Talent selection in youth football: Specific rather than general motor performance predicts future player status of football talents

Roland Sieghartsleitner[1,*], Claudia Zuber[1], Marc Zibung[1], Bryan Charbonnet[1], & Achim Conzelmann[1]

1 Institute of Sport Science, University of Bern, Bern, Switzerland

* Corresponding author: Institute of Sport Science, University of Bern, Bremgartenstrasse 145, CH-3012 Bern, Switzerland
Tel: +41 (0) 31 631 45 56
E-Mail: roland.sieghartsleitner@ispw.unibe.ch

ORIGINAL ARTICLE

Article History:
Submitted 2nd April 2019
Accepted 16th September 2019
Published 12th November 2019

Handling Editor:
Martin Kopp
University of Innsbruck, Austria

Editor-in-Chief:
Martin Kopp
University of Innsbruck, Austria

Reviewers:
Reviewer 1: Anonymous
Reviewer 2: Anonymous

ABSTRACT

Recommended multidimensional models for talent selection are difficult to implement for practitioners in the field. Furthermore, their application has not been established from a scientific point of view, with a lack of clarity concerning how to integrate manifold test results with respect to loading, interaction, and compensation phenomena. Consequently, the question of powerful single predictors for future player status are still of interest within talent research in order to determine promising content for less extensive selection procedures. The aim of the current study is a comparison of the prognostic validity of two frequently used areas within talent selection in youth football: physiologically driven general motor performance (GMP) capacities (40m sprint, agility, counter movement jump, Yo-Yo intermittent recovery test) and domain-specific motor performance (SMP) capacities (i.e., technical skills; dribbling, passing, juggling, shooting). The area under the curve (AUC) from the receiver operating characteristic was used to compare the prognostic validity of both motor performance areas at early and middle adolescence (predicting U20 player status: 17 professional vs. 116 non-professional players at U13/U14; 23 vs. 62 at U16/U17). Although no comparison at the four different age levels led to a significant difference ($.07 \leq p \leq .65$), there was a continuous superiority of SMP over GMP in descriptive AUC values ($.04 \leq \Delta AUC \leq .14$). These descriptive differences reached relevant extent within early adolescence ($\Delta AUC_{U13} = .09$; $\Delta AUC_{U14} = .14$) and were partially accounted for by the influence of biological maturation ($.31 \leq r \leq .50$ between maturation and performance in 40m and counter movement jump). In line with theoretical considerations and earlier research, these results provide further evidence of the superiority of SMP over GMP in predicting future player status. Until the applicability of multidimensional models is further established, SMP rather than GMP should be included in less extensive talent selection models, especially in early adolescence.

Keywords:
soccer – talent identification – physical fitness – technical skills – biological maturation

Citation:
Sieghartsleitner, R., Zuber, C., Zibung, M., Charbonnet, B., & Conzelmann, A. (2019). Talent selection in youth football: Specific rather than general motor performance predicts future player status of football talents. Current Issues in Sport Science, 4:011. doi: 10.15203/CISS_2019.011.

Introduction

Although early talent selection may result in the loss of certain potential talent, the tremendous popularity of youth football and limited resources within clubs and associations turn this unwanted procedure into a necessity. To minimize the risk of false selection decisions, research has advocated the use of multidimensional approaches which should maximize prognostic validity in terms of predicting the future player status of young talents with data from childhood or adolescence (Sieghartsleitner, Zuber, Zibung, & Conzelmann, 2019; Vaeyens, Lenoir, Williams, & Philippaerts, 2008). In their fundamental work on these approaches, Williams and Reilly (2000) suggested more than 25 potential predictors of talent in football and grouped this "shopping list of key criteria" (p. 658) into sociological, physical, physiological, and psychological dimensions. This pioneering work had significant impact and several researchers have subsequently executed multidimensional work on talent in football (Figueiredo, Gonçalves, Coelho-e-Silva, & Malina, 2009a; Forsman, Blomqvist, Davids, Liukkonen, & Konttinen, 2016; Huijgen, Elferink-Gemser, Lemmink, & Visscher, 2014; Vaeyens et al., 2006; Zuber, Zibung, & Conzelmann, 2016). However, there has been no scientific study or practical implementation that has integrated all of the suggested predictors from the work of Williams and Reilly (2000). Therefore, whilst conducting multidimensional measurements sounds promising, extensive data collection for talent selection may have certain limitations.

Beyond the obvious economic aspects, statistical considerations of common linear methods versus non-linear alternatives turn the inclusion of several variables within multidimensional modelling into a meaningful problem. In both statistical approaches, extensive models with high numbers of variables increase the probability of obtaining results that are difficult to interpret. For example, multicollinearity leads to unclear explanations of variance, whereby the loading and weighing of single variables also becomes unclear (Backhaus, Erichson, Plinke, & Weiber, 2018). Furthermore, a specific problem of linear statistical models is that they are enslaved by the general relations of "the higher (or lower) x, the higher (or lower) y" (Maszczyk et al., 2014). They may therefore fail to represent possible interaction and compensation phenomena between different talent predictors within developing talents (Conzelmann, Zibung, & Zuber, 2018; Meylan, Cronin, Oliver, & Hughes, 2010). Non-linear alternatives such as artificial neural networks and person-oriented approaches also face certain problems, particularly regarding impossible comparisons of different statistical model configurations, or difficulties in interpreting the obtained results (Pfeiffer & Hohmann, 2012; Pion, Hohmann, Liu, Lenoir, & Segers, 2017; Zibung, Zuber, & Conzelmann, 2016; Zuber et al., 2016). Indeed, since an artificial neural network is a kind of black box, the process behind the emergence of its results is hidden, and imposes a questionable blind explanation of an effect without prior insight into the processes that cause this effect (Zhang et al., 2018). On the other hand, person-oriented methods only manage to overcome interpretative difficulties by compromising part of their holistic aspiration: To be able to deliver interpretable clusters, they must be restricted to a relatively small number of variables, and therefore to less extensive models; i.e. four to six so-called *operating factors* (e.g., the Linking of Clusters after removal of a Residue (LICUR) method; Bergman, Magnusson, & El-Khouri, 2003; Bogat, von Eye, & Bergman, 2016). Overall, therefore, no satisfying solution is currently available for integrating extensive data collection into practical talent selection decisions on single players (such as whether to include a player in a talent development program).

As economic and methodological reasons hinder the implementation of multidimensional talent selection models to a certain degree, there is a need for less extensive solutions; i.e., models with a smaller number of integrated variables. In this context, the search for the most powerful predictors of later performance grows in importance. If current talent selection models can only handle a limited number of variables, this raises the question of which areas provide the most powerful way of discriminating between future performance levels of elite youth players, and which variables are worthy of inclusion in these less extensive models (e.g., a person-oriented model with four to six operating factors).

In general, motor performance has been one of the most considered predictors within talent research in youth football and is also frequently used by practitioners in the field (Höner, Leyhr, & Kelava, 2017; Sarmento, Anguera, Pereira, & Araújo, 2018). Whilst the overall value of motor performance for talent selection is not doubted, it is unclear whether physiologically driven general motor performance (GMP) capacities (e.g., speed, endurance, vertical jump) or domain-specific motor performance (SMP) capacities (i.e., technical skills) best predict future performance levels of young football players (Dodd & Newans, 2018; Forsman et al., 2016; Gonaus & Müller, 2012; Höner et al., 2017; Murr, Raabe, & Höner, 2017). In particular, the prognostic validity of GMP for long-term predictions from childhood or early adolescence has been vigorously questioned due to lower specificity of the task and development-related influences such as biological maturation and relative age, which may at least influence strength and speed abilities (Lidor, Côté, & Hackfort, 2009; Malina, Cumming, Coelho-e-Silva, & Figueiredo, 2017; Müller, Gonaus, Perner, Müller, & Raschner, 2017; Romann, Rössler, Javet, & Faude, 2018). For that reason, the SMP, which is more likely to be maturity-unbiased, is thought to provide higher prognostic validity than GMP, although the measurement reliability of the former is generally lower (Höner, Votteler, Schmid, Schultz, & Roth, 2015; Lidor et al., 2009; Vaeyens et al., 2008).

On the question of the prognostic validity of GMP versus SMP, the amount of relevant research is limited. As Murr, Feichtinger, Larkin, O'Connor, and Höner (2018) showed in their review, only the studies of two working groups can report on long-term prognostic validities (i.e., more than four years) of both GMP and SMP. Forsman et al. (2016) presented data from Finnish U14 players, which were used to predict player status at U19 level (elite vs. sub-elite, n = 114). SMP (dribbling and passing,

passing and centering; d_{mean} = 0.73) and GMP (30m sprint, agility, vertical jump, and endurance; d_{mean} = 0.67) separated groups with similar strong effect sizes on average. In reports on the German football talent identification and talent development program, GMP is reported by means of 20m sprint and agility, SMP with tests for dribbling, ball control, and shooting (Höner et al., 2015). Within U12 data, dribbling and ball control show highest effect sizes (η^2 = .02, n = 22,843) for discrimination between the performance levels of players at the middle and late adolescent stages (U16-U19; Höner & Votteler, 2016). A second study, also within U12 data, reports a higher prognostic validity for SMP (η^2_{mean} = .010) compared to speed abilities (η^2_{mean} = .006) in predicting adult performance levels (n = 14,187; Höner et al., 2017). In a third study, the working group was able to show that the 20m sprint was the only variable that did not significantly separate adult performance levels within U12 to U15 data (n = 1,134; Leyhr, Kelava, Raabe, & Höner, 2018).

In summary, research on the predictive values of GMP versus SMP for talent selection in youth football has produced some empirical results and certain theoretical considerations. However, an immediate pairwise comparison of the prognostic validity of these two areas of motor performance within longitudinal study designs is still missing. As methodological reasons hinder the use of multidimensional approaches with many variables, this comparative knowledge on the prognostic relevance of certain variables at different age groups seems to be necessary to facilitate the choice of predictors for less extensive talent selection models. For that reason, the aim of the current investigation was to research whether GMP or SMP showed higher prognostic validity within talent selection in different stages of youth football (early and middle adolescence).

Methods

Participants

The current research is part of the longitudinal project *Talent Selection and Talent Development in Swiss football*, which followed players born in 1999 throughout the talent promoting system of the Swiss Football Association by using various tests including notably measurements of the motor performance area (Sieghartsleitner et al., 2019, 2018; Zibung et al., 2016; Zuber et al., 2016). The current contribution used a total sample of 195 male players. During the season 2018/2019, 31 of these players (15.9%) participated in the first to third league within Switzerland or were nominated for the Swiss U20 junior national team (i.e., 15.9% were classified as professional players). The remaining 164 players took part in the fourth league or below and were classified as non-professionals. The total sample consists of two groups. A first group of 133 players (17 professionals, 12.8%) volunteered to participate in motor performance tests at early adolescence (U13/U14 age categories). A second group of 85 players (23 professionals, 27.1%) went through the same tests at middle adolescence (U16/U17 age categories). The two groups include 23 players (9 professionals) who participated in early and middle adolescence; hence in all four age groups. As the selection level increases through the ongoing talent promotion system, the U16/U17 group includes a higher percentage of professional players than the U13/U14 group (χ^2 = 7.06, p < .05). The study received approval from the Ethics Committee of the Faculty of Human Sciences of the University of Bern and all players and their legal representatives provided their written informed consent to participate.

Measures

During a single season, players participated twice (autumn and spring) in a test battery consisting of eight variables to determine motor performance. The season's performance was calculated using the mean value of both tests. If one of the two measurements was missed (e.g., through injury, sickness, or school activities), the other served as the test score for the age category (19.4% of cases). As there was only one difference (shooting at U14 age group; p = .03) between players with either one or two participations within 32 comparisons (eight tests at four age groups; t-Test; false discovery rate adjusted alpha level of significance from Benjamini and Hochberg (1995): α = .044) and similar procedures are common within long-term development analysis in football, this procedure was considered to be appropriate (Gonaus & Müller, 2012; Höner et al., 2015).

GMP was operationalized by the following four tests. Firstly, a 40m sprint was conducted with a twin photoelectric sensor (Microgate, Bolzano, Italy) at the starting and finishing line (r_{tt} = .96; Zuber et al., 2016). Secondly, for an agility test, players took a short sprint, ran around three poles with a change of direction, and repeated these actions mirror-inverted before finishing (Höner et al., 2015). As in the sprint test, twin photoelectric sensors measured times (r_{tt} = .83). Thirdly, in a vertical counter movement jump test (without arm swing), the highest value of five attempts served as the test score (Myotest, Sion, Switzerland; ICC = .96; Casartelli, Muller, & Maffiuletti, 2010). Finally, the Level 1 Yo-Yo intermittent recovery test measured the intermittent endurance performance (r_{tt} = .93; Bangsbo, Iaia, & Krustrup, 2008).

SMP was measured by means of an additional four tests. A dribbling test was executed with the same trajectory as the agility test, the only difference being that it was performed with a ball instead of without a ball (r_{tt} = .56; Höner et al., 2015). Secondly, a passing test was adapted from the passing test used by Höner et al. (2015). In this test, players passed the ball from a confined zone against four walls in turn, one in each direction. After the fourth pass, the same sequence was repeated in reverse order (reaching nine passes). Time served as the test score and was measured manually with stopwatches (r_{tt} = .68; Zuber et al., 2016). Thirdly, a juggling test required players to juggle along a course shaped like the figure 8 (left and right foot alternately). Players scored a point for each quarter of a circle they completed. The test was stopped after 45 seconds or, alternatively, as soon as a mistake was made (e.g., 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 (Höner et al., 2015), reaching a $r_{tt} = .79$ (Zuber et al., 2016). Finally, in a shooting test, players had to shoot eight times into target zones of the goal (2 targets, 2 feet, 2 attempts). Successful shots on the target were subjectively rated by speed on a three stage scale (low, medium, or high speed denote 1, 2, or 3 points), and the test score was the overall number of points ($r_{tt} = .31$; Höner et al., 2015).
The protocol for the test battery was standardized (warm-up, order of tests, trained team of testers) and it was executed on dry synthetic turf only. For the 40m sprint, agility, dribbling, passing, and juggling tests, the better of two attempts was used for data analysis. For the all-out Yo-Yo intermittent recovery test, only one attempt was possible.
Finally, an adult state prediction was assessed along with the eight motor performance tests to obtain an indicator of biological maturation by means of the percentage of predicted adult height (Sherar, Mirwald, Baxter-Jones, & Thomis, 2005).

Data Analysis

Due to missed, incorrect, or aborted attempts, 1.0% of all values were missing in the U13/U14 dataset (6.8% cases showed missing data; Little's missing completely at random test: $\chi^2 = 109.0$, $df = 88$, $p = .06$). In the U16/U17 dataset, 3.8% of the values were missing (20.0% cases showed missing data; Little's missing completely at random test: $\chi^2 = 131.0$, $df = 96$, $p = .01$). As missing values can lead to unwanted distortions in statistical analysis (e.g., biased parameter estimates and reduced sample size) and Little's test showed that current data points were missing at random rather than missing completely at random, multiple imputation with $m = 10$ imputations and a maximum of $k = 10$ iterations was carried out by means of the R package *mice* to impute missing values (Jekauc, Völkle, Lämmle, & Woll, 2012; Little, 1988; Stuart, Azur, Frangakis, & Leaf, 2009; van Buuren & Groothuis-Oudshoorn, 2011). All variables in the dataset were defined as predictors as well as imputation variables. After creating the complete datasets, all of the following data analysis procedures were conducted for each of the imputed datasets. Finally, the results of point estimates (mean of the estimates from completed datasets) and interval estimates (considering the within- and between-imputation variance of the completed datasets) were pooled with reference to Rubins' Rule (Jekauc et al., 2012).
The first step in data analysis calculated two classification models per age group to predict U20 player status (professional or non-professional): one for GMP (40m sprint, agility, counter movement jump, Yo-Yo intermittent recovery test) and one for SMP (dribbling, passing, juggling, shooting).
To calculate the likelihood of each individual being categorized as a professional or non-professional player, each of the models used robust classification from binary logistic regression (BLR) in R (Antonogeorgos, Panagiotakos, Priftis, & Tzonou, 2009; R Core Team, 2017). The subsequent receiver operating characteristic (ROC) from the R package *pROC* determined the discriminative power of this classification (Robin et al., 2011). To proof BLR models for superiority over a baseline model and for fitting of the data, a likelihood-ratio test (Omnibus tests of model coefficients) and the Hosmer-Lemeshow test were conducted (Hosmer, Lemeshow, & Sturdivant, 2013; Zeileis & Hothorn, 2002). For both tests, the alpha level for significance was set at $\alpha < .05$. According to the corresponding null hypothesis, superiority over a baseline model was indicated by $\alpha < .05$ (Omnibus tests of model coefficients) and an appropriate fit of the data by means of $\alpha > .05$ (Hosmer-Lemeshow test).
As a next step, the likelihood of each individual being categorized as a professional or non-professional player from BLR was used to create the ROC. The resulting area under the curve (AUC; an index for measuring the quality of classification), was used to compare the GMP and SMP models using the DeLong non-parametric test (DeLong, DeLong, & Clarke-Pearson, 1988; Robin et al., 2011). Again, the alpha level for significance was initially set to $\alpha < .05$. Due to the comparison between the classification models at each age group level, the false discovery rate was used to appropriately adjust the alpha level of significance for multiple testing (Benjamini & Hochberg, 1995)
Compared to the use of BLR only, ROC offers beneficial descriptive values of correctly identified talents (sensitivity), correctly identified non-talents (specificity), and an overall percentage of all correct selection decisions (accuracy; Robin et al., 2011). Each of these three descriptive values can be calculated for each single point on the ROC curve to describe the effectiveness of a certain discrimination threshold. According to these discrimination thresholds, an additional benefit of ROC over BLR and its setting of a fixed threshold is the search for the most powerful discrimination threshold, known as the Youden index (Youden, 1950). The Youden index describes that point of the ROC curve where the sum of sensitivity and specificity is maximized, and therefore may represent the most efficient talent selection threshold for inclusion in a talent development system.
Finally, to examine potential maturational influences on motor performance, the relationship between each performance test and the biological maturation indicator was estimated by using Pearson correlations. Again, the alpha level for significance was set to $\alpha < .05$ and the false discovery rate was used to adjust for multiple testing (Benjamini & Hochberg, 1995).

Results

Tables 1 and 2 provide an overview of the descriptive characteristics of the measured variables for professional and non-professional players. According to the results of the BLR analysis (see Table 3), only three models were significant (Omnibus tests of model coefficients: $p < .05$) and also appropriately calibrated (Hosmer-Lemeshow test: $p > .05$). These were the SMP models at the U13 ($p < .01$, $p = .60$, Nagelkerkes $R^2 = .22$) ; the U14 ($p < .01$, $p = .65$, Nagelkerkes $R^2 = .23$); and the U16 age groups ($p = .03$, $p = .42$, Nagelkerkes $R^2 = .17$). Whilst the SMP model at

Table 1: Means (±standard deviation) for professional players (PP) and non-professional players (NPP) for measured items in U13 / U14 age groups.

Item	U13 PP (*n* = 17)	U13 NPP (*n* = 116)	U13 Total (*n* = 133)	U14 PP (*n* = 17)	U14 NPP (*n* = 116)	U14 Total (*n* = 133)
Age (years)	12.52 ± 0.36	12.57 ± 0.32	12.56 ± 0.32	13.56 ± 0.32	13.56 ± 0.33	13.56 ± 0.33
Height (cm)	152.7 ± 5.0	153.9 ± 8.0	153.8 ± 7.7	160.2 ± 6.5	160.5 ± 8.2	160.5 ± 8.0
Weight (kg)	41.8 ± 5.4	42.8 ± 7.1	42.6 ± 6.9	48.8 ± 6.5	48.0 ± 7.9	48.1 ± 7.7
Maturation (% adult height)	84.95 ± 1.35	85.08 ± 1.99	85.06 ± 1.91	88.94 ± 2.54	88.59 ± 2.77	88.63 ± 2.73
40m sprint (sec)	6.54 ± 0.28	6.61 ± 0.36	6.60 ± 0.35	6.48 ± 0.34	6.41 ± 0.34	6.42 ± 0.34
Agility (sec)	8.20 ± 0.32	8.18 ± 0.32	8.18 ± 0.31	8.15 ± 0.25	8.11 ± 0.29	8.11 ± 0.28
CMJ (cm)	28.8 ± 2.3	30.2 ± 3.7	30.0 ± 3.6	29.3 ± 2.8	31.1 ± 3.9	30.9 ± 3.8
Yo-Yo (m)	905 ± 251	890 ± 284	892 ± 280	1123 ± 408	1100 ± 353	1103 ± 349
Dribbling (sec)	10.21 ± 0.49*	10.76 ± 0.80	10.69 ± 0.78	10.07 ± 0.42	10.33 ± 0.60	10.29 ± 0.58
Passing (sec)	17.21 ± 1.65*	18.39 ± 2.06	18.24 ± 2.06	15.86 ± 1.35*	16.84 ± 1.70	16.72 ± 1.69
Juggling (points)	4.7 ± 3.8*	3.0 ± 3.1	3.2 ± 3.2	11.3 ± 5.4*	5.9 ± 5.0	6.6 ± 5.4
Shooting (points)	9.0 ± 2.7*	7.4 ± 2.8	7.6 ± 2.8	10.1 ± 3.3	8.5 ± 2.6	8.7 ± 2.7

Note: CMJ = counter movement jump; Yo-Yo = Level 1 Yo-Yo intermittent recovery test; * = different from NPP (*t*-Test, $p < .05$)

Table 2: Means (±standard deviation) for professional players (PP) and non-professional players (NPP) for measured items in U16 / U17 age groups.

Item	U16 PP (*n* = 23)	U16 NPP (*n* = 62)	U16 Total (*n* = 85)	U17 PP (*n* = 23)	U17 NPP (*n* = 62)	U17 Total (*n* = 85)
Age (years)	15.45 ± 0.44	15.53 ± 0.40	15.51 ± 0.41	16.51 ± 0.31	16.61 ± 0.33	16.58 ± 0.33
Height (cm)	152.7 ± 5.0	153.9 ± 8.0	153.8 ± 7.7	160.2 ± 6.5	160.5 ± 8.2	160.5 ± 8.0
Weight (kg)	62.4 ± 7.2	64.6 ± 7.1	64.1 ± 7.2	66.9 ± 6.6	68.6 ± 7.1	68.2 ± 7.0
Maturation (% adult height)	96.98 ± 1.39	97.53 ± 1.34	97.38 ± 1.37	98.78 ± 0.74	99.05 ± 0.75	98.98 ± 0.75
40m sprint (sec)	5.67 ± 0.16	5.73 ± 0.20	5.71 ± 0.61	5.58 ± 0.18	5.66 ± 0.20	5.64 ± 0.58
Agility (sec)	7.77 ± 0.22	7.85 ± 0.28	7.83 ± 0.27	7.73 ± 0.27	7.86 ± 0.35	7.82 ± 0.34
CMJ (cm)	37.4 ± 3.7	36.9 ± 4.8	37.0 ± 4.5	38.1 ± 3.3	37.5 ± 4.7	37.7 ± 4.3
Yo-Yo (m)	2226 ± 432	2050 ± 429	2098 ± 434	2344 ± 359*	2111 ± 533	2174 ± 501
Dribbling (sec)	9.71 ± 0.53	9.92 ± 0.63	9.86 ± 0.61	9.65 ± 0.53	9.83 ± 0.59	9.78 ± 0.58
Passing (sec)	14.56 ± 1.16	14.84 ± 1.43	14.77 ± 1.36	13.39 ± 1.30*	13.82 ± 1.29	13.70 ± 1.28
Juggling (points)	14.0 ± 7.1*	9.0 ± 6.2	10.4 ± 6.8	16.6 ± 6.6*	11.5 ± 7.3	12.9 ± 7.5
Shooting (points)	6.9 ± 3.4	6.9 ± 2.8	6.9 ± 2.9	8.3 ± 3.3	7.7 ± 3.2	7.9 ± 3.2

Note: CMJ = counter movement jump; Yo-Yo = Level 1 Yo-Yo intermittent recovery test; * = different from NPP (*t*-Test, $p < .05$)

Table 3: Significance, calibration, and model fit values for the general and specific motor performance models.

	Age group	Omnibus tests of model coefficients			Hosmer-Lemeshow test			Model fit
		χ^2	*df*	*p*	χ^2	*df*	*p*	Nagelkerkes R^2
General motor performance	U13	6.59	4	.16	5.22	8	.73	.09
	U14	3.88	4	.42	10.17	8	.25	.05
	U16	4.37	4	.36	6.94	7	.44	.07
	U17	6.36	4	.17	7.34	7	.50	.11
Specific motor performance	U13	16.37	4	< .01	6.45	8	.60	.22
	U14	17.33	4	< .01	5.95	8	.65	.23
	U16	10.60	4	.03	7.13	7	.42	.17
	U17	8.21	4	.08	4.32	7	.74	.13

Table 4: Descriptive values of the receiver operating characteristic curves for the five classification models.

	Age group	AUC [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]	Accuracy [95% CI]	YI
General motor performance	U13	.68 [.53; .82]	.64 [.36; .93]	.80 [.29; 1.00]	.77 [.37; 1.00]	.44
	U14	.65 [.52; .78]	.76 [.66; 1.00]	.65 [.22; 1.00]	.66 [.32; 1.00]	.42
	U16	.68 [.55; .81]	.78 [.43; 1.00]	.61 [.46; .77]	.67 [.56; .79]	.40
	U17	.67 [.55; .80]	.79 [.56; 1.00]	.64 [.39; .89]	.68 [.53; .83]	.43
Specific motor performance	U13	.77 [.66; .88]	.88 [.61; 1.00]	.66 [.46; .85]	.68 [.52; .85]	.54
	U14	.79 [.67; .91]	.76 [.53; 1.00]	.78 [.54; 1.00]	.78 [.59; .99]	.55
	U16	.74 [.63; .85]	.82 [.51; 1.00]	.66 [.34; .98]	.70 [.51; .88]	.49
	U17	.71 [.59; .83]	.77 [.46; 1.00]	.67 [.35; .99]	.70 [.51; .89]	.44

Note: AUC = area under the curve, YI = Youden Index

the U17 age group was still close (p = .08, p = .74, Nagelkerkes R^2 = .13), all of the GMP models lacked clearer significance and showed lower model fits (.05 ≤ Nagelkerkes R^2 ≤ .11).

Table 4 presents the descriptive values from the ROC. The AUC [95% CI] indicates values from .65 [.52; .78] to .68 [.55; .81] for GMP models and from .71 [.59; .83] to .79 [.67; .91] for SMP models. Sensitivities of the GMP classification models indicate values between .64 [.36; .93] and .79 [.56; 1.00], which means that these models were able to identify a range from 10 of 17 (U13) up to 18 of 23 (U17) professional players correctly. The classification models for SMP identified a range from 12 of 17 (U14) up to 15 of 17 (U13) professional players correctly, which indicates sensitivities from .76 [.53; 1.00] to .88 [.61; 1.00]. Values for specificities ranged from .61 [.46; .77] to .80 [.29; 1.00] within GMP classification models. This means a correct identification ranging from 37 of 62 (U16) up to 93 of 116 (U13) non-professional players. Within SMP models, a range from 41 of 62 (U16) up to 90 of 116 (U14) non-professional players were identified correctly (specificities from .66 [.34; .98] to .78 [.54; 1.00]). Sensitivities and specificities together lead to Youden indices (YIs) from .40 to .44, with accuracies from .66 [.32; 1.00] to .77 [.37; 1.00] in GMP models. In other words, the GMP model with the lowest accuracy (U14) would predict 54 of 133 players becoming professional (13 valid predictions, 4 professional players missed). The GMP model with the highest accuracy (U13) would predict only 34 players (11 valid predictions, 6 professional players missed). SMP models reached YIs from .44 to .55, with accuracies of .68 [.52; .85] to .78 [.59; .99]. Overall selection decisions based on these models would predict 54 of 133 players (U13: 15 valid predictions, 2 professional players missed) in the worst case, while predicting 39 of 133 players (U14: 13 valid predictions, 4 professional players missed) represents the most effective selection decision from SMP models.

Table 5 displays the results of the non-parametric approach to compare the AUCs of GMP and SMP within each single age group. None of these four comparisons led to a significant difference. The highest *z*-value appeared within the U14 age group (AUC_{GMP} = .65 [.52; .78], AUC_{SMP} = .79 [.67; .91], z = 1.79, p = .07), and the lowest in the U17 age group (AUC_{GMP} = .67 [.55; .80], AUC_{SMP} = .71 [.59; .83], z = 0.46, p = .65).

To illustrate the impact of the single variables in the classification models, Table 6 shows the BLR regression coefficients for the SMP model for the age group with the highest model fit (U14). Juggling is the only variable with significant impact in that model, reaching an Odds Ratio (*OR*) [95% CI] of 2.06 [1.19; 3.55] ($p < .01$) for *z*-standardized data, which means that achieving one standard deviation better in juggling doubles the chances of becoming a professional player. The further variables included did not significantly influence the U14 SMP regression model (shooting: p = .20; dribbling: p = .46; and passing: p = .52).

To obtain further insight into the value of single GMP variables, Table 7 presents the BLR regression coefficients for this model at the U17 age group. The non-significant results for the overall model show that no single variable has a significant impact, with the Yo-Yo intermittent recovery test showing the highest *OR* of 1.52 [0.80; 2.89] (p = .20) for *z*-standardized data, whilst the 40m sprint (p = .28), agility (p = .64), and counter movement jump (p = .84) showed less important *OR*s.

Finally, Table 8 presents Pearson correlations between the percentage of predicted adult height and motor performance to examine the influence of biological maturation. The results do

Table 5: Comparison using the DeLong non-parametric test (DeLong et al., 1988) between the AUCs of the general and specific motor performance models for each age group.

Age group	General motor performance	Specific motor performance		
	AUC [95% CI]	**AUC [95% CI]**	***Z***	***p***
U13	.68 [.53; .82]	.77 [.66; .88]	1.38	.17
U14	.65 [.52; .78]	.79 [.67; .91]	1.79	.07
U16	.68 [.55; .81]	.74 [.63; .85]	0.80	.42
U17	.67 [.55; .80]	.71 [.59; .83]	0.46	.65

Note: AUC = area under the curve

Table 6: Coefficients of the U14 specific motor performance binary logistic regression model (for *z*-standardized data).

Item[1]	β	***SE***	**Wald**	***df***	***p***	***Odds Ratio [95% CI]***
Juggling	0.72	0.28	6.75	1	.01	2.06 [1.19; 3.55]
Shooting	0.38	0.30	1.67	1	.20	1.47 [0.82; 2.62]
Dribbling	–0.32	0.43	0.54	1	.46	0.73 [0.31; 1.69]
Passing	–0.29	0.45	0.41	1	.52	0.75 [0.31; 1.82]

Note: [1] Variables ranked by absolute value of beta coefficients

Table 7: Coefficients of the U17 general motor performance binary logistic regression model (for *z*-standardized data).

Item[1]	β	*SE*	Wald	*df*	*p*	*Odds Ratio [95% CI]*
Yo-Yo	0.42	0.33	2.06	1	.20	1.52 [0.80; 2.89]
40m sprint	–0.36	0.33	1.17	1	.28	0.70 [0.37; 1.34]
Agility	–0.15	0.32	0.23	1	.64	0.86 [0.46; 1.62]
CMJ	–0.06	0.31	0.04	1	.84	0.94 [0.52; 1.72]

Note: [1] Variables ranked by absolute value of beta coefficients. CMJ = counter movement jump; Yo-Yo = Level 1 Yo-Yo intermittent recovery test

Table 8: Pearson correlation coefficients between biological maturation (percentage of predicted adult height) and motor performance tests

	Specific motor performance				General motor performance			
Age Group	**Juggling**	**Shooting**	**Dribbling**	**Passing**	**Yo-Yo**	**40m sprint**	**Agility**	**CMJ**
U13	–.01	–.10	–.10	–.12	.05	.31*	–.08	.37*
U14	.01	.15	–.03	.08	.01	.50*	–.11	.33*
U16	–.12	–.12	–.11	.06	–.14	.23*	–.12	.14
U17	–.11	–.05	–.11	–.04	–.08	.07	–.28*	.06

Note: Positive correlations express better test performance with higher percentage of predicted adult height; CMJ = counter movement jump; Yo-Yo = Level 1 Yo-Yo intermittent recovery test; * = $p < .05$ (false discovery rate adjusted α: .045; Benjamini & Hochberg, 1995)

not show any significant correlations between SMP tests and biological maturation over all age groups. On the other hand, the GMP tests of the 40m sprint ($.31 \le r \le .50$) and counter movement jump ($.33 \le r \le .37$) are correlated with biological maturation in early adolescence (U13/U14). These correlations decline or disappear over time until middle adolescence, while a single correlation between biological maturation and agility appears at U17.

Discussion

The findings of the current study show that talent selection models with the single dimensions of GMP or SMP do not lead in general to significant predictions with the aim of an early differentiation between professional and non-professional players. Only three out of eight BLR models showed superiority over random predictions and were also calibrated appropriately (i.e., the SMP models for the U13, U14, and U16 age groups). Compared to the significance of predictions and the higher explained variance from the more extensive, multidimensional selection models within a similar group of participants, the less extensive models in the current study indicate substantially lower prognostic validity (Sieghartsleitner et al., 2019). This may underline the assumed advantages of multidimensionality over less extensive models (Vaeyens et al., 2008; Williams & Reilly, 2000; Zuber et al., 2016). However, as long as these multidimensional models are not easily applicable to talent selection in the field, the immediate comparison between possible predictors of later performance may be of certain relevance within talent research.

Prognostic validity and specificity of the task

Based on the immediate comparisons of different predictors of future performance in youth football, the current research contrasted the value of two different areas of motor performance for early talent selection. In doing so, there was a continuous superiority of SMP ($.71 \le AUC \le .79$) over GMP ($.65 \le AUC \le .68$) in descriptive values within each analysed age group through early (U13/U14) and middle adolescence (U16/U17), despite the higher measurement reliability of the GMP. None of the four comparisons at the different age groups led to a significant difference between AUCs of GMP and SMP ($.07 \le p \le .65$). However, the AUC assessment from Hosmer et al. (2013) underpins the relevance of a decile difference within this parameter (i.e., AUC of .50 = no discrimination; .70 = acceptable discrimination; .80 = excellent discrimination; and 1.00 = perfect discrimination). As this decile difference is the case between GMP and SMP in the U13 (AUC = .68/.77) and U14 (AUC = .65/.79) age groups, these differences may express a relevant but not significant difference within discrimination between professional and non-professional players. For this reason, the current findings of a slightly higher discriminative power for SMP over GMP seems to be in line with earlier research from the German football talent identification and talent development program (Höner et al., 2017; Höner & Votteler, 2016; Leyhr et al., 2018). For the discrimination between performance levels of players at late adolescence or early adulthood, they also found higher effect sizes within SMP compared to GMP in U12 to U15 data. Furthermore, the results underline the theoretical considerations of Lidor et al. (2009), who assumed higher reliability from GMP tests but higher prognostic validity from SMP because of the specificity of the task.

With regard to the value of single SMP variables within the discrimination models, juggling shows by far the highest impact on future performance within the example of the U14 BLR (*OR* = 2.06 [1.19; 3.55]). Shooting (*OR* = 1.47 [0.82; 2.62]), dribbling (*OR* = 0.73 [0.31; 1.69]), and passing (*OR* = 0.75 [0.31; 1.82]) also indicate that better test performances have the tendency to affect the chance of becoming a professional player positively (time scales in dribbling and passing are inverse). Compared to earlier research, this relevance of juggling is unexpected, because juggling tests have not been considered in many studies of the prognostic validity of SMP in youth football, nor have any studies reported on the long-term prognostic validity of juggling (Murr et al., 2018). Compared to shooting, dribbling, and passing, the skill of juggling is not a relevant task within football matches, which might explain why juggling has not received much attention within talent research (Ali, 2011). Apart from its prognostic validity, the higher reliability (r_{tt} = .79) compared to other SMP tests in the current study (.31 ≤ r_{tt} ≤ .68), and the substantial factor loading on the latent dimension *technical skills*, may be hints for the high value of juggling (Höner et al., 2015). Earlier results on its independence from biological maturation are confirmed in the current study, albeit this is also true for the other three SMP tests (Figueiredo, Goncalves, Silva, & Malina, 2009b; Matta, Figueiredo, Garcia, & Seabra, 2014). Following these results, the use of juggling tests within talent selection in youth football is highly recommended.

Regarding GMP, earlier research claimed a significant impact of test results from different stages of adolescence on adult performance levels (Dodd & Newans, 2018; Gonaus & Müller, 2012; Murr et al., 2017). However, the BLR models for GMP in the current study did not lead to any significant solutions, whilst descriptive values even show surprising inverse characteristics in early adolescence (i.e., descriptive statistics indicate better values for non-professional players in certain tests within early adolescence).

Prognostic validity and time span

In addition to the immediate comparison of the prognostic validity of GMP and SMP over all age groups and the value of single variables within these models, the current study enables insight into changes in prognostic validity over different time spans from early or middle adolescence to the U20 age group. In particular, the prognostic validity of the GMP appeared quite stable through all age groups (.65 ≤ AUC ≤ .68), which indicates that the value for predicting the U20 player status (professional vs. non-professional) from GMP data is similar over a period of three years (from middle adolescence: U17 to U20) or seven years (from early adolescence: U13 to U20) respectively. This seems to be unexpected, because biological maturation influences the long-term predictions from GMP in early adolescence by means of the 40m sprint and counter movement jump, and therefore these long-term predictions should suffer from lower prognostic validity compared to predictions over shorter time spans (Malina et al., 2017). Compared to this stability for predictions from GMP, SMP even shows a tendency for inverse patterns considering the time span. In early adolescence (six to seven years before U20), SMP variables differentiated players slightly better (.77 ≤ AUC ≤ .79) than they did in middle adolescence (three to four years before the U20 performance criterion; .71 ≤ AUC ≤ .74). Furthermore, the accuracy and specificity of both motor performance areas show, on average, a rather declining trend from early (U13/U14) to middle adolescence (U16/U17). Accordingly, it seems increasingly challenging to identify correctly who will (not) become a professional, which is inconsistent with the expected pattern (i.e., the shorter the time span, the more accurate the prognosis; Güllich, 2014).

These unexpected patterns of prognostic validity over different time spans may be due to certain aspects of the methodology used in the current research. Firstly, the mixed longitudinal and cross-sectional data, which resulted from recurring selections in the pyramidal standard model of talent development within the system of the Swiss Football Association, could lead to a different selection level between early (U13/U14) and middle adolescence data (U16/U17; Bailey & Collins, 2013). The change in the percentage of professional players within these different groups of participants supports this consideration (13% professional players at U13/14 vs. 27% at U16/U17; χ^2 = 7.06, $p < .05$).

Secondly, as the level of selection increases over time, the heterogeneity among players' potential may decrease (Baker & Wattie, 2018). In combination with the progressive attenuation of the inequalities linked to the difference in biological maturation in the 40m sprint and counter movement jump, this greater homogeneity may redistribute the worth of each predictive area to correctly identify future professionals (sensitivity; Baker, Wattie, & Schorer, 2019). This would explain why, on average, the sensitivities of GMP start to gradually compensate for the corresponding loss of discriminative power attributed to maturity-independent factors (i.e., SMP) as adulthood approaches.

Thirdly, the BLR and ROC statistics compensate for the different directions of descriptive differences in GMP between early and middle adolescence (Backhaus et al., 2018). Whilst descriptive mean values for non-professional players were better in more than 50% of GMP tests during early adolescence, professional players performed better in every test during middle adolescence. Therefore, the prognostic validity of GMP tests seems to be questionable in early adolescence. Overall, the overlapping effects of time spans (the assumed lower prognostic validity for longer time spans), different selection levels (the assumed higher prognostic validity for heterogeneous groups at lower selection levels), decreasing inequalities from biological maturation, as well as further non-linear influences, may have led to the stable AUC values over time within GMP, and to the slightly higher but decreasing AUC values of SMP.

General limitations

Apart from the above limitations, which address the specific topic of prognostic validity and time span, there are also general aspects of the current study that might limit its worth. Firstly, all data analysis was carried out using exploratory dataset models only. Regarding the statistical machine learning practice, this would be only the first step of an analysis, which should be completed by validating the metric model through a second dataset (Till et al., 2016). However, because of the limited number of participants in the current study, they could not be split into exploratory and validation datasets without violating other requirements of the statistical analysis. Therefore, the limited number of participants as well as the small number of professional players which emerged as a result, especially within the early adolescence sample, may hinder a more comprehensive and detailed insight into the value of GMP versus SMP for talent selection in youth football.

Secondly, the data analysis ignored non-significant results from BLR within the initial step of analysis. However, as the results have shown, a non-significant result from BLR within GMP data does not necessarily mean that the discriminative power of such a model is significantly lower than the discriminative power of a model with significant results in BLR within SMP data. Furthermore, as stated in the introduction, talent research aims for multidimensional models to explain the future performance of players (Williams & Reilly, 2000). Therefore, the use of single dimensions only (i.e., GMP or SMP) may have its weaknesses, such that significant results from BLR set rigid criteria for unidimensional models, which would be difficult to meet.

Thirdly, the data analysis is based on the curve linear model of the BLR. As mentioned in the introduction, linear statistical models rely on the higher the x, the higher the y relations and may therefore fail to represent intra-individual interactions that allow weaknesses in one predictor to be compensated for by strengths in another (Conzelmann et al., 2018; Maszczyk et al., 2014; Meylan et al., 2010). Apart from that, non-linear alternatives (e.g. person-oriented methods or artificial neural networks) are unable to deliver immediate comparisons between different model configurations (Bogat et al., 2016). Therefore, the use of the current methodology was a corollary from earlier considerations of the research question.

Conclusion

This study of an immediate comparison of the prognostic validity of GMP (40m sprint, agility, counter movement jump, YoYo intermittent recovery test) versus SMP (dribbling, passing, juggling, shooting) for talent selection in youth football seems to provide certain evidence that the latter is more useful for predicting future player status. This is in line with theoretical considerations and earlier research on the topic (Höner et al., 2017; Höner & Votteler, 2016; Leyhr et al., 2018; Lidor et al., 2009). SMP showed promising results with significant BLR models, especially for long-term predictions from early adolescence (U13/U14), whereas the prognostic validity of GMP over this longer time span of six to seven years seems to be unclear (for instance, descriptive statistics indicate better values for non-professional players in certain tests within early adolescence). This weak prognostic relevance of the GMP is at least partly explained by the influence of biological maturation. According to changes over time, the influence of biological maturation tends to disappear and the prognostic validity of GMP becomes more evident in middle adolescence (i.e., descriptive statistics indicate better values for professional players in each test), though SMP still discriminates players slightly better.

Consequently, until multidimensional models are a) less difficult to implement for practitioners in the field and b) able to process manifold variables from different dimensions for overall selection decisions on single players, then SMP should be included in less extensive talent selection models in early adolescence, as GMP may have more questionable prognostic validity. For selection models in middle adolescence, SMP is still preferable, though its superiority over GMP decreases.

Acknowledgements

We would like to thank the Swiss Football Association for supporting and funding this research project.

Competing Interests

The authors have declared that no competing interests exist.

Data Availability Statement

The datasets analysed during the current study are available from the corresponding author on request.

References

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

Antonogeorgos, G., Panagiotakos, D. B., Priftis, K. N., & Tzonou, A. (2009). Logistic regression and linear discriminant analyses in evaluating factors associated with asthma prevalence among 10- to 12-years-old children: divergence and similarity of the two statistical methods. *International Journal of Pediatrics, 2009*, 1–6. doi.org/10.1155/2009/952042

Bailey, R., & Collins, D. (2013). The standard model of talent development and its discontents. *Kinesiology Review, 2*(4), 248–259. doi.org/10.1123/krj.2.4.248

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2018). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (15th ed.) [Multivariate analysis methods. An application-oriented introduction]. Berlin: Springer.

Baker, J., & Wattie, N. (2018). Innate talent in sport: separating myth from reality. *Current Issues in Sport Science, 3*, 1–9. doi: 10.15203/CISS_2018.006

Baker, J., Wattie, N., & Schorer, J. (2019). A proposed conceptualization of talent in sport: the first step in a long and winding road. *Psychology of Sport and Exercise, 43*, 27–33. doi.org/10.1016/j.psychsport.2018.12.016

Bangsbo, J., Iaia, F. M., & 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. doi.org/10.2165%2F00007256-200838010-00004

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*(1), 289–300.

Bergman, L., Magnusson, D., & El-Khouri, B. M. (2003). *Studying individual development in an interindividual context: a person-oriented approach.* Mahwah, N. J.: Lawrence Erlbaum Associates Publishers.

Bogat, G. A., von Eye, A. & Bergman, L. (2016). Person-oriented approaches. In D. Cicchetti (Eds.), *Developmental Psychopathology, Volume 1: Theory and Method* (3rd ed., pp. 1–49). Hoboken: Jon Wiley & Sons.

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

Conzelmann, A., Zibung, M., & Zuber, C. (2018). Talente finden und fördern im Sport [Talent identification and talent development in sports]. In A. Ritz & N. Thom (Eds.), *Talent management* (pp. 87–106). Wiesbaden: Springer Fachmedien. doi.org/10.1007%2F978-3-658-19100-9_5

DeLong, E., DeLong, D., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics, 44*(3), 837–845. doi.org/10.2307/2531595

Dodd, K. D., & Newans, T. J. (2018). Talent identification for soccer: physiological aspects. *Journal of Science and Medicine in Sport, 21*, 1073-1087. doi.org/10.1016/j.jsams.2018.01.009

Figueiredo, A., Gonçalves, C. E., Coelho-e-Silva, M., & Malina, R. (2009a). Characteristics of youth soccer players who drop out, persist or move up. *Journal of Sports Sciences, 27*(9), 883–891. doi.org/10.1080/02640410902946469

Figueiredo, A., Goncalves, C. E., Coelho-e-Silva, M., & Malina, R. (2009b). Youth soccer players, 11-14 years: maturity, size, function, skill and goal orientation. *Annuals of Human Biology, 36*(1), 60–73. doi.org/10.1080/03014460802570584

Forsman, H., Blomqvist, M., Davids, K., Liukkonen, J., & Konttinen, N. (2016). Identifying technical, physiological, tactical and psychological characteristics that contribute to career progression in soccer. *International Journal of Sports Science & Coaching, 11*(4), 505–513. doi.org/10.1177/1747954116655051

Gonaus, C., & Müller, E. (2012). Using physiological data to predict future career progression in 14- to 17-year-old Austrian soccer academy players. *Journal of Sports Sciences, 30*(15), 1673–1682. doi.org/10.1080/02640414.2012.713980

Güllich, A. (2014) Selection, de-selection and progression in German football talent promotion. *European Journal of Sport Science, 14*(6), 530–537. doi.org/10.1080/17461391.2013.858371

Höner, O., Leyhr, D., & 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*, 1–15. doi.org/10.1371/journal.pone.0182211

Höner, O., & 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. doi.org/10.1080/02640414.2016.1177658

Höner, O., Votteler, A., Schmid, M., Schultz, F., & 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. doi.org/10.1080/02640414.2014.928416

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). New York: John Wiley.

Huijgen, B. C. H., Elferink-Gemser, M. T., Lemmink, K., & Visscher, C. (2014). Multidimensional performance characteristics in selected and deselected talented soccer players. *European Journal of Sport Science, 14*(1), 2–10. doi.org/10.1080/17461391.2012.725102

Jekauc, D., Völkle, M., Lämmle, L., & Woll, A. (2012). Fehlende Werte in sportwissenschaftlichen Untersuchungen [Missing values in sports science studies]. *Sportwissenschaft, 42*(2), 126–136. doi.org/10.1007/s12662-012-0249-5

Leyhr, D., Kelava, A., Raabe, J., & 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*, 1–16. doi.org/10.1371/journal.pone.0196324

Lidor, R., Côté, J., & Hackfort, D. (2009). ISSP position stand: to test or not to test? The use of physical skill tests in talent detection and in early phases of sport development. *International Journal of Sport and Exercise Psychology, 7*, 131–146. doi.org/10.1080/1612197X.2009.9671896

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), 1198–1202. doi.org/10.2307/2290157

Malina, R., Cumming, S. P., Coelho-e-Silva, M., & Figueiredo, A. (2017). Talent identification and development in the context of "growing up". In J. Baker, S. Cobley, J. Schorer, & N. Wattie (Eds.), *Routledge handbook of talent identification and development in Sport* (pp. 150–168). London: Routledge. doi.org/10.4324/9781315668017

Maszczyk, A., Gołaś, A., Pietraszewski, P., Roczniok, R., Zając, A., & Stanula, A. (2014). Application of neural and regression models in sports results prediction. *Procedia - Social and Behavioral Sciences, 117*, 482–487. doi.org/10.1016/j.sbspro.2014.02.249

Matta, M. D. O., Figueiredo, A., Garcia, E. S., & Seabra, A. (2014). Perfil morfológico, maturacional, funcional e técnico de jovens futebolistas Brasileiro [Morphological, maturational, functional and technical profile of young Brazilian footballers]. *Revista Brasileira De Cineantropometria E Desempenho Humano, 16*(3), 277-286. doi.org/10.5007/1980-0037.2014v16n3p277

Meylan, C., Cronin, J., Oliver, J. L., & 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. doi.org/10.1260/1747-9541.5.4.571

Müller, L., Gonaus, C., Perner, C., Müller, E., & Raschner, C. (2017). Maturity status influences the relative age effect in national top level youth alpine ski racing and soccer. *PloS One, 12*, 1–11. doi.org/10.1371/journal.pone.0181810

Murr, D., Feichtinger, P., Larkin, P., O'Connor, D., & 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*, 1–24. doi.org/10.1371/journal.pone.0205337

Murr, D., Raabe, J., & Höner, O. (2017). The prognostic value of physiological and physical characteristics in youth soccer: a systematic review. *European Journal of Sport Science, 18*(1), 1–13. doi.org/10.1080/17461391.2017.1386719

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

Pion, J., Hohmann, A., Liu, T., Lenoir, M., & Segers, V. (2017). Predictive models reduce talent development costs in female gymnastics. *Journal of Sports Sciences, 35*(8), 806–811. doi.org/10.1080/02640414.2016.1192669

R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). Proc: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics, 12,* 1–8. doi.org/10.1186/1471-2105-12-77

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

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

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

Sieghartsleitner, R., Zuber, C., Zibung, M., & Conzelmann, A. (2018). "The early specialised bird catches the worm!" - a specialised sampling model in the development of football talents. *Frontiers in Psychology, 9*, 1–12. doi.org/10.3389/fpsyg.2018.00188

Sieghartsleitner, R., Zuber, C., Zibung, M., & Conzelmann, A. (2019). "Science or coaches' eye? – both!" Beneficial collaboration of multidimensional measurements and coach assessments for efficient talent selection in in elite youth football. *Journal of Sports Science and Medicine, 18*, 32–43.

Stuart, E. A., Azur, M., Frangakis, C., & Leaf, P. (2009). Multiple imputation with large data sets: a case study of the children's mental health Initiative. *American Journal of Epidemiology, 169*(9), 1133–1139. doi.org/10.1093/aje/kwp026

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

Vaeyens, R., Lenoir, M., Williams, M., & Philippaerts, R. M. (2008). Talent identification and development programmes in sport. *Sports Medicine, 38*(9), 703–714. doi.org/10.1080/02640410903110974

Vaeyens, R., Malina, R., Janssens, M., van Renterghem, B., Bourgois, J., Vrijens, J., & 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. doi.org/10.1136/bjsm.2006.029652

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3). doi.org/10.18637/jss.v045.i03

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

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*, 32–35. doi.org/10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3

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

Zhang, Z., Beck, M. W., Winkler, D. A., Huang, B., Sibanda, W., & Goyal, H. (2018). Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. *Annals of Translational Medicine, 6*(11), 216. doi.org/10.21037/atm.2018.05.32

Zibung, M., Zuber, C., & Conzelmann, A. (2016). The motor subsystem as a predictor of success in young football talents: a Person-Oriented Study. *PloS One, 11*, 1–11. doi.org/10.1371/journal.pone.0161049

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