighted the relevance of tracking skeletal maturation limited to younger age categories (U13 to U15) given the potential variability in maturation stages (Figure 1). To illustrate this further from a practical standpoint, we shall consider the cases of two student-athletes training and competing in the same chronological age category. Estimated age peak at velocity of 13.62 yr (95% CI, 13.55 to 13.70 yr) and peak height velocity of 9.9 $\text{cm}\cdot\text{yr}^{-1}$ (95% CI, 9.5 to 10.3 $\text{cm}\cdot\text{yr}^{-1}$) for this population (13) are also considered as complementary information to relative skeletal maturity. Demographic, anthropometric, and skeletal age characteristics were obtained for a 12.2-year old player with a measured standing height at the time of the x-ray scan of 169.7 cm, annual height velocity of 4.8 $\text{cm}\cdot\text{yr}^{-1}$, a RUS score of 813 au, and TW-II skeletal age of 16.4 years. Test scores for 10-m sprinting, 40-m sprinting, CMJ, and VAM performance were 1.9 s, 5.8 s, 32 cm, and 14.6 $\text{km}\cdot\text{h}^{-1}$, respectively. In the other case, we consider a 12.7-year old player with standing height of 148.9 cm, annual height velocity of 4.8 cm·yr^{-1}, a RUS score of 332 au, and TW-II skeletal age of 11 years. Performance in 10-m sprinting, 40-m sprinting, CMJ, and VAM assessments was 2.0 s, 6.4 s, 29.5 cm, and 14.5 km·h^{-1}, respectively. Notably, sprinting and lower-limb explosive strength attributes would appear different between the two cases on the basis of our pre-defined criteria for test performance interpretations. When contextualised, these differences in performance are consistent with differences in relative skeletal maturity ($\Delta = +4.2$ yr *versus* $\Delta = -1.7$ yr), together with the fact the two subjects are passing through contrasting phases of the growth process. From a real-world perspective, such information can serve as valuable tools for coaches and practitioners to arrive at more context-specific decisions for talent identification and development purposes. While also relevant to accurate estimations of predicted adult height (13), our findings substantiated further the importance of tracking proxy measures of biological maturation to inform context-specific player development strategies, particularly for U13 to U15 age categories (Figure 1).

Our study addressed the current evidence base extending knowledge about the extent of relative maturity status evaluation and its application for grouping in soccer. From an applied perspective, our findings and the current literature suggested the need for expert consensus on the construct definition of relative maturity status. The re-appraisal of Todd atlas-based skeletal ages from youth baseball players guided the definition of conventional maturity status classifications criteria in this domain (24). Yet, these and other criteria were discussed by researchers in other fields (21, 22, 40, 41). Bayley provided the first example of early, on-time, and late maturity grouping in boys and girls with skeletal ages determined as per Todd standards in 1943 (40). Boys were classified into three groups *based on the age at which they attained a skeletal age of 17 years and 3 months* (40). Classifications were determined using a retrodictive approach in which the means of the chronological ages for the three maturity groups were expected approximately one year apart at maturity (40). Pyle and colleagues defined maturity status based on the progression of skeletal age-by-chronological age longitudinal curves for a sample of 133 children (chronological age range: 1 to 18 yr) interpreted against sample-specific norms (41). According to this procedure, all the available skeletal ages for a given subject must remain above or below a zone limited by the spread of ±1 SD to be advanced or delayed in maturity, respectively (41). Similar criteria were applied to describe the rate of development (41). Using data from South African children from the urban conurbation of Johannesburg–Soweto, Hawley and colleagues (21) explored predictors of relative maturity, calculated as TW-III skeletal age *minus* chronological age, using criteria similar to those adopted illustrated by Krogman and other researchers in this field (5, 24). In a clinical study exploring the association between insulin-like growth factor-1 and skeletal maturation before and after growth hormone treatment, Zhao and colleagues (42) defined *late (delayed)* maturity for a given subject if the Greulich-Pyle skeletal age *minus* chronological age (Δ) value fell below 2-SD based on data of 783 short children and adolescents from China. In other medical disciplines as orthodontics, calculation of relative maturity generally informs treatment planning and dentofacial orthopaedics (22). Using the Greulich-Pyle method, Suri and colleagues divided 572 serial hand-wrist radiographs of 68 white children (chronological age range: 9 to 18 yr) with normal facial growth into five categories spaced by a pre-defined margin of error of Δ = ± 0.5 yr (22). Adding complexity to the set of operational classifications based on skeletal age (21, 22, 40, 41), researchers also defined early and late maturation on a different conceptual basis using alternative instruments such as, for example, a classical growth chart (43, 44). In this context, Tanner and Davies (43) defined *late and early maturers* children whose standing height lay below or above the 5th and 95th centiles on a height-on-chronological age growth standard. Likewise, more recently, Cameron (44) defined children whose height centile status moved from the 50th to below the 10th and above the 90th centiles as *late and early* *developers*, respectively. Overall, the lack of precise guidelines on skeletal age assessment protocol, the inconsistency on classification criteria and definitions likely have contributed to potential misclassifications of subjects and heterogeneity of findings in this and other research settings. With this in mind, evidence from previous research in this population informed the adoption of a principled approach in our study keeping the relative skeletal maturity variable as a continuous measurement to avoid the shortcomings of categorisation and any a priori approach influencing our results (25).

We conducted the largest study exploring the influence of differences in relative skeletal maturity on performance test outcomes in the field of sports and exercise sciences (n=199), yet not without limitations. Our investigation examined the influence of differences in relative maturity status using data for a limited number of performance test outcomes. While reporting in this field is diverse, we selected outcomes based on reliability and academy strategy-based criteria to maximize the practical context of our findings. Researchers in sports science and medicine also discussed the potential utility of other criteria for grouping athletic populations via, for example, the bio-banding strategy established on the percentage of predicted adult height index (45). Despite the potential utility of this approach, recent explorations revealed how relative skeletal maturity constitutes the overarching criterion given the heterogeneous distributions of youth players within bands at relatively lower percentages of predicted adult height (45). Likewise, we used ratings limited to left-hand and wrist roentgenograms for assessing skeletal maturity. Biologists and anthropologists discussed the value of other assessment methods and different anatomical sites to determine skeletal age (46). Rating of hand x-rays remain, nonetheless, more practically feasible. Accordingly, the notion of maturation as a measure of progressive development towards adulthood deserves careful consideration as it may be a cause of more misunderstanding than clarity (47). Any advancement or delay in maturation is generally extrapolated as the difference between skeletal age and chronological age. Calculation of this indicator has the sole advantage of negating the need to control for chronological age in any model for describing the degree to which a youth athlete is advanced or delayed in their skeletal maturity (21). Any difference that may be positive or negative in sign merely reflects the progression of skeletal development relative to chronological age, precluding any conclusion regarding potential factors that may underlie any advancement or delay in biological maturation (48). Also, the fact that one year of skeletal age is not biologically equivalent to one year of chronological age deserves consideration for the calculation and generalisation of relative skeletal maturity (49). Collectively, the nature of this measurement suggests caution with the use and application of terms such as "*early maturer*" or "*late maturer*" (47). The scrutiny of a selected indicator or anatomical site in isolation is unlikely to provide an unbiased reflection of the overall developmental process (47). Marshall stated that the term "*early maturer*" applies only to someone who reaches full maturity at an early (chronological) age and depends on the maturity indicator someone assesses (47). Any change in the neuroendrocrine system leading to development of secondary sexual characteristics may not occur simultaneously with mechanisms regulating maturation and closure of different centres of ossification (47). The general assessment of skeletal age may also remain constrained in the applied settings of a sporting academy as a non-medical human imaging requiring formal justification for benefit by authorities for sports organizations, players, medical professionals, and regulatory bodies (50).

5. Conclusion

Outcomes from the integration of skeletal age and performance assessments suggested conventional maturity status classification criteria lacked empirical support for applications relevant to player grouping and development in our study context. Differences in test performance among youth players were inconsistent across different test protocols, whose extent may depend on a substantial delay in skeletal maturation ($\Delta \leq -1.5$ years) and the test performance measurement. Our study advanced knowledge on the role of skeletal maturity determination applied for tracking test performance to an underexplored population of youth athletes.

Acknowledgments

The authors wish to thank the staff of the Football Performance & Science Department at Aspire Academy for the invaluable support in the collection of the routine assessments data.

Conflict of Interest

The authors declare no conflicts of interest. This study is not funded. The results of this study do not constitute endorsement by ACSM. The results of this study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

References

1. Svensson M, Drust B. Testing soccer players. *J Sports Sci*. 2005;23(6):601-18.
2. Hulse MA, Morris JG, Hawkins RD, Hodson A, Nevill AM, Nevill ME. A field-test battery for elite, young soccer players. *Int J Sports Med.* 2013;34(4):302-11.
3. Beunen G, Malina RM. Growth and physical performance relative to the timing of the adolescent spurt. *Exerc Sport Sci Rev.* 1988;16:503-40.
4. Malina RM, Eisenmann JC, Cumming SP, Ribeiro B, Aroso J. Maturity-associated variation in the growth and functional capacities of youth football (soccer) players 13-15 years. *Eur J Appl Physiol.* 2004;91(5-6):555-62.

5. Malina RM. Skeletal age and age verification in youth sport. *Sports Med.* 2011;41(11):925-47.

6. Cameron N. Measuring maturity. In: L Molinari, N Cameron, RC Hauspie editors. *Methods in Human Growth Research*. Cambridge: Cambridge University Press; 2004, pp. 108-40.

7. Carling C, Le Gall F, Malina RM. Body size, skeletal maturity, and functional characteristics of elite academy soccer players on entry between 1992 and 2003. *J Sports Sci*. 2012;30(15):1683-93.

8. Coelho ESMJ, Figueiredo AJ, Simões F et al. Discrimination of u-14 soccer players by level and position. *Int J Sports Med.* 2010;31(11):790-6.

9. Figueiredo AJ, Gonçalves CE, Coelho ESMJ, Malina RM. Youth soccer players, 11-14 years: maturity, size, function, skill and goal orientation. *Ann Hum Biol.* 2009;36(1):60-73.

10. Gouvea M, Cyrino ES, Ribeiro AS et al. Influence of skeletal maturity on size, function and sport-specific technical skills in youth soccer players. *Int J Sports Med.* 2016;37(6):464-9.

11. Hirose N. Relationships among birth-month distribution, skeletal age and anthropometric characteristics in adolescent elite soccer players. *J Sports Sci*. 2009;27(11):1159-66.

12. Itoh R, Hirose N. Relationship among biological maturation, physical characteristics, and motor abilities in youth elite soccer players. *J Strength Cond Res.* 2020;34(2):382-8.

13. Lolli L, Johnson A, Monaco M, Cardinale M, Di Salvo V, Gregson W. Tanner–Whitehouse and Modified Bayley–Pinneau adult height predictions in elite youth soccer players from the Middle East. *Med Sci Sports Exerc.* 2021;53(12):2683-2690.

14. Malina RM, Chamorro M, Serratosa L, Morate F. TW3 and Fels skeletal ages in elite youth soccer players. *Ann Hum Biol.* 2007;34(2):265-72.

15. Malina RM, Coelho ESMJ, Figueiredo AJ et al. Tanner-Whitehouse skeletal ages in male youth soccer players: TW2 or TW3? *Sports Med.* 2018;48(4):991-1008.

16. Teixeira AS, Guglielmo LGA, Fernandes-da-Silva J et al. Skeletal maturity and oxygen uptake in youth soccer controlling for concurrent size descriptors. *PLoS One*. 2018;13(10):e0205976.

17. Teixeira AS, Valente-dos-Santos J, Coelho ESMJ et al. Skeletal maturation and aerobic performance in young soccer players from professional academies. *Int J Sports Med.* 2015;36(13):1069-75.

18. Valente-dos-Santos J, Coelho-e-Silva MJ, Martins RA et al. Modelling developmental changes in repeated-sprint ability by chronological and skeletal ages in young soccer players. *Int J Sports Med.* 2012;33(10):773-80.

19. Tanner JM, Healy M, Goldstein H, Cameron N. *Assessment of skeletal maturity and prediction of adult height (TW3 Method)*. 3rd ed. London: Academic Press; 2001.

20. Tanner JM, Whitehouse RH, Cameron N, Marshall WA, Healy M, Goldstein H. *Assessment of skeletal maturity and prediction of adult height (TW2 method)* 2nd ed. London: Academic Press; 1983.

21. Hawley NL, Rousham EK, Johnson W, Norris SA, Pettifor JM, Cameron N. Determinants of relative skeletal maturity in South African children. *Bone*. 2012;50(1):259-64.

22. Suri S, Prasad C, Tompson B, Lou W. Longitudinal comparison of skeletal age determined by the Greulich and Pyle method and chronologic age in normally growing children, and clinical interpretations for orthodontics. *Am J Orthod Dentofacial Orthop.* 2013;143(1):50-60.

23. McCormack SE, Chesi A, Mitchell JA et al. Relative skeletal maturation and population ancestry in nonobese children and adolescents. *J Bone Miner Res.* 2017;32(1):115-24.

24. Krogman WM. Maturation age of 55 boys in the little league world series, 1957. *Res Q*. 1959;30(1):54-6.

25. Harrell FE, Jr. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis.* New York: Springer; 2015.

26. Malina RM. A consideration of factors underlying the selection of methods in the assessment of skeletal maturity. *Am J Phys Anthropol.* 1971;35(3):341-6.

27. Roche AF, Davila GH, Eyman SL. A comparison between Greulich-Pyle and Tanner-Whitehouse assessments of skeletal maturity. *Radiology*. 1971;98(2):273-80.

28. Martin DD, Schittenhelm J, Thodberg HH. Validation of adult height prediction based on automated bone age determination in the Paris Longitudinal Study of healthy children. *Pediatr Radiol.* 2016;46(2):263-9.

29. Chinn S. Mixed longitudinal studies: their efficiency for the estimation of trends over time. *Ann Hum Biol.* 1988;15(6):443-54.

30. Wood SN. *Generalized additive models: an introduction with R*. 2nd ed. Boca Raton: CRC Press; 2017.

31. Fasiolo M, Nedellec R, Goude Y, Wood SN. Scalable visualization methods for modern generalized additive models. *J Comput Graph Stat.* 2020;29(1):78-86.

32. Lenth RV, Buerkner P, Herve M, Love J, Riebl H, Singmann H. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.4. https://cran.r-project.org/web/packages/emmeans/index.html. In: 2021.

33. Francq BG, Lin D, Hoyer W. Confidence, prediction, and tolerance in linear mixed models. *Stat Med*. 2019;38(30):5603-22.

34. Marra G, Wood SN. Coverage properties of confidence intervals for generalized additive model components. *Scand J Stat*. 2012;39(1):53-74.

35. Datson N, Lolli L, Drust B, Atkinson G, Weston M, Gregson W. Inter-methodological quantification of the target change for performance test outcomes relevant to elite female soccer players. *Sci Med Footb*. 2021:1-14.

36. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ*. 1995;311(7003):485.

37. Figueiredo AJ, Gonçalves CE, Coelho ESMJ, Malina RM. Characteristics of youth soccer players who drop out, persist or move up. *J Sports Sci*. 2009;27(9):883-91.

38. Boeyer ME, Sherwood RJ, Deroche CB, Duren DL. Early maturity as the new normal: a century-long study of bone age. *Clin Orthop Relat Res.* 2018;476(11):2112-22.

39. Buchheit M, Mendez-Villanueva A. Reliability and stability of anthropometric and performance measures in highly-trained young soccer players: effect of age and maturation. *J Sports Sci*. 2013;31(12):1332-43.

40. Bayley N. Skeletal maturing in adolescence as a basis for determining percentage of completed growth. *Child Dev*. 1943;14(1):1-46.

41. Pyle SI, Reed RB, Stuart HC. Patterns of skeletal development in the hand. *Pediatrics*. 1959;24:886-903.

42. Zhao Q, Zhang M, Chu Y et al. Association between insulin-like growth factor-1 and relative skeletal maturation: a retrospective cohort study of short children and adolescents. *Biomed Res Int.* 2020;2020:8052143.

43. Tanner JM, Davies PS. Clinical longitudinal standards for height and height velocity for North American children. *J Pediatr.* 1985;107(3):317-29.

44. Cameron N. The biology of growth. *Nestle Nutr Workshop Ser Pediatr Program*. 2008;61:1-19.

45. Malina RM, Cumming SP, Rogol AD et al. Bio-banding in youth sports: background, concept, and application. *Sports Med.* 2019;49(11):1671-85.

46. Pyle SI, Stuart HC, Cornoni J, Reed RB. Onsets, completions, and spans of the osseous stage of development in representative bone growth centers of the extremities. *Monogr Soc Res Child Dev.* 1961;26:1-126.

47. Marshall WA. Interrelationships of skeletal maturation, sexual development and somatic growth in man. *Ann Hum Biol.* 1974;1(1):29-40.

48. Malina RM. Growth and maturity status of young soccer players. In: T Reilly, AM Williams editors. *Science and Soccer*. London: Routledge; 2003, pp. 287-306.

49. Cameron N. Can maturity indicators be used to estimate chronological age in children? *Ann Hum Biol.* 2015;42(4):302-7.

50. *INTERNATIONAL ATOMIC ENERGY AGENCY. Radiation Safety of X Ray Generators and Other Radiation Sources Used for Inspection Purposes and for Non-medical Human Imaging, IAEA Safety Standards Series No. SSG-55, IAEA, Vienna. 2020.* Vienna: International Atomic Energy Agency; 2020.

Figure captions

Figure 1. Density plot showing distribution for absolute (a) and relative (b) measures of skeletal maturity by age category.

Figure 2. Mean effects (Δ) in 10-m sprinting, 40-m sprinting, CMJ, and VAM performance by pairwise comparisons for differences in relative maturity status. The colour intensity of the density strip represents the degree of uncertainty around the point estimate for the mean effect.

Table captions

Table 1. Estimated marginal means for performance test outcomes by Tanner-Whitehouse II skeletal age