©Journal of Sports Science and Medicine (2022) **21,** 640-657
http://www.jssm.org DOI: https://doi.org/10.52082/jssm.2022.639

Research article

Maturity-based correction mechanism for talent identification: When is it needed, does it work, and does it help to better predict who will make it to the pros?

Bryan Charbonnet ✉, Roland Sieghartsleitner, Jürg Schmid, Claudia Zuber, Marc Zibung and Achim Conzelmann
Institute of Sport Science, University of Bern, Switzerland

Abstract

When identifying talent, the confounding influence of maturity status on motor performances is an acknowledged problem. To solve this problem, correction mechanisms have been proposed to transform maturity-biased test scores into maturity-unbiased ones. Whether or not such corrections also improve predictive validity remains unclear. To address this question, we calculated correlations between maturity indicators and motor performance variables among a sample of 121 fifteen-year-old elite youth football players in Switzerland. We corrected motor performance scores identified as maturity-biased, and we assessed correction procedure efficacy. Subsequently, we examined whether corrected scores better predicted levels of performance achievement 6 years after data collection (47 professionals vs. 74 non-professional players) compared with raw scores using point biserial correlations, binary logistic regression models, and DeLong tests. Expectedly, maturity indicators correlated with raw scores ($0.16 \leq | r | \leq 0.72$; $ps < 0.05$), yet not with corrected scores. Contrary to expectations, corrected scores were not associated with an additional predictive benefit (univariate: no significant r-change; multivariate: $0.02 \leq \Delta AUC \leq 0.03$, $ps > 0.05$). We do not interpret raw and corrected score equivalent predictions as a sign of correction mechanism futility (more work for the same output); rather we view them as an invitation to take corrected scores seriously into account (same output, one fewer problem) and to revise correction-related expectations according to initial predictive validity of motor variables, validity of maturity indicators, initial maturity-bias, and selection systems. Recommending maturity-based corrections is legitimate, yet currently based on theoretical rather than empirical (predictive) arguments.

Key words: Soccer, motor skills, physical fitness, growth and development, confounding variable, predictive value of tests.

Introduction

Identifying talent in youth football ultimately comes down to making a binary developmental prediction: Does a specific youth player have what it takes to become a professional as an adult or not? To make this prediction, which aims to support coaches' opinions in the selection process (Sieghartsleitner et al., 2019b; Lath et al., 2020), researchers use a "shopping list of key criteria" (Williams and Reilly, 2000) to build talent identification models based on general linear models (GLM) (for an exceptional example of non-linear approaches, see Siener et al., 2021; Zuber et al., 2016; Pfeiffer and Hohmann, 2012). In essence, they break down talent into multidisciplinary components (Vaeyens et al., 2006), such as physiological (Dodd and Newans, 2018), skill-related (Murr et al., 2018), psychological (Ivarsson et al., 2020), and sociological (Reeves et al., 2018b) criteria, all of which explain some part of career outcome variability (professional vs. non-professional) (Sarmento et al., 2018; Williams et al., 2020). Once related tests are completed, all test scores are usually either converted into *z*-scores (Turner, 2014; Figueiredo et al., 2011; Souza-Lima et al., 2020), weighted in a scoring system (Fuchslocher et al., 2011; Höner et al., 2015), or aggregated within linear predictive models (Sieghartsleitner et al., 2019b; Höner et al., 2021). Such procedures build a total score (for example, total number of points or forecasted probability of becoming a professional in a binary logistic regression model) and position players on a talent continuum, such as ranking lists. Regardless of the calculation procedure, the ground rule of such a summative approach remains unchanged: "the better the *input* (observed test scores), the better the *output* (total score)" (Overton, 2014; Maszczyk et al., 2014).

Distortion problems during puberty

Theoretical (Cumming et al., 2012; Baxter-Jones et al., 2005) and empirical reasons (Meylan et al., 2010; Malina et al., 2015) suggest that the input-output operations of additive-linear scientific models encounter major distortion problems during puberty. From a biological standpoint, individuals belonging to the same age group can differ strongly because some are up to 5 years older/younger than others (Malina et al., 2004). It is only consequential that inter-individual maturity differences affect indicators of athletic performance and potential, favoring either late or early maturing athletes to different degrees according to maturation stage, age, task or sport considered (Baxter-Jones et al., 2005; Mitchell et al., 2017; Hill et al., 2020; Javet et al., 2022). In sports like football, athletes who are closer to fully mature status are repeatedly shown to perform better in motor tests than others who are less developed (Malina et al., 2005; Malina and Cumming, 2004; Albaladejo-Saura et al., 2021; Sieghartsleitner et al., 2019a; with exception of the motor awkwardness hypothesis; Quatman-Yates et al., 2012). As a consequence, failing to consider issues of inter-individual maturity differences means players are erroneously positioned on talent continuums: The more mature, the better the *input* (motor performances), the better the *output* (total score, game evaluation, estimated potential, selection and success chances"; Cripps et al., 2016; Hill et al., 2021; Peña-González et al., 2021; Johnson et al., 2017; with exception of the underdog hypothesis; Smith and Weir, 2020; Cumming et al., 2018b). Therefore, while collecting motor performance data was an important first step, the next logical one is to

Received: 31 May 2022 / Accepted: 30 November 2022 / Published (online): 01 December 2022

move forward with a "game plan" to address these issues (Cumming, 2018; Baxter-Jones et al., 2005).

Potential solutions

On the side of the coaches' eye, some solutions such as bio-banding—grouping players by maturity status (Cumming et al., 2018a; Cumming et al., 2017; Malina et al., 2019; Reeves et al., 2018a)—or maturation-ordered shirt numbers (Lüdin et al., 2022; Mann, 2020) are already in place in different sport organizations around the world to support better rating of in-game performance. On the side of scientific models, linear (Larochelambert et al., 2022) and curvilinear (Abbott et al., 2020) correction mechanisms have been proposed to deal with inter-individual developmental differences (maturity or relative age, for the distinction see Towlson et al., 2021b). They have been implemented in alpine skiing (Larochelambert et al., 2022), athletic sprinting (Romann and Cobley, 2015; Brustio and Boccia, 2021), long jumping (Brustio et al., 2022), or swimming (Hogan et al., 2022). Correction mechanisms estimate how good adolescent athletes would be if they had competed or performed some motor tasks without developmental advances or delays. As such, they provide better informational *inputs* (corrected test scores, in the sense of maturity-unbiased). From a pedagogical viewpoint, corrected scores aspire to equalize the selection chances by promising (biologically) equivalent, and thus fair, inter-individual performance comparisons (Abbott et al., 2021b; Cobley et al., 2020). However, the main goal of talent identification is not (only) fairness, it is (at its core) accurate predictions of a player's future chance of success in adulthood (Sieghartsleitner et al., 2019b; Bergkamp et al., 2019).

The present study

To our knowledge, the accuracy differences of long-term predictions from corrected and uncorrected scores are presently unexamined. Our study's purpose provided insight into this topic. To this end, we investigated whether (1) maturity status influences our motor talent criteria (if not, no correction would be necessary). Then, for the sake of parsimony and implementation simplicity, we corrected the maturity-biased talent criteria with a linear correction mechanism and examined if (2) the correction mechanism worked (if not, the correction would be useless). Finally, we examined if (3) corrected test scores (univariate prediction) or (4) models including several corrected motor scores (multivariate prediction) make better predictions of adult performance levels of youth football players when compared with raw scores/model.

Methods

Participants

As part of the Institute of Sport Science at the University of Bern's *Talent selection and development in Swiss Football* project, we studied motor performances of 15-year-old male elite football players born in 1999 (N = 121; M_{age} = 15.13, SD = 0.34), who played at the highest Swiss national level in their age category. On the one hand, choosing this age group was motivated by biological considerations (it should still be considerably affected by maturity differences; Malina et al., 2015) and on the other hand, by (selection-relevant) cultural realities (from the age group U15 onwards, players enter a new phase in the national football-specific youth development program where performance and exposure to international competition gains in importance; Schweizer Fussballverband, 2014).

The study received approval from the Ethics Committee of the Faculty of Human Sciences of the University of Bern. All players and their parents or guardians provided their written informed consent to participate.

Motor performances

Based on a football-specific literature review (Williams et al., 2020) and subjective talent criteria mentioned by coaches (Jokuschies and Conzelmann, 2016; Bergkamp et al., 2022), we attempted to account for the diversity of motor abilities necessary to excel in football by including a series of eight motor performance tests (see Table 1). Their practical relevance is demonstrated by the fact that they are used by both the Swiss and German Football Association as part of their talent identification program (Höner et al., 2015; Schweizer Fussballverband, 2016). Moreover, these tests have repeatedly been found to validly predict future performance levels (Höner et al., 2021; Leyhr et al., 2018; Höner et al., 2017; Höner and Votteler, 2016; Sieghartsleitner et al., 2019a).

Maturity indicators

We operationalized player maturity status at age 15 with 6 validated pragmatic somatic prediction methods (Mirwald, Moore-1, Moore-2, Fransen-1, Fransen-2) and attained percentage of predicted adult height (%PAH) (see Table 2 for equations and references). Like other research groups (Leyhr et al., 2020a), we call these prediction methods "pragmatic" because they can be used easily and prospectively. Six years later (at the age of 21), all players were recontacted and asked to complete a questionnaire which aimed to collect their real (in the sense of observed as opposed to predicted) adult height. In 37% of the cases, they did not respond, so we searched transfermarkt. Based on the principle of crowd wisdom (Peeters, 2018), online platform transfermarkt is regularly used as a source of information on talent (Leyhr et al., 2020b; Skorski et al., 2016; Doyle and Bottomley, 2018) and sport management research (Herm et al., 2014; Prockl and Frick, 2018; Müller et al., 2017). By knowing players' real adult height, we calculated a seventh maturity indicator retrospectively—the attained percentage of the real adult height (%RAH). Although we are aware that %RAH represents only a sub-dimension of biological maturity (the somatic dimension; Baxter-Jones, 2017), we consider it a more reliable and objective method compared with other indicators since "the one closer to adult height is more mature than the other who is further from adult height" (Malina et al., 2015).

Adult performance level (career outcome)

We collected adult performance level in the 2020–2021 season in the follow-up questionnaire. By knowing adult performance level, we objectively classified players into two groups: professionals (n = 47) who played in the 1st to 3rd league within Switzerland or held professional contracts abroad and non-professionals (n = 74).

Table 1. **Talent criteria in the motor performance domain.**

Test/Variable	Description	Reliability coefficient*	Reference
40-m sprint (sec)	40-m sprint test performed without start signal. Twin photoelectric sensors at the start triggered chronometry and an identical device recorded the time after 40 meters.	0.96	Zuber et al. (2016)
Jumping, CMJ (cm)	An accelerometric system (Myotest, Sitten, Switzerland) measured five attempts in a vertical counter movement jump test (CMJ, without arm swing). We retained the height of the best attempt.	0.96	Casartelli et al. (2010)
Agility (sec)	Players sprinted a short distance, ran around three poles with a change of direction (right, left, right), and repeated these actions mirror-inverted before finishing. The time to complete this run was recorded.	0.83	Höner et al. (2015)
Yo-Yo (m)	Players ran back and forth between two lines 20 meters apart with a 10-second pause between two runs. Acoustic signals determined the pace. Pace increased successively. When players fell behind the pace, the run was stopped. We recorded the distance covered prior to last regularly 20 m lap as the test result.	0.93	Bangsbo et al. (2008)
Dribbling (sec)	Course and procedure of dribbling test was identical to the agility test except that the dribbling test was performed with a ball.	0.56	Höner et al. (2015)
Passing (sec)	Players quickly passed balls in a confined zone and bounced balls off four walls in turn, one in each direction. After the fourth pass, the same sequence was repeated in reverse order (reaching a total of nine passes). Time was measured manually with stopwatches.	0.68	Zuber et al. (2016)
Shooting (points)	Players shot balls eight times into target zones of goals (2 targets, 2 feet, 2 attempts). Successful shots on the target were subjectively rated by speed on a three-point rating scale (1 = low, 2 = medium, and 3 = high speed). The test score was the overall number of points.	0.31	Höner et al. (2015)
Juggling (points)	Players juggled balls along a figure 8-shaped course (alternating left and right foot). Players scored 1 point for each quarter of a circle they completed. We stopped the test after 45 seconds or as soon as a mistake was made, such as one foot twice in succession; the ball touching the ground or any other part of the body. The number of points served as the test score.	0.79	Höner et al. (2015)

40-m sprint = 40-meter sprint; CMJ = counter movement jump; Yo-Yo = Level 1 Yo-Yo intermittent recovery test.
*In all instances, test-retest reliability (r_{tt}) was calculated, except for the CMJ, where the ICC was used.

Table 2. **Maturity indicators and calculation procedures.**

Indicator	Calculation procedure	Reference
Mirwald	Maturity offset = –9.236 + (0.0002708 · leg length and sitting height interaction) + (–0.001663 · age and leg length interaction) + (0.007216 · age and sitting height interaction) + (0.02292 · weight by height ratio)	Mirwald et al. (2002)
Moore-1	Maturity offset = –8.125741 + (0.0070346 · age · sitting height)	Moore et al. (2015)
Moore-2	Maturity offset = –7.999994 + (0.0036124 · age · height)	Moore et al. (2015)
Fransen-1	Maturity ratio = 6.986547255416 + (0.115802846632 · age) + (0.001450825199 · age^2) + (0.004518400406 · weight) – (0.000034086447 · weight^2) – (0.151951447289 · height) + (0.000932836659 · height^2) – (0.000001656585 · height^3) + (0.032198263733 · leg length) – (0.000269025264 · leg length^2) – (0.000760897942 · height · age)	Fransen et al. (2018b)
Fransen-2	Maturity ratio = 6.99 + (0.154 · age – 0.242) + (0.00452 · weight) – (0.0000341 · weight^2) – (0.152 · height) + (0.000933 · height^2) – (0.00000166 · height^3) + (0.0322 · leg length) – (0.000269 · leg length^2) – (0.000761 · height · age)	Fransen et al. (2018a)
%PAH	%PAH = (height at the age of 15 · 100) / predicted adult height according to the method of Sherar et al. (2005)	Sherar et al. (2005)
%RAH	%RAH = (height at the age of 15 · 100) / real adult height at the age of 21	Baxter-Jones et al. (2005)

%PAH = attained percentage of predicted adult height; %RAH = attained percentage of the real adult height. Measurement units: centimeters (leg length, sitting height, height), kilograms (weight), and years (age).

Data analysis

Due to absence, injury, missed, incorrect, or aborted test series, there were missing scores in the study variables (range: 0.8%–6.6%). Little's test (1988) showed these data were not randomly missing ($\chi^2 = 172.91$, $df = 137$, $p = 0.020$). Since there was no clear indication the MAR hypothesis could not be retained, we used EM-imputation based on all study variables to complete the dataset. We analyzed the imputed dataset.

Research question 1: Does maturity status influence motor performance scores?

Pearson's correlations served as the measure of association between raw or uncorrected test scores and each maturity indicator. Accordingly, we identified motor performance tests confounded by maturation and requiring correction.

Correction mechanism

To correct maturity-biased raw scores, we used a four-step correction mechanism. Table 3 illustrates these steps using the example of the sprint scores, which were corrected using %RAH as a maturity indicator. Table 3 also shows how the same sprint score (6.20 sec) of three fictitious players with different maturity statuses (85%, 95%, and 100%, respectively) differed after correction (5.41, 6.17, and 6.63 sec, respectively).

Step 1 (linear regression): We conducted a simple linear regression that predicted raw test score (RS) of specific motor tasks based on maturity status (MS, Table 2).

Step 2 (create new variables): We stored parameters (b_0 and b_1) defining the corresponding regression line. The coefficient b_0 is the constant in the equation. The coefficient b_1 indicates the change in the expected raw score associated with an increase of one unit in the maturity status (MS). By means of these two parameters (b_0 and b_1), we computed two new variables:

1) The test score expected from each player based solely on maturity status: expected raw score (ERS) = $b_0 + b_1 MS_{player}$.

2) The test score expected from the average maturing player in the sample: expected score of the average maturing player (ESA) = $b_0 + b_1 MS_{mean}$.

Step 3 (calculate correction factor): We divided RS of players by biologically expected RS (ERS, see step 2) for an individual correction factor (CF).

Step 4 (generate corrected test score): We multiplied CF by ESA for corrected scores (CS). Since this calculation was based on the ESA, it led to a biological uniformization of the group and provided test scores an exact replica each player with normal development was expected to achieve. By "normal," we mean average as opposed to early or late.

We repeated this procedure for each motor performance test and maturity indicator. In this way, there are corrected test scores for each player in each motor performance test using Mirwald, Moore-1, Moore-2, Fransen-1, Fransen-2, %PAH, or %RAH as maturity indicators.

Research question 2: Is the correction mechanism effective?

To ensure the correction mechanism served its purpose, we calculated Pearson's correlations between each maturity indicator and their corresponding corrected test scores. If $r = 0$, then $R^2 = 0$, thus implying that the proportion of the variance that a maturity indicator and a motor performance test previously shared was successfully partialed out.

Table 3. Four-step correction mechanism and example (three fictitious players with same raw test score but different maturity status).

Procedure				Example			
					Fictitious player 1: Late maturer	**Fictitious player 2: Average maturer**	**Fictitious player 3: Early maturer**
Initialization	**Variable 1**	Indicator of maturity status (MS)	Use MS of all the players	MS = %RAH	MS = 85%	MS = 95%	MS = 100%
	Variable 2	Raw test score (RS)	Use RS of all the players	RS = 40-meter sprint	RS = 6.20 sec	RS = 6.20 sec	RS = 6.20 sec
			Calculate MS_{Mean} (sample mean)	MS_{Mean} = 95.4 %			
Correction mechanism	**Step 1**	Compute a simple linear regression with MS as independent variable and RS as dependent variable and save the regression coefficients	b_0	$b_0 = 13.74$			
			b_1	$b_1 = –0.08$			
	Step 2	Compute the expected raw score (ERS) of each player	$ERS = b_0 + b_1 \cdot MS_{Player}$	$ERS = 13.74 – 0.08 \cdot MS_{Player}$	ERS = 13.74 – 0.08 · 85 = 6.74 sec	ERS = 13.74 – 0.08 · 95 = 5.91 sec	ERS = 13.74 – 0.08 · 100 = 5.50 sec
		Compute the expected score of the average maturing player (ESA)	$ESA = b_0 + b_1 \cdot MS_{Mean}$	ESA = 13.74–0.08 · 95.40 = 5.88 sec			
	Step 3	Compute a correction factor (CF) for each player	CF = RS / ERS		CF = 6.20 / 6.74 = 0.92	CF = 6.20 / 5.91 = 1.05	CF = 6.20 / 5.50 = 1.13
	Step 4	Compute the corrected score (CS) for each player	CS = ESA · CF		CS = 5.88 · 0.9 = 5.41 sec	CS = 5.88 · 1.05 = 6.17 sec	CS = 5.88 · 1.13 = 6.63 sec

MS = maturity status; RS = raw score; ERS = expected raw score; ESA = expected score of the average maturing player; CF = correction factor; CS = corrected score; %RAH = attained percentage of the real adult height.

Research question 3: Does each univariate prediction of adult performance level improve after applying the correction mechanism?

To determine whether each corrected talent criterion was more closely linked with adult performance level (professional vs. non-professional) than each raw talent criterion, we computed point-biserial correlations—a special case of the ordinary Pearson correlation. If the correlation coefficients had noticeable descriptive differences, we tested them for significance using the *R* package cocor (Diedenhofen and Musch, 2015).

Research question 4: Does the multivariate prediction of the adult performance level improve significantly after applying the correction mechanism?

Answering the fourth question required three main analytical steps: (1) regression modelling, (2) estimating receiver operating characteristic curves, and (3) applying the DeLong-Test.

Regression modelling: We estimated eight binary logistic regression models that included eight motor talent criteria (Table 1; *z*-values). Model one served as the reference model; raw test scores were used as input (raw motor performance model). Depending on results from research question 1, the seven other models included motor performance scores corrected based on each corresponding maturity indicator (e.g., model with test scores corrected according to Mirwald). We compared predictions from these multivariate regression models with a baseline model—a model without predictors (Field, 2018). We tested the significance of the coefficients (improvement over the baseline model) with the omnibus test of the model coefficients (Zeileis and Hothorn, 2002). Its null hypothesis (H_0) is that both models (that is, the baseline model and a regression model with eight predictors) predict the adult performance level of the players equally well. We tested appropriate calibration with the Hosmer-Lemeshow test (Hosmer et al., 2013). The model fit was quantified using Nagelkerke-R^2 (R^2_N). Each regression model generated the probability of being classified as professional or non-professional for all players based on a combination of (raw or corrected) motor performance. If a player's predicted probability exceeded 0.5, the player was classified as professional.

Receiver operating characteristic analysis (ROC): The predicted probabilities we obtained in each regression analysis were used to create non-parametric ROC curves. In other words, each ROC curve represented one of the regression models. The area under each curve (AUC) can be used as an index of the classification's (discriminative ability) overall quality since it reflects the relationship between sensitivity (correctly identified professionals) and specificity (correctly identified non-professionals) (Hanley, 1989). The more players are correctly classified by a regression model, the closer the AUC is to 1. The following guidelines have been suggested (Hosmer et al., 2013) to facilitate the interpretation of the AUC: 0.9–1.0 (outstanding discrimination), 0.8–0.9 (excellent discrimination), 0.7–0.8 (acceptable discrimination), and 0.5–0.7 (poor discrimination).

DeLong-Test: We compared the AUC of each of the seven corrected regression models along with the raw regression model by employing the DeLong-Test. The H_0 states that the raw and each corrected model do not differ in their ability to predict a dichotomous criterion (professional vs. non-professional) over 6 years (H_0: $\Delta AUC = 0$).

We set alpha levels for all significance tests to 0.05. However, in view of the seven model comparisons, we adjusted the significance level for research question 4 using Bonferroni's method ($\alpha = 0.05/7 = 0.007$). We analyzed data using IBM SPSS version 27.

Results

We provide means and standard deviations of all variables in the supplementary material (Table S1). According to all maturity indicators except %PAH ($r = 0.21$; $p = 0.023$), at the age of 15, the maturity status of the 47 professionals and 74 non-professionals did not differ significantly (Table 4).

Research question 1: Does maturity status influence motor performance scores?

Table 5 confirms the need for a correction mechanism by displaying significant maturational influences ($0.16 \leq |r| \leq 0.72$; $ps < 0.05$), which were more pronounced in functional capacities than in motor skills and followed the same principle: the more mature, the better the player performed. All negative correlations referred to motor tasks with reverse coding (the shorter the time needed, the faster and the better) and therefore do not contradict this principle.

Research question 2: Is the correction mechanism effective?

As expected, all significant correlations in Table 5 were reduced to (almost) zero and became insignificant ($0.00 \leq r \leq 0.03$) after applying the correction mechanism.

Research question 3: Does each univariate prediction of adult performance level improve after applying the correction mechanism?

Table 6 illustrates significant correlations between raw motor talent criteria of sprinting, agility, dribbling, juggling, and shooting and adult performance level ($0.16 \leq |r| \leq 0.29$, $ps < 0.05$). Thus, some motor variables appear useful for making predictions. In most cases, the predictive validity of these variables barely changed after we applied the correction mechanism. We only observed slight descriptive improvements or deteriorations in the correlations with adult performance levels after application of the correction mechanism (e.g., sprint raw: $r = -0.24$ vs. sprint corrected with %RAH: $r = -0.29$ vs. sprint corrected with %PAH: $r = -0.13$). However, none of these differences were significant.

Research question 4: Does the multivariate prediction of the adult performance level improve significantly after applying the correction mechanism?

All eight regression models (one with raw scores, and seven with corrected scores when needed; Table 7) were statistically significant (omnibus test of model coefficients:

$ps < 0.05$) and appropriately calibrated (Hosmer-Lemeshow test: $ps > 0.05$). The models explained between 19% and 28% (Nagelkerke's R^2) of the variance among adult performance level and correctly classified 64.5% to 68.6% of the cases.

However, DeLong tests revealed that no model based on the corrected scores was significantly better than those based on raw ones (see Table 8). Consequently, all regression models (raw or corrected) demonstrated similar predictive validity ($-0.02 \leq \Delta AUC \leq 0.03$, $ps > 0.05$).

Table 4. Point-biserial correlations between each maturity indicator and adult performance level (professional vs. non-professional; N = 121).

	Mirwald	Moore-1	Moore-2	Fransen-1	Fransen-2	%PAH	%RAH
Adult performance level	0.13	0.14	0.10	0.13	0.13	**0.21**	0.02

%PAH = attained percentage of the predicted adult height; %RAH = attained percentage of the real adult height. Coding of the adult performance level (0 = non-professional, 1 = professional). Critical value of Pearson's r (α = 5%, one-tailed, df = 119): $|r_{crit}| = 0.15$.

Table 5. Pearson's correlation coefficients between raw performance scores and maturity status (N = 121).

Indicators of maturity status	Performance test							
	Sprint	Jumping	Yo-Yo	Agility	Dribbling	Passing	Juggling	Shooting
Mirwald	**–.71**	**.37**	**.32**	**–.25**	**–.20**	**–.18**	.08	.00
Moore-1	**–.70**	**.37**	**.33**	**–.27**	**–.20**	**–.18**	.11	.01
Moore-2	**–.65**	**.35**	**.29**	**–.16**	–.13	–.05	.03	–.01
Fransen-1	**–.67**	**.34**	**.32**	**–.27**	**–.23**	**–.24**	.11	.02
Fransen-2	**–.66**	**.33**	**.31**	**–.27**	**–.23**	**–.25**	.11	.02
%PAH	**–.72**	**.37**	**.35**	**–.28**	**–.25**	**–.22**	.10	–.01
%RAH	**–.63**	**.40**	**.36**	**–.18**	**–.24**	–.13	.06	.00

Yo-Yo = Level 1 Yo-Yo intermittent recovery test; %PAH = attained percentage of the predicted adult height; %RAH = attained percentage of the real adult height. Measurement units: seconds (sprint, agility, dribbling, passing), centimeters (jumping), meters (Yo-Yo), and points (shooting, juggling). Critical value of Pearson's r (α = 5%, one-tailed, df = 119): $|r_{crit}| = 0.15$.

Table 6. Point-biserial correlations between each talent criterion (raw and corrected test scores) and the adult performance level (professional vs. non-professional; N = 121).

Type of correction	Talent criterion							
	Sprint*	Jumping	Yo-Yo	Agility*	Dribbling*	Passing*	Juggling	Shooting
None (raw scores)	**–.24**	.02	.13	**–.16**	**–.26**	–.15	**.20**	**–.18**
Mirwald	**–.20**	–.03	.10	–.13	**–.24**	–.13	—	—
Moore-1	**–.20**	–.03	.10	–.12	**–.24**	–.13	—	—
Moore-2	**–.23**	–.02	.12	–.14	—	—	—	—
Fransen-1	**–.20**	–.03	.10	–.12	**–.24**	–.12	—	—
Fransen-2	**–.20**	–.03	.10	–.12	**–.24**	–.12	—	—
%PAH	–.13	–.06	.07	–.10	**–.22**	–.11	—	—
%RAH	**–.29**	.01	.13	**–.16**	**–.26**	—	—	—

Each cell represents the correlation between adult performance level and scores in a test (corrected or uncorrected). For example, the value in the last cell of the first column (r = –.29) represents the correlation between adult performance level and scores in the sprint test corrected according to %RAH. Yo-Yo = Level 1 Yo-Yo intermittent recovery test; %PAH = attained percentage of the predicted adult height; %RAH = attained percentage of the real adult height. — = not needed since the correlation between motor performance test and maturity indicator was not significant in Table 5. Coding of the adult performance level (0 = non-professional, 1 = professional). Measurement units: seconds (sprint, agility, dribbling, passing), centimeters (jumping), meters (Yo-Yo), and points (shooting, juggling). Critical value of Pearson's r (α = 5%, one-tailed, df = 119): $|r_{crit}| = 0.15$. *Motor tasks with reverse coding (the shorter the time needed, the faster and the better).

Table 7. Goodness-of-fit and classification accuracy of the binary logistic predictive models predicting adult performance level (professional vs. non-professional; N = 121).

Type of correction used in predictive motorperformance model	Omnibus test of model coefficients		Hosmer-Lemeshow test		Model fit	Correct classification
	$\chi^2(8)$	p	$\chi^2(8)$	p	R^2_N	%
None (raw scores)	**23.53**	.003	6.02	.645	0.24	64.5
Mirwald	**21.40**	.006	4.84	.775	0.22	66.9
Moore-1	**21.39**	.006	12.11	.146	0.22	67.8
Moore-2	**23.02**	.003	6.12	.634	0.24	64.5
Fransen-1	**21.53**	.006	2.45	.964	0.22	65.3
Fransen-2	**21.51**	.006	4.24	.835	0.22	66.1
%PAH	**18.40**	.018	8.79	.361	0.19	68.6
%RAH	**28.23**	<.001	3.79	.876	0.28	66.1

%PAH = attained percentage of the predicted adult height, %RAH = attained percentage of the real adult height.

Table 8. Comparison of the discrimination ability of the predictive models with DeLong-Test (*N* = 121).

Types of correction mechanisms compared in motor performance models	AUC [95% CI]$_{corrected\ scores}$*	Δ = AUC [95% CI]$_{corrected\ scores}$ - AUC [95% CI]$_{raw\ scores}$†	*z*	*p*
Mirwald	.73 [.63; .82]	.00 [–.05; .04]	–0.10	.922
Moore-1	.73 [.64; .82]	.00 [–.04; .05]	0.06	.950
Moore-2	.73 [.64; .82]	.00 [–.03; .03]	0.15	.879
Fransen-1	.73 [.64; .82]	.00 [–.05; .05]	–0.02	.981
Fransen-2	.73 [.64; .82]	.00 [–.05; .05]	0.04	.972
%PAH	.71 [.61; .80]	-.02 [–.07; .02]	–0.92	.356
%RAH	.76 [.67; .84]	.03 [–.01; .07]	1.33	.182

%PAH = attained percentage of the predicted adult height, %RAH = attained percentage of the real adult height; AUC = area under the curve. Adjusted α-level (Bonferroni correction) = 0.007. *AUC [95% CI]$_{corrected\ scores}$ = These statistics represent the AUC based on the regression model using corrected scores. †AUC [95% CI]$_{raw\ scores}$ = This statistic represents the AUC based on the binary regression model using raw scores and corresponds to 0.73 [0.64; 0.82].

Discussion

Our results replicate existing data and information about maturity-associated variations: Maturity status correlates with motor performance scores and affects functional capacities, such as sprinting and jumping, more strongly than motor skills, such as juggling (Meylan et al., 2010; Albaladejo-Saura et al., 2021). For scientific talent identification models operating within a summative approach, our results imply that input-output operations are confounded—the more mature the player, the better input (test scores) and output (total score: predicted probability of being classified as professional in a binary logistic regression). In line with previous studies (e.g., Romann and Cobley, 2015; Abbott et al., 2021b; Larochelambert et al., 2022), our correction mechanism served its purpose: Any correlation between motor performance test scores and the confounder (maturity status) was removed. Yet contrary to our expectations, univariate and multivariate predictions of adult performance level did not substantially improve with corrected scores.

Why did the predictions not improve? Revising expectations and speculating about when the correction mechanism might (not) work

An answer to this question emanates from two lines of thought: (1) characteristics of the variables included in the correction and prediction process (initial predictive validity of the motor variable, the initial maturity bias and the validity of the maturity indicators) and (2) the selection system in which predictions have been made.

Characteristics of the variables included in the correction and prediction process

In the case of univariate prediction, our results indicate that the correction mechanism failed generating better predictive variables. We suggest three possible reasons for this failure (Figure 1): 1) the variables have no predictive validity per se (the correction procedure cannot improve something, i.e. the predictive value, that doesn't exist); 2) the corrected and raw test scores are very similar (similar scores do not entail different predictive results; y-axis); 3) the corrected scores are not meaningful (they still contain at least the same number of errors as the raw scores; x-axis).

The first reason mentioned seems conceivable (the motor variables considered have no predictive validity), however, we can hardly argue in this direction—on the one hand, because most of our results (Table 6), and on the other hand, because the empirical discourse related to our motor variables does not support it (Leyhr et al., 2018; Höner et al., 2021; Höner et al., 2017; Sieghartsleitner et al., 2019a). Consequently, we elaborate more deeply on the two other reasons that came to mind by means of Figure 1.

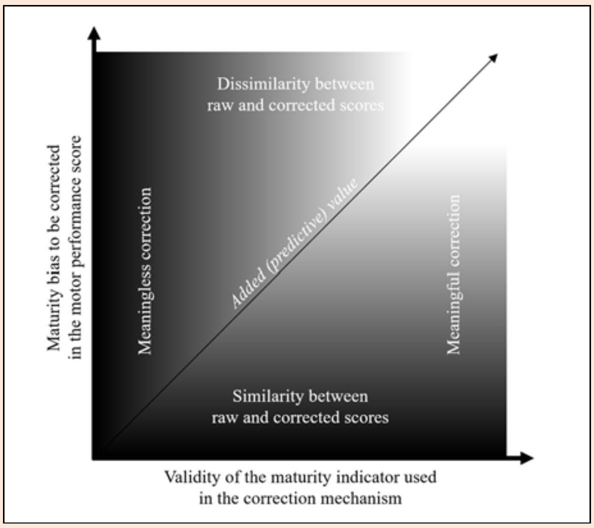

Figure 1. Hypothetical relationship between the validity of the maturity indicator used in the correction mechanism (x-axis), the maturity bias in the motor performance test to be corrected (y-axis), and their respective consequences (degree of meaningfulness and dissimilarity) regarding the generated corrected scores.

The first reason mentioned seems conceivable (the motor variables considered have no predictive validity), however, we can hardly argue in this direction—on the one hand, because most of our results (Table 6), and on the other hand, because the empirical discourse related to our motor variables does not support it (Leyhr et al., 2018; Höner et al., 2021; Höner et al., 2017; Sieghartsleitner et al., 2019a). Consequently, we elaborate more deeply on the two other reasons that came to mind by means of Figure 1.

The y-axis of Figure 1 implies that the similarity of raw and corrected scores impact their differences in predictive validity. The lower the correlations between motor performance tests and maturity indicators (Table 5), the smaller the confounding influence and (need for) adjustment and by extension, the greater the similarity between raw and corrected scores and their predictive validity. In other words, no maturity-bias does not require correction, no correction means unchanged scores, and unchanged scores imply unchanged prediction. Accordingly, since the sprint test is the only one displaying strong correlations with maturity status (see Table 5), it is probably the best candidate for raw and corrected scores that differ sufficiently to show a noticeable change in predictive validity.

The x-axis of Figure 1 suggests that a positive change in predictive validity depends on meaningful correction mechanism foundations (validity of the chosen maturity indicator). An invalid maturity indicator is unlikely to generate a valid (meaningful) corrected score to make a valid prediction. Even though some literature shows that pragmatic maturity indicators are sufficient to economically inform coaches about youth player maturity status (Romann et al., 2017; Leyhr et al., 2020a; Malina et al., 2012), serious validity concerns as a function of age and stage of maturation have also been raised (Malina and Kozieł, 2014; Teunissen et al., 2020; Parr et al., 2020). The lower the validity of the chosen maturity indicator (left side, Figure 1), the greater the loss of the correction and predictive process with respect to meaningful informative value. Since our sample characteristics (15 years-old, highly selective, more mature than same-aged non-athletes) match the criteria when pragmatic maturity indicators (especially the offset methods) become less accurate (Malina and Kozieł, 2014; Malina et al., 2015), the loss in meaningful informative/predictive value may be substantial. Notably, the predictive validity of all the scores corrected with a pragmatic maturity indicator tends to be slightly lower than that of the raw scores (see Table 6). Hence, if one is seeking a talent criterion that is expected to exhibit a noticeable increase in its predictive validity after applying the correction mechanism, it will probably be found among the ones that have been corrected with more objective and reliable methods (here, %RAH; or in general, skeletal age or real age at PHV; Malina, 2017).

Against the backdrop of these reflections, our expectation to systematically create notably better predictive variables by simply applying a correction mechanism may have been unrealistic. Not all test variables were predisposed to a noticeable improvement in their predictive validity after applying the correction mechanism. In fact, combining the three determining factors (1. Has the variable some predictive validity? 2. How similar are corrected and uncorrected scores? 3. How meaningful is the foundation of the correction procedure?), it appears that—among all test variables—the sprint test when corrected according to our most objective and reliable method (%RAH) was the only candidate where an improvement in predictive validity after the correction mechanism could be envisaged. Unsurprisingly, it is also the only corrected variable with a (descriptively) better predictive validity than its raw version (see Table 6). Although this predictive benefit remains trivial and non-significant, our data and reasoning suggest the closer to the top right corner of Figure 1, the more promising the application of a correction mechanism might be.

Role of selection systems in which predictions have been made

The selection system in use when we conducted our study might also be responsible for a lack of predictive improvement in the multivariate context (research question 4; Table 8). It is clear that the selection-related decision-making process depends primarily on evaluations of match performance and subjective insights from coaches and scouts (Roberts et al., 2019; Christensen, 2009). Nevertheless, it seems reasonable to assume that coaches' selection decisions are based on available information—the players' characteristics (Lath et al., 2021; Bergkamp et al., 2022; Lüdin et al., 2022). Thus, information systematically collected (for selection recommendation purposes) or visually noticeable (in training and competition contexts) probably have more influence (in the sense of decision-making power) on coach selection decisions than uncollected or unnoticed information (Lüdin et al., 2022). This distinction is particularly important because raw motor performances are observed by coaches on a day-to-day basis and systematically tracked as part of the talent identification program in Switzerland (Schweizer Fussballverband, 2016; Fuchslocher et al., 2011), while the corrected scores were nonexistent during the study duration. In other words, it was simply impossible for corrected scores to influence the probability of survival of early and late maturing players in the elite-sport system. Contrastingly, through the (maturity-biased) tracks raw performances left behind (the more mature, the better the raw scores, the better the impression; Hancock et al., 2013; Hill et al., 2021; Cripps et al., 2016; Furley and Memmert, 2016), they could actively shape coaches' opinions, selection chances, final decisions, and by extension, chances to access continuous professional support (Meylan et al., 2010; Malina et al., 2015). However, whether and the extent coaches were actually influenced by maturity-biased raw motor performances is not the focal point here. The core issue is raw scores, unlike corrected scores, had decision-making power. Unfortunately, developmental research has no way of experimentally manipulating the temporal component to compare whom coaches would select if they had access to sources of information not confounded by maturity status. Yet, it seems difficult to refute the hypothesis that maturity-biased and -unbiased performance considerations may lead to different conclusions about players and thus, to different selection decisions (Lüdin et al., 2022). In turn, it suggests that raw scores had a perceptible (and non-compensable) advantage over corrected scores. Is this advantage somehow reflected statistically? We suggest it is—with more decision-making power, raw scores carried more (predictive) information than corrected scores.

The potential of a multivariate regression model to discriminate between future professionals and non-professionals (AUC) is expected to improve, the better the model fits the data (Steyerberg, 2019). Therefore, our attention should be focused on the extent, if at all, the correction mechanism leads to better-fitting models (Nagelkerke-R^2). Typically, adding relevant information, such as additional predictors, to models leads to improvements in model fit (Field, 2018). The more information a statistical model can work with, the better it is expected to perform (Field, 2018). For corrected scores, it is rather unfortunate because the correction mechanism does not add, but rather extracts everything related to biological maturation in each motor performance test (b + e; see Figure 2) (i.e., the correlation between corrected score and maturity status is equal to zero; research question 2). Figure 2 highlights three important points:

1) Before applying the correction mechanism, the motor scores contained information related to maturity status—the maturity and motor information overlap (in the sense of confounded; raw scores = a + b + e + f).
2) After applying the correction mechanism, the maturity and motor information are separated—the motor performance tests are free from this confounding influence (corrected scores = a + f).
3) The dependent variable (career outcome: professional vs. non-professional) remains unchanged during the entire process (a + b + c + d).

If we accept these three points, we have to assume that the extent of change in the model fit (R^2) after the correction process primarily relies on the correction-related absence of components b and e (however, no one is really interested in component e, because e has nothing to do with the career outcome). Accordingly, the bigger component b is (the more biological maturation embedded in raw scores has predictive validity), the more the motor scores lose their predictive power after applying the correction mechanism when compared with raw scores. Since existing literature generally acknowledges that early maturers are more likely to be selected in football academies than late maturers (Johnson et al., 2017), the size of component b is likely to be substantial. Thus, as long as the selection process is (even slightly) biased by the maturity status, the model fit of a regression model using corrected scores will theoretically always be worse than the one using the raw scores only. Does our data support this view? It seems so.
Apart from one outlier (motor performance model corrected with %RAH; $R^2_N = .28$), which might just be an example of capitalizing on chance, our model fits always decreased when using corrected ($.19 \leq R^2_N \leq .24$) instead of raw test scores ($R^2_N = .24$) to predict player career outcomes. In concrete terms, the stronger the correlation between maturity indicator and adult performance level (component b in Figure 2; Table 4), the greater the decrease in model fit (Table 7). In line with this, the largest reduction we observed in model fit ($R^2_N = .19$) after applying the correction mechanism was in the multivariate regression model using test scores corrected with maturity indicators having the highest correlation with adult performance level (%PAH, Table 4). Consequently, it suggests a methodological artifact: Raw scores had a (methodological and non-compensable) predictive advantage over corrected scores.

To sum up, expecting the application of a correction mechanism to automatically improve predictions from multivariate regression models was unrealistic. It overlooked existing selection system influence during the study duration, which produced a methodological artifact inflating raw score predictive power. The methodological artifact can seemingly be avoided only if the career outcome (professional vs. non-professional; a + b + c + d in Figure 2) emerges from an intervention manipulating the selection system during puberty (control vs. intervention group).

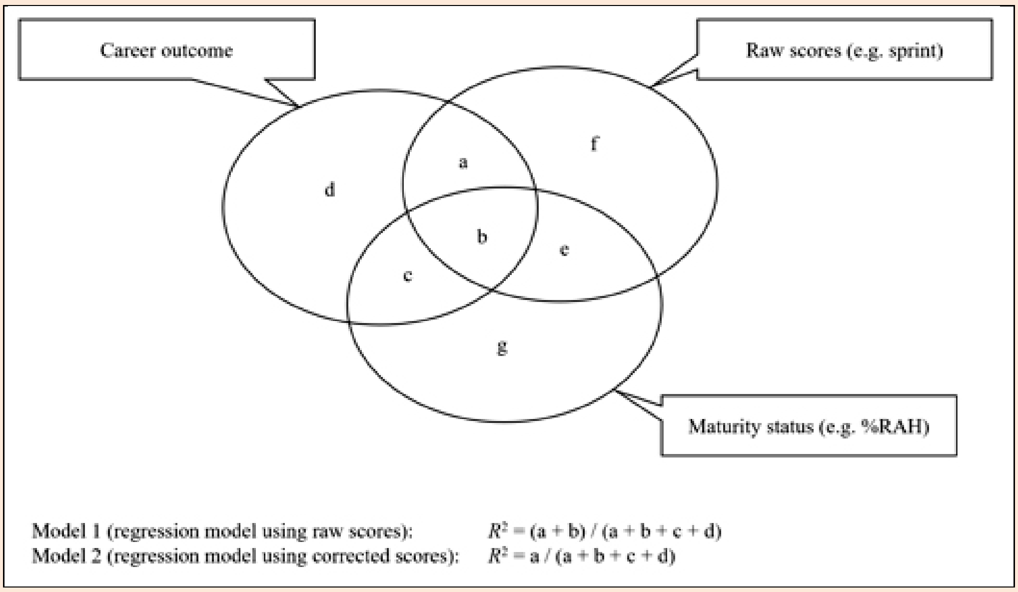

Figure 2. Variance components for the prediction of adult performance level (career outcome) illustrated with the sprint scores confounded by the attained percentage of the real adult height (%RAH).

In the control group, players are selected annually according to their raw scores (without applying the correction mechanism). Players in the intervention group are selected annually according to their corrected scores (with applying the correction mechanism). Since the selection process is not affected by maturity-related selection errors in the intervention group, the right players—the players with the highest (motor) potential—are chosen. Accordingly, on average, players selected from the intervention group are expected to reach a higher adult performance level than players selected from the control group. However, to implement such an intervention, clubs and federations would need to relinquish control over talent selection decisions and agree to use a unidimensional talent identification model based on motor performance, which is inconceivable and definitively outdated (Williams et al., 2020). Therefore, the approach adopted in this study may be the only way to empirically introduce and theoretically discuss potential benefits associated with corrected scores.

Practical implications

Taken together, our results indicate that correction mechanism produce principally-sound corrected scores as valid as raw scores from a predictive point of view (Table 6 and Table 8). What practical implications can be drawn from this predictive equivalence, or rather, is a correction recommended or not?

To correct or not to correct? That is the question

Because coaches are encouraged to use scientific models in the talent identification process (Sieghartsleitner et al., 2019b; Williams et al., 2020) and corrected scores are available, it becomes necessary to reflect upon what kind of scientific test scores (raw or corrected) should be relied on. From a methodological perspective, it can be argued that compared with raw scores, corrected scores have limited utility since more work from implementing a correction mechanism is needed for the same overall predictive output. This violates the principle of parsimony in building models (Field, 2018), which "tells us to remove what is unnecessary" (Sober, 1981, p. 145) and speaks against their application.

Although one can see the glass as half empty (same output for more work), one can also see it as half full (same output, but one fewer problem). In the words of Karl Popper (2000):

> *A new hypothesis is only taken seriously if it explains at least everything that was successfully explained by its predecessor, and if; in addition, it either promises to avoid particular errors of the old hypothesis or makes new predictions—where possible, testable predictions. (p. 39)*

Accordingly, from a critical rationalist point of view, corrected scores should be favored because they predict at least everything successfully predicted by their predecessors (raw scores). In addition, they promise to avoid the acknowledged problem of maturity-biased inter-individual performance comparisons in scientific models and make new testable predictions (for example, see Figure 1). By correcting, and thus changing each player's scores, the correction mechanism changes the position of each player on any kind of talent continuum based on one of several summative approaches (z-scores addition, Turner, 2014; Souza-Lima et al., 2020; weighted scoring system, Fuchslocher et al., 2011; Höner et al., 2015; regression, Sieghartsleitner et al., 2019a). Those who previously (and erroneously) did not emerge as candidates for selection could receive this status after applying a correction mechanism (and vice versa). At the same time, it may increase the acceptance of scientific forecasting in practice. Indeed, the more our models avoid mispositioning players on the talent continuum due to early or late maturation, the more the models will be taken seriously.

Correction recommendations are in vogue: similar rationale, different implementation objects

Interestingly, on some level the correction mechanism is not drastically different from the bio-banding strategy, which has garnered a reputation among applied sports scientists in recent years (Cumming et al., 2017; Cumming et al., 2018a; Reeves et al., 2018a; Towlson et al., 2021a). Both approaches "correct" something to solve a maturity-related problem for one of the complementary parts of the talent identification process, that is coaches and scientific models (Lath et al., 2020; Sieghartsleitner et al., 2019b). Bio-banding "corrects" the playing environment and allows coaches to observe how good each player would be if they played in an environment without maturity-related (dis-) advantages (Cumming et al., 2017; Malina et al., 2019; Rogol et al., 2018). It is supposed to enable coaches to better detect talented players. The correction mechanism seeks to achieve a similar goal. It corrects the test scores mathematically and allows scientific talent identification models to estimate how good each player would be in each test, if he was as biologically developed as the others. In turn, it is supposed to help scientific models better detect talented players in the future.

Furthermore, from a talent development perspective, just as bio-banding changes the behaviors and experiences of each player on the field (Cumming et al., 2018a), the correction mechanism would not be impervious to side effects. Specifically, according to Vallerand and Losier's (1999) integrative analysis of motivation, each (social) context influences the basic psychological needs (e.g., perceived competence; Deci and Ryan, 1985), which in turn increase or undermine one's intrinsic motivation and lead to a host of consequences (e.g., dropout intention, negative or positive emotions; Balish et al., 2014; Pelletier et al., 2001; Sarrazin et al., 2002). Thus, a context which is characterized by comparisons based on corrected (instead of raw) scores might help late maturers build confidence in their ability (satisfaction of need for competence), which in turn might increase their intrinsic motivation and decrease their possible dropout intentions. In contrast, corrected scores might threaten early maturers' experience of competence, leading to less intrinsic motivation and negative emotions such as frustration, dissatisfaction, and fear of disappointing significant others. Therefore, players should not be left alone with their corrected scores, rather coaches have to explain beforehand the meaning behind such correction (in the sense of individualized feedback). In particular, correction-related feedback should make late maturers realize that their feeling incompetent might be unwarranted and that in all likelihood they will catch up with the others. Inversely, the correction-related feedback should make early maturers realize that their feeling competent might be, to some extent, unjustified and that they need to continue to work upon a particular aspect of their game, because in all likelihood the other players will catch up. In this respect, implementing a correction mechanism responds to a recent call for strategies creating learning contexts that encourage early and late maturing boys to develop more adaptative psychological skill sets (Cumming et al., 2018b). It is worth noting that such maturity-unbiased contexts may change players', parents' and coaches' expectations regarding achievement. Theoretically, these expectations are not devoid of consequences, since they may actively shape talent development paths through self-fulfilling prophecies, such as the Pygmalion, the Galatea and the Golem effect (when higher/lower other-/self-expectations lead to higher/poorer career outcomes; Hancock et al., 2013; Leonardo Filho, 2016; Babad et al., 1982; Rosenthal and Jacobson, 1968).

To sum up, similar to bio-banding strategy (Cumming et al., 2017), corrected scores provide a new piece of information and mirror reality in a restructured, more nuanced way (Larochelambert et al., 2022; Abbott et al., 2021b). They refine judgment and might affect players' personality and expectations. As a result, corrected scores may also impact career outcomes. Given their perceived highly subjective usefulness for maturity-unbiased, inter- and intra-individual performance comparisons, and equivalent predictive validity, we recommend the use of corrected scores to complement current talent identification and development methods. In this context, one of the strengths of maturity-based correction mechanisms is their application in all sports, which responds to a recent call for scientific solutions with an impact beyond a single sport (Abt et al., 2022).

Limitations

There are four noteworthy study limitations. First, we focused on isolated skills and abilities in the tests we used (see Table 1). Thus, we could not completely capture the unpredictable and complex nature of football. For instance, the test assessing dribbling skill, not unlike a slalom around stationary markers, places a high reliance on the player's ability to accelerate. Unfortunately, it misses some aspects of the cognitive, perceptual and motor skills that are essential to dribbling the ball in match conditions (McDermott et al., 2015; Ali, 2011). The same kind of criticism applies to the other tests. However, limited ecological validity was the price to pay to maximize closeness to practice—i.e., to investigate the predictive validity of the specific tests used in German and Swiss talent identification programs (Höner et al., 2015; Schweizer Fussballverband, 2016). Furthermore, while it is true that an increase in the complexity of the test battery should improve (ecological or predictive) validity, it could at the same time impair the reliability of the tests (see reliability-validity dilemma; Slomp and Fuite, 2004; Höner and Roth, 2002). Indeed, we had serious concerns that less reliable tests would provide less trustworthy corrected scores.

Second, we considered biological maturation as a finite and linear resource, which can only have a unidirectional effect on motor performances (the more mature, the better). Yet clearly our conceptualization was reductionist and failed to consider possible problems, such as maturity-related altered motor control also known as "adolescent awkwardness," emerging at certain points during development (Quatman-Yates et al., 2012). Indeed, some athletes experience performance drops due to accelerated periods of growth, which could not be captured or considered in our correction procedure. Therefore, it remains unclear to what extent it is legitimate to linearize the phenomenon of maturation, which is primarily regarded as non-linear (Boeyer et al., 2020) and to correct its influence on talent criteria with linear methods. Other modelling approaches may be adopted in future studies (e.g., curvilinear; Abbott et al., 2021b; Abbott et al., 2021a).

Third, we interpreted results guided by the assumption that %RAH was a highly valid indicator of the somatic biological maturation of the players. However, its validity in our study is in question. We did not directly measure somatic biological maturation; we asked about it in a follow-up survey or collected it through transfermarkt. Both data collection methods are not without problems. Unlike other studies in talent research (Leyhr et al., 2020a), we assumed players did not overestimate their current adult height. Thus, we decided not to adjust self-reported adult height according to Epstein's equation (Epstein et al., 1995). We believe these adjustments would have had limited bearing on our study findings since they are generally small in nature. Transfermarkt relies on the principle of crowd wisdom to generate information (Peeters, 2018). Consequently, even though it provides reliable and precise information about players (Prockl and Frick, 2018), the validity of its data depends on the crowd's wisdom. To ascertain validity, we calculated the correlation between self-reported adult height and adult height found on transfermarkt for a sub-sample of 12 players for whom data were available ($r_s = 0.99$). In our case, the crowd seemed wise and our assumptions about %RAH are valid. However, future studies should reflect on the necessity of Epstein's formula in the context of talent research and further examine the margin of error associated with adult height indicated on transfermarkt.

Fourth, our sample was (a) highly selected and (b) relatively small. It resulted in a reduced dispersion of maturity status, which may have deflated, for example, the reported relationships between maturity status and motor performances. Our small sample size meant we could not split the dataset into exploratory and validation subsamples without violating important requirements of statistical analysis. For this reason, we evaluated present models using only the exploratory dataset and failed meeting current standards of machine learning practice (Till et al., 2016). Consequently, our regression models are only internally valid and may not generalize to the population (Steyerberg et al., 2010). Considering these four limitations, our results seem trustworthy, yet need replication.

Future directions

On a methodological level, further research could attempt to reproduce our findings with different operationalizations of maturity status, such as with Khamis-Roche method (Khamis and Roche, 1994), sexual and/or skeletal maturity indicators (Malina et al., 2015), different sample characteristics, like those where pragmatic maturity indicators are more accurate (Malina and Kozieł, 2014), different correction mechanisms (e.g., curvilinear instead of linear; Abbott et al., 2021b or size- and/or body-mass-related instead of maturity-related; Valente dos Santos et al., 2014a; Cunha et al., 2011), or other possibly confounded talent criteria, like psychological skill (Cumming et al., 2012; Cumming et al., 2018b). It would also be interesting to extend the correction mechanism to further variables confounding the input-output operations of additive-linear scientific models and explore how the long-term predictive validity responds. Potential candidates are relative (chronological) age (Leyhr et al., 2021; Romann et al., 2018; Votteler and Höner, 2014) or training age (Johnston and Baker, 2020; Guimarães et al., 2019).

We invite sports physiologists to further disentangle the direct, indirect, and mediating effects of biological maturation on performance in motor tests (e.g., body size descriptors, such as stature and body mass, not only predict maturity status, but also motor performance, Valente dos Santos et al., 2012; Valente dos Santos et al., 2014b). Sports psychologists might be interested in determining whether knowledge of corrected scores results in changes in player evaluation (as this is the case for knowledge of maturity status on shirt number; Lüdin et al., 2022), player personality (e.g., basic psychological needs, football-related self-concept, self-efficacy, motivational characteristics), support from parents or coaches, or team role and status (Cumming et al., 2012; Schmidt et al., 2015; Cumming et al., 2006; Eisenmann et al., 2020).

The correction mechanism also has applications in training science, namely as complement to developmental fitness curves (Myburgh et al., 2020; Owen et al., 2022; Till et al., 2022; Cumming, 2018) for (a) profiling strengths and weaknesses and (b) monitoring performance development during puberty. Performance profiling is meant to guide coaches so they better tailor training to player needs (Till et al., 2018; Eisenmann et al., 2020). Figure 3 shows a fictitious (late-maturing) player who appears to have a raw score-based major sprinting weakness. After correction (in red), it becomes clear that the weakness is not problematic. His corrected sprint performance is above average. However, without information provided by the correction mechanism, strength and conditioning coaches might invest time for sprint training, which could be better utilized elsewhere. Regarding the issue of monitoring, strength and conditioning coaches struggle to separate training effects from maturation effects after training intervention, such as strength training (van Hooren and Ste Croix, 2020; Moran et al., 2017; Till et al., 2018). By making pre-post comparisons with corrected rather than uncorrected scores, coaches better capture true training effects and better understand athlete progress (Eisenmann et al., 2020; Till et al., 2018).

Finally, on a more general note and beyond the realm of youth elite sports, the correction mechanism could be applied in school sport settings. Indeed, knowing maturity status influences physical activity patterns and motor performance, which, in turn, influences personality development and creates motivational basis for lifelong sports participation (Cumming et al., 2020; Cumming et al., 2012; Conzelmann and Schmidt, 2020), is it fair or expedient when physical education teachers evaluate students according to their raw performance during puberty (the more mature the adolescent, the better the performance; Nevill et al., 2021; Vist Hagen et al., 2022)?

Thus, a series of new research questions and options for practical application arise with available corrected scores.

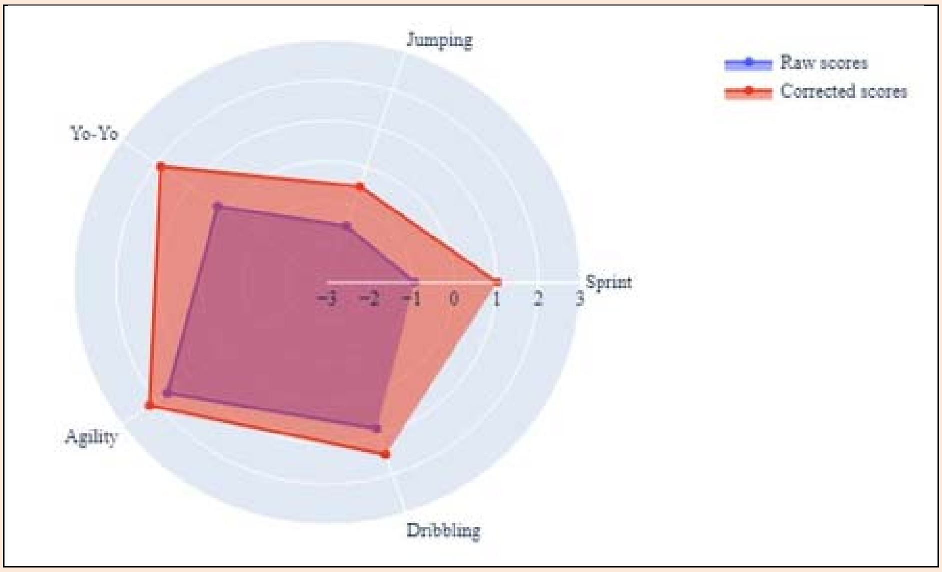

Figure 3. **Fictitious performance profile (*z*-scores) of a player before and after the application of the correction mechanism (raw scores = blue; corrected scores = red; correction based on %RAH). Coding of sprint, agility and dribbling performance has been reversed so that a higher value corresponds to better performance.**

Conclusion

Our study confirmed (1) that maturity status influenced some motor talent criteria and (2) maturity-based correction mechanisms worked. Contrary to our expectations, corrected scores (3) considered one by one (univariate prediction) or (4) in combination (multivariate prediction) were not found to be better predictors of who became a professional athlete six years later than uncorrected scores. In our opinion, predictive equivalence does not imply the futility of correction mechanisms (more work for the same output), but it underlines their potential importance (same output, one fewer problem) and the need to revise our correction-related expectations according to four factors: initial predictive validity of considered motor variables, validity of chosen maturity indicators, initial maturity-bias, and analyzed cohort selection system. Even though the recommendation to use maturity-based correction mechanisms seems theoretically legitimate, its empirical basis (from a predictive point of view for talent identification) remains to be established.

Acknowledgements

We thank the Swiss Football Association for supporting and funding this research project. We thank Kristin Marie Bivens for her editorial guidance and support. The experiments comply with the current laws of the country in which they were performed. The authors have no conflicts of interests to declare. The datasets generated and analyzed during the current study are not publicly available but are available from the corresponding author who was an organizer of the study.

References

Abbott, S., Castiglioni, M., Cobley, S., Halaki, M., Hogan, C., Mitchell, L.J.G., Romann, M., Salter, J. and Yamauchi, G. (2021a) Removing maturational influences from female youth swimming: the application of corrective adjustment procedures. *Journal of Science and Medicine in Sport* **24**, 38. https://doi.org/10.1016/j.jsams.2021.09.098

Abbott, S., Hogan, C., Castiglioni, M.T., Yamauchi, G., Mitchell, L.J.G., Salter, J., Romann, M. and Cobley, S. (2021b) Maturity-related developmental inequalities in age-group swimming: The testing of 'Mat-CAPs' for their removal. *Journal of Science and Medicine in Sport* **24(4)**, 397-404. https://doi.org/10.1016/j.jsams.2020.10.003

Abbott, S., Moulds, K., Salter, J., Romann, M., Edwards, L. and Cobley, S. (2020) Testing the application of corrective adjustment procedures for removal of relative age effects in female youth swimming. *Journal of Sports Sciences* **38(10)**, 1077-1084. https://doi.org/10.1080/02640414.2020.1741956

Abt, G., Jobson, S., Morin, J.-B., Passfield, L., Sampaio, J., Sunderland, C. and Twist, C. (2022) Raising the bar in sports performance research. *Journal of Sports Sciences* **40(2)**, 125-129. https://doi.org/10.1080/02640414.2021.2024334

Albaladejo-Saura, M., Vaquero-Cristóbal, R., González-Gálvez, N. and Esparza-Ros, F. (2021) Relationship between biological maturation, physical fitness, and kinanthropometric variables of young athletes: a systematic review and meta-analysis. *International Journal of Environmental Research and Public Health* **18(1)**, 328. https://doi.org/10.3390/ijerph18010328

Ali, A. (2011) Measuring soccer skill performance: a review. *Scandinavian Journal of Medicine and Science in Sports* **21(2)**, 170-183. https://doi.org/10.1111/j.1600-0838.2010.01256.x

Babad, E.Y., Inbar, J. and Rosenthal, R. (1982) Pygmalion, Galatea, and the Golem: investigations of biased and unbiased teachers. *Journal of Educational Psychology* **74(4)**, 459-474. https://doi.org/10.1037/0022-0663.74.4.459

Balish, S.M., McLaren, C., Rainham, D. and Blanchard, C. (2014) Correlates of youth sport attrition: a review and future directions. *Psychology of Sport & Exercise* **15(4)**, 429-439. https://doi.org/10.1016/j.psychsport.2014.04.003

Bangsbo, J., Iaia, F.M. and Krustrup, P. (2008) The Yo-Yo intermittent recovery test: a useful tool for evaluation of physical performance in intermittent sports. *Sports Medicine* **38(1)**, 37-51. https://doi.org/10.2165/00007256-200838010-00004

Baxter-Jones, A.D.G. (2017) Growth and maturation. In: Oxford textbook of children's sport and exercise medicine. Ed: Armstrong, N. and van Mechelen, W. Oxford: Oxford University Press. 13-23. https://doi.org/10.1093/med/9780198757672.003.0002

Baxter-Jones, A.D.G., Eisenmann, J.C. and Sherar, L.B. (2005) Controlling for maturation in pediatric exercise science. *Pediatric Exercise Science* **17(1)**, 18-30. https://doi.org/10.1123/pes.17.1.18

Bergkamp, T.L.G., Frencken, W.G.P., Niessen, S., Meijer, R.R. and Den Hartigh, R.J.R. (2022) How soccer scouts identify talented players. *European Journal of Sport Science* **22(7)**, 994-1004. https://doi.org/10.1080/17461391.2021.1916081

Bergkamp, T.L.G., Niessen, S., Den Hartigh, R.J.R., Frencken, W.G.P. and Meijer, R.R. (2019) Methodological issues in soccer talent identification research. *Sports Medicine* **49(9)**, 1317-1335. https://doi.org/10.1007/s40279-019-01113-w

Boeyer, M.E., Leary, E.V., Sherwood, R.J. and Duren, D.L. (2020) Evidence of the non-linear nature of skeletal maturation. *Archives of Disease in Childhood* **105(7)**, 631-638. https://doi.org/10.1136/archdischild-2019-317652

Brustio, P.R. and Boccia, G. (2021) Corrective procedures remove relative age effect from world-class junior sprinters. *Journal of Sports Sciences* **39(22)**, 2603-2610. https://doi.org/10.1080/02640414.2021.1947618

Brustio, P.R., Cobley, S., Abbott, S., La Torre, A., Moisè, P., Rainoldi, A. and Boccia, G. (2022) Corrective adjustment procedures as a strategy to remove relative age effects: validation across male and female age-group long jumping. *Journal of Science and Medicine in Sport* **25(8)**, 678-683. https://doi.org/10.1016/j.jsams.2022.04.007

Casartelli, N., Muller, R. and Maffiuletti, N.A. (2010) Validity and reliability of the Myotest accelerometric system for the assessment of vertical jump height. *Journal of Strength & Conditioning Research* **24(11)**, 3186-3193. https://doi.org/10.1519/JSC.0b013e3181d8595c

Christensen, M.K. (2009) "An eye for talent": talent identification and the "practical sense" of top-level soccer coaches. *Sociology of Sport Journal* **26**, 365-382. https://doi.org/10.1123/ssj.26.3.365

Cobley, S., Abbott, S., Moulds, K., Hogan, C. and Romann, M. (2020) Re-balancing the relative age effect scales: Meta-analytical trends, causes, and corrective adjustment procedures as a solution. In: Relative age effects in sport. Ed: Dixon, J.C. et al. New York: Routledge. 136-153. https://doi.org/10.4324/9781003030737-11

Conzelmann, A. and Schmidt, M. (2020) Persönlichkeitsentwicklung durch Sport [Personality development through sport]. In: Sportpsychologie. Ed: Schüler, J., Wegner, M. and Plessner, H. Berlin, Heidelberg: Springer. 337-354. https://doi.org/10.1007/978-3-662-56802-6_14

Cripps, A.J., Hopper, L.S. and Joyce, C. (2016) Coaches' perceptions of long-term potential are biased by maturational variation. *International Journal of Sports Science and Coaching* **11(4)**, 478-481. https://doi.org/10.1177/1747954116655054

Cumming, S.P. (2018) A game plan for growth: how football is leading the way in the consideration of biological maturation in young male athletes. *Annals of Human Biology* **45(5)**, 373-375. https://doi.org/10.1080/03014460.2018.1513560

Cumming, S.P., Battista, R.A., Standage, M., Ewing, M.E. and Malina, R.M. (2006) Estimated maturity status and perceptions of adult autonomy support in youth soccer players. *Journal of Sports Sciences* **24(10)**, 1039-1046. https://doi.org/10.1080/02640410500386142

Cumming, S.P., Brown, D.J., Mitchell, S.B., Bunce, J., Hunt, D., Hedges, C., Crane, G., Gross, A., Scott, S., Franklin, E., Breakspear, D., Dennison, L., White, P., Cain, A., Eisenmann, J.C. and Malina, R.M. (2018a) Premier League academy soccer players' experiences of competing in a tournament bio-banded for biological maturation. *Journal of Sports Sciences* **36(7)**, 757-765. https://doi.org/10.1080/02640414.2017.1340656

Cumming, S.P., Harrington, D.M., Davis, M.J., Edwardson, C.L., Gorely, T., Khunti, K., Rowlands, A.V., Yates, T. and Sherar, L.B. (2020) Maturational timing, physical self-perceptions and physical activity in UK adolescent females: investigation of a mediated effects model. *Annals of Human Biology* **47(4)**, 384-390. https://doi.org/10.1080/03014460.2020.1784277

Cumming, S.P., Lloyd, R.S., Oliver, J.L., Eisenmann, J.C. and Malina, R.M. (2017) Bio-banding in sport: applications to competition, talent identification, and strength and conditioning of youth athletes. *Strength & Conditioning Journal* **39(2)**, 34-47. https://doi.org/10.1519/SSC.0000000000000281

Cumming, S.P., Searle, C., Hemsley, J.K., Haswell, F., Edwards, H., Scott, S., Gross, A., Ryan, D., Lewis, J., White, P., Cain, A., Mitchell, S.B. and Malina, R.M. (2018b) Biological maturation, relative age and self-regulation in male professional academy soccer players: A test of the underdog hypothesis. *Psychology of Sport & Exercise* **39**, 147-153. https://doi.org/10.1016/j.psychsport.2018.08.007

Cumming, S.P., Sherar, L.B., Pindus, D.M., Coelho-e-Silva, M.J., Malina, R.M. and Jardine, P.R. (2012) A biocultural model of maturity-associated variance in adolescent physical activity. *International Review of Sport and Exercise Psychology* **5(1)**, 23-43. https://doi.org/10.1080/1750984X.2011.630481

Cunha, G.D.S., Lorenzi, T., Sapata, K., Lopes, A.L., Gaya, A.C. and Oliveira, Á. (2011) Effect of biological maturation on maximal oxygen uptake and ventilatory thresholds in soccer players: an allometric approach. *Journal of Sports Sciences* **29(10)**, 1029-1039. https://doi.org/10.1080/02640414.2011.570775

Deci, E.L. and Ryan, R.M. (1985) Intrinsic motivation and self-determination in human behavior. New York: Plenum. https://doi.org/10.1007/978-1-4899-2271-7

Diedenhofen, B. and Musch, J. (2015) cocor: A comprehensive solution for the statistical comparison of correlations. *Plos One* **10(3)**, e0121945. https://doi.org/10.1371/journal.pone.0121945

Dodd, K.D. and Newans, T.J. (2018) Talent identification for soccer: physiological aspects. *Journal of Science and Medicine in Sport* **21(10)**, 1073-1078. https://doi.org/10.1016/j.jsams.2018.01.009

Doyle, J.R. and Bottomley, P.A. (2018) Relative age effect in elite soccer: more early-born players, but no better valued, and no paragon clubs or countries. *Plos One* **13(2)**, 1-13. https://doi.org/10.1371/journal.pone.0192209

Eisenmann, J.C., Till, K. and Baker, J. (2020) Growth, maturation and youth sports: issues and practical solutions. *Annals of Human Biology* **47(4)**, 324-327. https://doi.org/10.1080/03014460.2020.1764099

Epstein, L.H., Valoski, A.M., Kalarchian, M.A. and McCurley, J. (1995) Do children lose and maintain weight easier than adults: a comparison of child and parent weight changes from six months to ten years. *Obesity Research* **3(5)**, 411-417. https://doi.org/10.1002/j.1550-8528.1995.tb00170.x

Field, A. (2018) Discovering statistics using IBM SPSS Statistics. 5th edition. Los Angeles: Sage.

Figueiredo, A.J., Coelho-e-Silva, M.J. and Malina, R.M. (2011) Predictors of functional capacity and skill in youth soccer players. *Scandinavian Journal of Medicine and Science in Sports* **21(3)**, 446-454. https://doi.org/10.1111/j.1600-0838.2009.01056.x

Fransen, J., Baxter-Jones, A.D.G. and Woodcock, S. (2018a) Responding to the commentary on the article: "Improving the prediction of maturity from anthropometric variables using a maturity ratio". *Pediatric Exercise Science* **30(2)**, 7-9. https://doi.org/10.1123/pes.2017-0249

Fransen, J., Bush, S., Woodcock, S., Novak, A., Baxter-Jones, A.D.G., Deprez, D., Vaeyens, R. and Lenoir, M. (2018b) Improving the prediction of maturity from anthropometric variables using a maturity ratio. *Pediatric Exercise Science* **30(2)**, 296-307. https://doi.org/10.1123/pes.2017-0009

Fuchslocher, J., Romann, M., Rüdisüli, R., Birrer, D. and Hollenstein, C. (2011) Das Talentidentifikationsinstrument PISTE - Wie die Schweiz Nachwuchsathleten auswählt [The talent identification tool PISTE - how Switzerland selects junior athletes]. *Leistungssport* **4(2)**, 22-27.

Furley, P. and Memmert, D. (2016) Coaches' implicit associations between size and giftedness: implications for the relative age effect. *Journal of Sports Sciences* **34(5)**, 459-466. https://doi.org/10.1080/02640414.2015.1061198

Guimarães, E., Ramos, A., Janeira, M.A., Baxter-Jones, A.D.G. and Maia, J. (2019) How does biological maturation and training experience impact the physical and technical performance of 11-14-year-old male basketball players? *Sports* **7(12)**, 1-13. https://doi.org/10.3390/sports7120243

Hancock, D.J., Adler, A.L. and Côté, J. (2013) A proposed theoretical model to explain relative age effects in sport. *European Journal of Sport Science* **13(6)**, 630-637. https://doi.org/10.1080/17461391.2013.775352

Hanley, J.A. (1989) Receiver operating characteristic (ROC) methodology: the state of the art. *Critical Reviews in Diagnostic Imaging* **29(3)**, 307-335.

Herm, S., Callsen-Bracker, H.M. and Kreis, H. (2014) When the crowd evaluates soccer players' market values: accuracy and evaluation attributes of an online community. *Sport Management Review* **17(4)**, 484-492. https://doi.org/10.1016/j.smr.2013.12.006

Hill, M., Scott, S., McGee, D. and Cumming, S.P. (2021) Are relative age and biological ages associated with coaches' evaluations of match performance in male academy soccer players? *International Journal of Sports Science & Coaching* **16(2)**, 227-235. https://doi.org/10.1177/1747954120966886

Hill, M., Scott, S., McGee, D. and Cumming, S.P. (2020) Coaches' evaluations of match performance in academy soccer players in relation to the adolescent growth spurt. *Journal of Science in Sport and Exercise* **2(4)**, 359-366. https://doi.org/10.1007/s42978-020-00072-3

Hogan, C., Abbott, S., Halaki, M., Torres Castiglioni, M., Yamauchi, G., Mitchell, L., Salter, J., Romann, M. and Cobley, S. (2022) Maturation-based Corrective Adjustment Procedures (Mat-CAPs) in youth swimming: Evidence for restricted age-group application in females. Plos One **17(10)**, e0275797. https://doi.org/10.1371/journal.pone.0275797

Höner, O., Leyhr, D. and Kelava, A. (2017) The influence of speed abilities and technical skills in early adolescence on adult success in soccer: a long-term prospective analysis using ANOVA and SEM approaches. Plos One **12(8)**, 1-15. https://doi.org/10.1371/journal.pone.0182211

Höner, O., Murr, D., Larkin, P., Schreiner, R. and Leyhr, D. (2021) Nationwide subjective and objective assessments of potential talent predictors in elite youth soccer: An investigation of prognostic validity in a prospective study. *Frontiers in Sports and Active Living* **3**. https://doi.org/10.3389/fspor.2021.638227

Höner, O. and Roth, K. (2002) Klassische Testtheorie: Die Gütekriterien sportwissenschaftlicher Erhebungsmethoden [Classical test theory: the quality criteria of sports science data collection methods]. In: Sozialwissenschaftliche Forschungsmethoden in der Sportwissenschaft. Ed: Singer, R. and Willimczik, K. Hamburg: Czwalina. 67-97.

Höner, O. and Votteler, A. (2016) Prognostic relevance of motor talent predictors in early adolescence: a group- and individual-based evaluation considering different levels of achievement in youth football. *Journal of Sports Sciences* **34(24)**, 2269-2278. https://doi.org/10.1080/02640414.2016.1177658

Höner, O., Votteler, A., Schmid, M., Schultz, F. and Roth, K. (2015) Psychometric properties of the motor diagnostics in the German football talent identification and development programme. *Journal of Sports Sciences* **33(2)**, 145-159. https://doi.org/10.1080/02640414.2014.928416

Hosmer, D.W., Lemeshow, S. and Sturdivant, R.X. (2013) Applied logistic regression. New York: John Wiley. https://doi.org/10.1002/9781118548387

Ivarsson, A., Kilhage-Persson, A., Martindale, R., Priestley, D., Huijgen, B., Ardern, C. and Mccall, A. (2020) Psychological factors and future performance of football players: a systematic review with meta-analysis. *Journal of Science and Medicine in Sport* **23(4)**, 415-420. https://doi.org/10.1016/j.jsams.2019.10.021

Javet, M., Fröhlich, S., Bruhin, B., Frey, W.O., Romann, M. and Spörri, J. (2022) Swiss-Ski Power Test results in youth competitive alpine skiers are associated with biological maturation and skiing performance. *International Journal of Sports Physiology and Performance* **17(6),** 961-968. https://doi.org/10.1123/ijspp.2021-0184

Johnson, A., Farooq, A. and Whiteley, R. (2017) Skeletal maturation status is more strongly associated with academy selection than birth quarter. *Science and Medicine in Football* **1(2)**, 157-163. https://doi.org/10.1080/24733938.2017.1283434

Johnston, K. and Baker, J. (2020) Waste reduction strategies: factors affecting talent wastage and the efficacy of talent selection in sport. *Frontiers in Psychology* **10**, 1-11. https://doi.org/10.3389/fpsyg.2019.02925

Jokuschies, N. and Conzelmann, A. (2016) "Das sieht man doch, dass das ein Talent ist!": Subjektive Talentkriterien von Trainern im Spitzenfußball ["It is obvious that this is a talent!": subjective talent criteria of coaches in top-level football]. *Zeitschrift für Sportpsychologie* **23(2)**, 44-55. https://doi.org/10.1026/1612-5010/a000161

Khamis, H.J. and Roche, A.F. (1994) Predicting adult stature without using skeletal age: the Khamis-Roche method. *Pediatrics* **94(4)**, 504-507.

Larochelambert, Q. de, Difernand, A., Antero-Jacquemin, J., Sedeaud, A., Toussaint, J.-F., Pierre Yves, L. and Coulmy, N. (2022) Relative age effect in French alpine skiing: problem and solution. *Journal of Sports Sciences* **40(10)**, 1137-1148. https://doi.org/10.1080/02640414.2022.2052428

Lath, F., Den Hartigh, R.J.R., Wattie, N. and Schorer, J. (2020) Talent selection. In: Talent identification and development in sport. Ed: Baker, J., Cobley, S. and Schorer, J.: Routledge. 50-65. https://doi.org/10.4324/9781003049111-4

Lath, F., Koopmann, T., Faber, I., Baker, J. and Schorer, J. (2021) Focusing on the coach's eye; towards a working model of coach decision-making in talent selection. *Psychology of Sport & Exercise* **56**, 102011. https://doi.org/10.1016/j.psychsport.2021.102011

Leonardo Filho, L.A. (2016) The Pygmalion and Galatea effects in the coaching process from the perspective of high-performance volleyball athletes. *Sports Coaching Review* **5(2)**, 195-197. https://doi.org/10.1080/21640629.2016.1201355

Leyhr, D., Bergmann, F., Schreiner, R., Mann, D., Dugandzic, D. and Höner, O. (2021) Relative age-related biases in objective and subjective assessments of performance in talented youth soccer players. *Frontiers in Sports and Active Living* **3**, 664231. https://doi.org/10.3389/fspor.2021.664231

Leyhr, D., Kelava, A., Raabe, J. and Höner, O. (2018) Longitudinal motor performance development in early adolescence and its relationship to adult success: an 8-year prospective study of highly talented soccer players. *Plos One* **13(5)**, 1-16. https://doi.org/10.1371/journal.pone.0196324

Leyhr, D., Murr, D., Basten, L., Eichler, K., Hauser, T., Lüdin, D., Romann, M., Sardo, G. and Höner, O. (2020a) Biological maturity status in elite youth soccer players: a comparison of pragmatic diagnostics with magnetic resonance imaging. *Frontiers in Sports and Active Living* **2**, 587861. https://doi.org/10.3389/fspor.2020.587861

Leyhr, D., Raabe, J., Schultz, F., Kelava, A. and Höner, O. (2020b) The adolescent motor performance development of elite female soccer players: A study of prognostic relevance for future success in adulthood using multilevel modelling. *Journal of Sports Sciences* **38(11-12)**, 1342-1351. https://doi.org/10.1080/02640414.2019.1686940

Little, R.J.A. (1988) A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association* **83(404)**, 1198. https://doi.org/10.1080/01621459.1988.10478722

Lüdin, D., Donath, L., Cobley, S., Mann, D. and Romann, M. (2022) Player-labelling as a solution to overcome maturation selection biases in youth football. *Journal of Sports Sciences* **40(14)**, 1641-1647. https://doi.org/10.1080/02640414.2022.2099077

Malina, R.M. (2017) Assessment of biological maturation. In: Oxford textbook of children's sport and exercise medicine. Ed: Armstrong, N. and van Mechelen, W. Oxford: Oxford University Press. 3-9. https://doi.org/10.1093/med/9780198757672.003.0001

Malina, R.M., Bouchard, C. and Bar-Or, O. (2004) Growth, maturation, and physical activity. Champaign: Human Kinetics. https://doi.org/10.5040/9781492596837

Malina, R.M., Coelho-e-Silva, M.J., Figueiredo, A.J., Carling, C. and Beunen, G.P. (2012) Interrelationships among invasive and non-invasive indicators of biological maturation in adolescent male soccer players. *Journal of Sports Sciences* **30(15)**, 1705-1717. https://doi.org/10.1080/02640414.2011.639382

Malina, R.M. and Cumming, S.P. (2004) Maturity-associated variation in functional and sport-specific skill tests: implications for adolescent football players. *Insight* **7(2)**, 37-39.

Malina, R.M., Cumming, S.P., Kontos, A.P., Eisenmann, J.C., Ribeiro, B. and Aroso, J. (2005) Maturity-associated variation in sport-specific skills of youth soccer players aged 13-15 years. *Journal of Sports Sciences* **23(5)**, 515-522. https://doi.org/10.1080/02640410410001729928

Malina, R.M., Cumming, S.P., Rogol, A.D., Coelho-e-Silva, M.J., Figueiredo, A.J., Konarski, J.M. and Kozieł, S.M. (2019) Bio-banding in youth sports: background, concept, and application. *Sports Medicine* **49(11)**, 1671-1685. https://doi.org/10.1007/s40279-019-01166-x

Malina, R.M. and Kozieł, S.M. (2014) Validation of maturity offset in a longitudinal sample of Polish boys. *Journal of Sports Sciences* **32(5)**, 424-437. https://doi.org/10.1080/02640414.2013.828850

Malina, R.M., Rogol, A.D., Cumming, S.P., Coelho-e-Silva, M.J. and Figueiredo, A.J. (2015) Biological maturation of youth athletes: assessment and implications. *British Journal of Sports Medicine* **49(13)**, 852-859. https://doi.org/10.1136/bjsports-2015-094623

Mann, D. (2020) Approaches to help coaches and talent scouts overcome relative age effects. In: Relative age effects in sport. Ed: Dixon, J.C. et al. New York: Routledge. 117-135. https://doi.org/10.4324/9781003030737-10

Maszczyk, A., Go, A. and Roczniok, R. (2014) Application of neural and regression models in sports results prediction. *Procedia - Social and Behavioral Sciences* **117**, 482-487. https://doi.org/10.1016/j.sbspro.2014.02.249

McDermott, G., Burnett, A.F. and Robertson, S.J. (2015) Reliability and validity of the Loughborough Soccer Passing Test in adolescent males: implications for talent identification. *International Journal of Sports Science & Coaching* **10(2-3)**, 515-527. https://doi.org/10.1260/1747-9541.10.2-3.515

Meylan, C., Cronin, J., Oliver, J. and Hughes, M. (2010) Talent identification in soccer: The role of maturity status on physical, physiological and technical characteristics. *International Journal of Sports Science & Coaching* **5(4)**, 571-592. https://doi.org/10.1260/1747-9541.5.4.571

Mirwald, R.L., Baxter-Jones, A.D.G., Bailey, D.A. and Beunen, G.P. (2002) An assessment of maturity from anthropometric measurements. *Medicine and Science in Sports & Exercise* **34(4)**, 689-694. https://doi.org/10.1097/00005768-200204000-00020

Mitchell, S.B., Haase, A.M., Cumming, S.P. and Malina, R.M. (2017) Understanding growth and maturation in the context of ballet: a biocultural approach. *Research in Dance Education* **18(3)**, 291-300. https://doi.org/10.1080/14647893.2017.1387525

Moore, S.A., McKay, H.A., Macdonald, H., Nettlefold, L., Baxter-Jones, A.D.G., Cameron, N. and Brasher, P.M. (2015) Enhancing a somatic maturity prediction model. *Medicine and Science in Sports and Exercise* **47(8)**, 1755-1764. https://doi.org/10.1249/MSS.0000000000000588

Moran, J., Sandercock, G.R., Ramírez-Campillo, R., Meylan, C., Collison, J. and Parry, D.A. (2017) A meta-analysis of maturation-related variation in adolescent boy athletes' adaptations to short-term resistance training. *Journal of Sports Sciences* **35(11)**, 1041-1051. https://doi.org/10.1080/02640414.2016.1209306

Müller, O., Simons, A. and Weinmann, M. (2017) Beyond crowd judgments: Data-driven estimation of market value in association football. *European Journal of Operational Research* **263(2)**, 611-624. https://doi.org/10.1016/j.ejor.2017.05.005

Murr, D., Feichtinger, P., Larkin, P., O'Connor, D. and Höner, O. (2018) Psychological talent predictors in youth soccer: a systematic review of the prognostic relevance of psychomotor, perceptual-cognitive and personality-related factors. Plos One **13(10)**, 1-24. https://doi.org/10.1371/journal.pone.0205337

Myburgh, G.K., Cumming, S.P., Coelho-e-Silva, M.J. and Malina, R.M. (2020) Developmental fitness curves: assessing sprint acceleration relative to age and maturity status in elite junior tennis players. *Annals of Human Biology* **47(4)**, 336-345. https://doi.org/10.1080/03014460.2020.1781250

Nevill, A.M., Negra, Y., Myers, T.D., Duncan, M.J., Chaabene, H. and Granacher, U. (2021) Are early or late maturers likely to be fitter in the general population? *International Journal of Environmental Research and Public Health* **18(2)**, 497-513. https://doi.org/10.3390/ijerph18020497

Overton, W.F. (2014) Relational developmental systems and developmental science: A focus on methodology. In: Handbook of developmental systems theory and methodology. Ed: Molenaar, P.C.M., Lerner, R.M. and Newell, K. New York: Guilford. 19-65.

Owen, C., Till, K., Phibbs, P., Read, D.J., Weakley, J., Atkinson, M., Cross, M., Kemp, S., Sawczuk, T., Stokes, K., Williams, S. and Jones, B. (2022) A multidimensional approach to identifying the physical qualities of male English regional academy rugby union players; considerations of position, chronological age, relative age and maturation. *European Journal of Sport Science* 1-11. https://doi.org/10.1080/17461391.2021.2023658

Parr, J., Winwood, K., Hodson-Tole, E., Deconinck, F.J.A., Parry, L., Hill, J.P., Malina, R.M. and Cumming, S.P. (2020) Predicting the timing of the peak of the pubertal growth spurt in elite male youth soccer players: evaluation of methods. *Annals of Human Biology* **47(4)**, 400-408. https://doi.org/10.1080/03014460.2020.1782989

Peeters, T. (2018) Testing the wisdom of crowds in the field: transfermarkt valuations and international soccer results. *International Journal of Forecasting* **34(1)**, 17-29. https://doi.org/10.1016/j.ijforecast.2017.08.002

Pelletier, L.G., Fortier, M.S., Vallerand, R.J. and Brière, N.M. (2001) Associations among perceived autonomy support, forms of self-regulation, and persistence: a prospective study. *Motivation and Emotion* **25(4)**, 279-306. https://doi.org/10.1023/A:1014805132406

Peña-González, I., García-Calvo, T., Cervelló, E.M. and Moya-Ramón, M. (2021) The coaches' efficacy expectations of youth soccer players with different maturity status and physical performance. *Journal of Human Kinetics* **79**, 289-299. https://doi.org/10.2478/hukin-2021-0083

Pfeiffer, M. and Hohmann, A. (2012) Applications of neural networks in training science. *Human Movement Science* **31(2)**, 344-359. https://doi.org/10.1016/j.humov.2010.11.004

Popper, K. (2000) In search of a better world: Lectures and essays from thirty years. London: Routledge.

Prockl, F. and Frick, B. (2018) Information precision in online communities: player valuations on www.transfermarkt.de. *International Journal of Sport Finance* **13(4)**, 319-335.

Quatman-Yates, C.C., Quatman, C.E., Meszaros, A.J., Paterno, M.V. and Hewett, T.E. (2012) A systematic review of sensorimotor function during adolescence: a developmental stage of increased motor awkwardness? *British Journal of Sports Medicine* **46(9)**, 649-655. https://doi.org/10.1136/bjsm.2010.079616

Reeves, M.J., Enright, K.J., Dowling, J. and Roberts, S.J. (2018a) Stakeholders' understanding and perceptions of bio-banding in junior-elite football training. *Soccer & Society* **19**, 1166-1182. https://doi.org/10.1080/14660970.2018.1432384

Reeves, M.J., McRobert, A.P., Littlewood, M.A. and Roberts, S.J. (2018b) A scoping review of the potential sociological predictors of talent in junior-elite football: 2000-2016. *Soccer & Society* **19(8)**, 1085-1105. https://doi.org/10.1080/14660970.2018.1432386

Roberts, A.H., Greenwood, D., Stanley, M., Humberstone, C., Iredale, F. and Raynor, A. (2019) Coach knowledge in talent identification: a systematic review and meta-synthesis. *Journal of Science and Medicine in Sport* **22(10)**, 1163-1172. https://doi.org/10.1016/j.jsams.2019.05.008

Rogol, A.D., Cumming, S.P. and Malina, R.M. (2018) Biobanding: A new paradigm for youth sports and training. *Pediatrics* **142(5)**, e20180423. https://doi.org/10.1542/peds.2018-0423

Romann, M. and Cobley, S. (2015) Relative age effects in athletic sprinting and corrective adjustments as a solution for their removal. *Plos One* **10(4)**, 1-12. https://doi.org/10.1371/journal.pone.0122988

Romann, M., Javet, M. and Fuchslocher, J. (2017) Coaches' eye as a valid method to assess biological maturation in youth elite soccer. *Talent Development & Excellence* **9(1)**, 3-13.

Romann, M., Rössler, R., Javet, M. and Faude, O. (2018) Relative age effects in Swiss talent development-a nationwide analysis of all sports. *Journal of Sports Sciences* **36(17)**, 2025-2031. https://doi.org/10.1080/02640414.2018.1432964

Rosenthal, R. and Jacobson, L. (1968) Pygmalion in the classroom. *The Urban Review* **3(1)**, 16-20. https://doi.org/10.1007/BF02322211

Sarmento, H., Anguera, M.T., Pereira, A. and Araújo, D. (2018) Talent identification and development in male football: a systematic review. *Sports Medicine* **48(4)**, 907-931. https://doi.org/10.1007/s40279-017-0851-7

Sarrazin, P., Vallerand, R., Guillet, E., Pelletier, L. and Cury, F. (2002) Motivation and dropout in female handballers: a 21-month prospective study. *European Journal of Social Psychology* **32(3)**, 395-418. https://doi.org/10.1002/ejsp.98

Schmidt, M., Blum, M., Valkanover, S. and Conzelmann, A. (2015) Motor ability and self-esteem: the mediating role of physical self-concept and perceived social acceptance. *Psychology of Sport & Exercise* **17**, 15-23. https://doi.org/10.1016/j.psychsport.2014.11.006

Schweizer Fussballverband, (2016) Testmanual des Schweizerischen Fussballverbandes [Test manual of the Swiss Football Association]. Available from URL: https://editor.football.ch/portal-data/27/Resources/dokumente/nachwuchsfoerderung/de/Dokumentationen/SFV-Testmanual_Talentfoerderung.pdf [Accessed 2 August 2022].

Schweizer Fussballverband, (2014) Das Nachwuchsförderungskonzept des Schweizerischen Fussballverbandes [The youth development concept of the Swiss Football Association]. Available from URL: https://football.ch/portaldata/27/Resources/dokumente/nachwuchsfoerderung/de/Dokumentationen/SFV-Nachwuchsfoerderungskonzept.pdf [Accessed 2 August 2022].

Sherar, L.B., Mirwald, R.L., Baxter-Jones, A.D.G. and Thomis, M. (2005) Prediction of adult height using maturity-based cumulative height velocity curves. *Journal of Pediatrics* **147(4)**, 508-514. https://doi.org/10.1016/j.jpeds.2005.04.041

Sieghartsleitner, R., Zuber, C., Zibung, M., Charbonnet, B. and Conzelmann, A. (2019a) Talent selection in youth football: technical skills rather than general motor performance predict future player status of football talents. *Current Issues in Sport Science* **4**. https://doi.org/10.15203/CISS_2019.011

Sieghartsleitner, R., Zuber, C., Zibung, M. and Conzelmann, A. (2019b) Science or coaches' eye?—both! beneficial collaboration of multidimensional measurements and coach assessments for efficient talent selection in elite youth football. *Journal of Sports Science and Medicine* **18(1)**, 32-43. https://pubmed.ncbi.nlm.nih.gov/30787649/

Siener, M., Faber, I. and Hohmann, A. (2021) Prognostic validity of statistical prediction methods used for talent identification in youth tennis players based on motor abilities. *Applied Sciences* **11(15)**, 7051. https://doi.org/10.3390/app11157051

Skorski, S., Faude, O., Hammes, D. and Meyer, T. (2016) The relative age effect in German elite youth soccer: implications for a successful career. *International Journal of Sports Physiology and Performance* **11(3)**, 370-376. https://doi.org/10.1123/ijspp.2015-0071

Slomp, D.H. and Fuite, J. (2004) Following Phaedrus: alternate choices in surmounting the reliability/validity dilemma. *Assessing Writing* **9(3)**, 190-207. https://doi.org/10.1016/j.asw.2004.10.001

Smith, K. and Weir, P. (2020) Late birthday benefits: The "underdog hypothesis". In: Relative age effects in sport. Ed: Dixon, J.C. et al. New York: Routledge. 71-82. https://doi.org/10.4324/9781003030737-7

Sober, E. (1981) The principle of parsimony. *British Journal for the Philosophy of Science* **32(2)**, 145-156. https://doi.org/10.1093/bjps/32.2.145

Souza-Lima, J. de, Zamora, J.L., Yáñez-Sepúlveda, R., Matsudo, V.K.R. and Mahecha-Matsudo, S. (2020) Detecting sporting talents with z-strategy: cross sectional study. *Revista Brasileira de Medicina do Esporte* **26(2)**, 147-152. https://doi.org/10.1590/1517-869220202602195735

Steyerberg, E.W. (2019) Clinical prediction models: a practical approach to development, validation, and updating. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-16399-0

Steyerberg, E.W., Vickers, A.J., Cook, N.R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M.J. and Kattan, M.W. (2010) Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* **21(1)**, 128-138. https://doi.org/10.1097/EDE.0b013e3181c30fb2

Teunissen, J.W., Rommers, N., Pion, J., Cumming, S.P., Rössler, R., D'Hondt, E., Lenoir, M., Savelsbergh, G.J. and Malina, R.M. (2020) Accuracy of maturity prediction equations in individual elite male football players. *Annals of Human Biology* **47(4)**, 409-416. https://doi.org/10.1080/03014460.2020.1783360

Till, K., Collins, N., McCormack, S., Owen, C., Weaving, D. and Jones, B. (2022) Challenges and solutions for physical testing in sport: the ProPQ (profiling physical qualities) tool. *Strength & Conditioning Journal Publish Ahead of Print*. https://doi.org/10.1519/SSC.0000000000000710

Till, K., Jones, B.L., Cobley, S., Morley, D., Hara, J.O., Chapman, C., Cooke, C. and Beggs, C.B. (2016) Identifying talent in youth sport: a novel methodology using higher-dimensional analysis. Plos One 11(5), 1-18. https://doi.org/10.1371/journal.pone.0155047

Till, K., Morris, R., Emmonds, S., Jones, B. and Cobley, S. (2018) Enhancing the evaluation and interpretation of fitness testing data within youth athletes. *Strength & Conditioning Journal* **40(5)**, 24-33. https://doi.org/10.1519/SSC.0000000000000414

Towlson, C., Macmaster, C., Gonçalves, B., Sampaio, J., Toner, J., Macfarlane, N., Barrett, S., Hamilton, A., Jack, R., Hunter, F., Myers, T. and Abt, G. (2021a) The effect of bio-banding on physical and psychological indicators of talent identification in academy soccer players. *Science & Medicine in Football* **5(4)**, 280-292. https://doi.org/10.1080/24733938.2020.1862419

Towlson, C., Macmaster, C., Parr, J. and Cumming, S.P. (2021b) One of these things is not like the other: time to differentiate between relative age and biological maturity selection biases in soccer? *Science and Medicine in Football* **6(3)**, 273-276. https://doi.org/10.1080/24733938.2021.1946133

Turner, A. (2014) Total score of athleticism: a strategy for assessing an athlete's athleticism. *Professional Strength and Conditioning* **33**, 13-17.

Vaeyens, R., Malina, R.M., Janssens, M., van Renterghem, B., Bourgois, J., Vrijens, J. and Philippaerts, R.M. (2006) A multidisciplinary selection model for youth soccer: the ghent youth soccer project. *British Journal of Sports Medicine* **40(11)**, 928-934. https://doi.org/10.1136/bjsm.2006.029652

Valente dos Santos, J., Coelho-e-Silva, M.J., Duarte, J., Pereira, J., Rebelo-Gonçalves, R., Figueiredo, A.J., Mazzuco, M.A., Sherar, L.B., Elferink-Gemser, M.T. and Malina, R.M. (2014a) Allometric multilevel modelling of agility and dribbling speed by skeletal age and playing position in youth soccer players. *International Journal of Sports Medicine* **35(9)**, 762-771. https://doi.org/10.1055/s-0033-1358469

Valente dos Santos, J., Coelho-e-Silva, M.J., Martins, R.A., Figueiredo, A.J., Cyrino, E.S., Sherar, L.B., Vaeyens, R., Huijgen, B.C., Elferink-Gemser, M.T. and Malina, R.M. (2012) Modelling developmental changes in repeated- sprint ability by chronological and skeletal ages in young soccer players. *International Journal of Sports Medicine* **33(10)**, 773-780. https://doi.org/10.1055/s-0032-1308996

Valente dos Santos, J., Coelho-e-Silva, M.J., Vaz, V., Figueiredo, A.J., Capranica, L., Sherar, L.B., Elferink-Gemser, M.T. and Malina, R.M. (2014b) Maturity-associated variation in change of direction and dribbling speed in early pubertal years and 5-year developmental changes in young soccer players. *Journal of Sports Medicine and Physical Fitness* **54(3)**, 307-316.

Vallerand, R.J. and Losier, G.F. (1999) An integrative analysis of intrinsic and extrinsic motivation in sport. *Journal of Applied Sport Psychology* **11(1)**, 142-169. https://doi.org/10.1080/10413209908402956

van Hooren, B. and Ste Croix, M. de (2020) Sensitive periods to train general motor abilities in children and adolescents: do they exist? A critical appraisal. *Strength & Conditioning Journal* **42(6)**, 7-14. https://doi.org/10.1519/SSC.0000000000000545

Vist Hagen, R., Haga, M., Sigmundsson, H. and Lorås, H. (2022) The association between academic achievement in physical education and timing of biological maturity in adolescents. *Plos One* **17(3)**, e0265718. https://doi.org/10.1371/journal.pone.0265718

Votteler, A. and Höner, O. (2014) The relative age effect in the german football TID programme: biases in motor performance diagnostics and effects on single motor abilities and skills in groups of selected players. *European Journal of Sport Science* **14(5)**, 433-442. https://doi.org/10.1080/17461391.2013.837510

Williams, A.M., Ford, P.R. and Drust, B. (2020) Talent identification and development in soccer since the millennium. *Journal of Sports Sciences* **38(11-12)**, 1199-1210. https://doi.org/10.1080/02640414.2020.1766647

Williams, A.M. and Reilly, T. (2000) Talent identification and development in soccer. *Journal of Sports Sciences* **18(9)**, 657-667. https://doi.org/10.1080/02640410050120041

Zeileis, A. and Hothorn, T. (2002) Diagnostic checking in regression relationships. *R News* **2(3)**, 7-10.

Zuber, C., Zibung, M. and Conzelmann, A. (2016) Holistic patterns as an instrument for predicting the performance of promising young soccer players─A 3-years longitudinal study. *Frontiers in Psychology* **7(1088)**, 1-10. https://doi.org/10.3389/fpsyg.2016.01088

Key points

- In summative approaches to identify talent, correction mechanisms are needed, and they can be successfully implemented. In this study, however, they could not improve predictions of future performance level (compared with raw scores).
- We do not interpret raw and corrected score equivalent predictions as a sign of correction mechanism futility (more work for the same output), instead we see them as an invitation to take corrected scores seriously into account (same output, one fewer problem).
- Expectations related to corrected scores must be revised according to four factors: initial predictive validity; initial maturity bias of considered variables; validity of the maturity indicator; and current selection system.
- The added value of corrected scores for talent identification and development, such as personality development, environmental support, performance profiling or monitoring, currently resides on rather theoretical grounds.

AUTHOR BIOGRAPHY

Bryan CHARBONNET
Employment
PhD Student at the Institute of Sport Science, University of Bern
Degree
MSc
Research interests
Talent Selection, Talent Development, Developmental Sciences
E-mail: bryan.charbonnet@unibe.ch

Roland SIEGHARTSLEITNER
Employment
Coach at the Upper Austrian Football Association
Degree
PhD
Research interests
Talent Selection and Talent Development in Football
E-mail: sieghartsleitner@ooefv.at

Jürg SCHMID
Employment
Lecturer at the Institute of Sport Science, University of Bern
Degree
PhD
Research interests
Research Methods, Test Development
E-mail: juerg.schmid@unibe.ch

	**Claudia ZUBER** **Employment** Post-Doc Researcher at the Institute of Sport Science, University of Bern **Degree** PhD **Research interests** Psychological Characteristics in Talent Selection and Talent Development **E-mail:** claudia.zuber@unibe.ch
	**Marc ZIBUNG** **Employment** Lecturer at the Institute of Sport Science, University of Bern **Degree** PhD **Research interests** Person-oriented Research in Talent Selection and Talent Development **E-mail:** marc.zibung@unibe.ch
	**Achim CONZELMANN** **Employment** Full Professor at the Institute of Sport Science at the University of Bern **Degree** PhD **Research interests** Developmental Sciences, Personality and Talent Research **E-mail:** achim.conzelmann@unibe.ch

✉ **Bryan Charbonnet**
Institute of Sport Science, University of Bern, Bremgartenstrasse 145, CH-3012 Bern, Switzerland