UNIVERSITÄT
BAYREUTH

Kulturwissenschaftliche Fakultät

Validitätsstudie zur Talentforschung

Prüfung der prognostischen Validität von Talentidentifikation

Maximilian Siener

Die vorliegende Dissertation wurde am 27. Juli 2022 von der Promotionskommission der Kulturwissenschaftlichen Fakultät der Universität Bayreuth als Dissertation angenommen.
Gutachtende Professoren waren
Professor Dr. Andreas Hohmann, Universität Bayreuth (Erstgutachter),
Professor Dr. Mark Pfeiffer, Johannes Gutenberg-Universität Mainz (Zweitgutachter),
Professor Dr. Martin Lames, Technische Universität München (Drittgutachter).

Herzlichen Dank für den fachlichen Rat, den kritischen Diskurs sowie die Begleitung und Bewertung meiner Dissertation.

Danksagung

Die vorliegende Arbeit stellt als bedeutenden beruflichen Lebensabschnitt das Resümee einer lehrreichen und intensiven Zeit dar. Auf dem Weg dorthin haben mich viele Menschen begleitet und mir mit Rat und Tat zur Seite gestanden. Dafür möchte ich ihnen meinen Dank aussprechen.

Lieber *Andreas*, ich danke dir, dass du zu jeder Zeit ein offenes Ohr für mich hattest und all meine Ideen und Planungen mit mir verwirklicht hast. Du hast mich auf meinem Werdegang in hohem Maße unterstützt, mir viel ermöglicht und auch in manch schwieriger Situation den Rücken gestärkt und vertraut. Gemeinsam haben wir spannende Projekte realisiert und diese auf Kongressen vorgestellt. Ich werde es vermissen, mit dir im Büro zu sitzen und über die regionalen und (inter-)nationalen Sportereignisse, die Sportförderung und anstehenden Studien zu diskutieren. Danke für diese großartige und lehrreiche Zeit.

Liebe *Elena*, du hast mich immer unterstützt, mich motiviert und mir den Rücken freigehalten. Ohne dich wäre ich nicht so weit gekommen. Danke, dass du immer für mich da bist.

Ein großer Dank gilt auch meiner *(Schwieger-)Familie*, die jederzeit zum gedanklichen Austausch bereit war und mich bei meiner Arbeit mit Vorschlägen und Ideen sowie mit der Herstellung von Kontakten sehr geholfen hat.

Ebenfalls möchte ich allen Arbeitskollegen danken, die mir bei meinen Studien und meiner Promotion behilflich waren und von denen ich viel lernen durfte. Liebe *Andrea*, lieber *Frischi*, lieber *Stefan*, lieber *Alex*, liebe *Susanne*, lieber *Clemens*, lieber *Johan*, liebes *Institut*, vielen Dank für eure Unterstützung.

Folgende Artikel wurden für die Promotion eingereicht:

1. Hohmann*, A., Fehr, U., Siener, M., & Hochstein, S. (2017). Talentscreening und Talentorientierung: Zur Validität von Bewegungs-Checks im Grundschulalter. *Leistungssport*, *47*(4), 11–21.
2. Hohmann, A., Siener*, M., & He, R. (2018). Prognostic validity of talent orientation in soccer. *German Journal of Exercise and Sport Research*, *48*(4), 478–488. https://doi.org/10.1007/s12662-018-0549-5
3. Siener*, M., & Hohmann, A. (2019). Talent orientation: the impact of motor abilities on future success in table tennis. *German Journal of Exercise and Sport Research*, *49*(3), 232–243. https://doi.org/10.1007/s12662-019-00594-1
4. Hohmann*, A., & Siener, M. (2021). Talent Identification in Youth Soccer: Prognosis of U17 Soccer Performance on the Basis of General Athleticism and Talent Promotion Interventions in Second-Grade Children. *Frontiers in Sports and Active Living*, *3*, 625645. https://doi.org/10.3389/fspor.2021.625645
5. Siener*, M., Ferrauti, A., & Hohmann, A. (2021). Early talent identification in tennis: A retrospective study. *International Journal of Racket Sports Science*, 3(2).
6. Siener*, M., Faber, I. R., & 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

* Corresponding Author

Inhaltsverzeichnis

1. Einleitung und Problemstellung 1
2. Theoretischer Rahmen 4
2.1 Talent 4
2.1.1 Der Talentbegriff 4
2.1.2 Talentmodelle 7
2.1.3 Talentidentifikation und Talententwicklung 9
2.2 Gütekriterien 13
3. Validierungsschritte prognostischer Validität 16
3.1 Auswahl der Prädiktoren (Predictor) 17
3.2 Auswahl der Stichprobe (Range) 28
3.3 Bestimmung des Kriteriums (Criterion) 30
3.4 Bestimmung der Selektionsrate (Selection Ratio) 34
3.5 Bestimmung der Prognosemethode (Methods) 37
3.6 Prognosegüte (Prediction) 42
3.7 Nützlichkeit (Utility) 48
4. Einordnung der Publikationen in die wissenschaftliche Talentforschung 58
4.1 Talentscreening und Talentorientierung: Zur Validität von Bewegungs-Checks im Grundschulalter 63
4.2 Prognostic validity of talent orientation in soccer 65
4.3 Talent orientation: the impact of motor abilities on future success in table tennis 66
4.4 Talent Identification in Youth Soccer: Prognosis of U17 Soccer Performance on the Basis of General Athleticism and Talent Promotion Interventions in Second-Grade Children 68
4.5 Early talent identification in tennis: A retrospective study 69

4.6 Prognostic Validity of Statistical Prediction Methods Used for Talent Identification in Youth Tennis Players Based on Motor Abilities 70
5. Diskussion 72
6. Zusammenfassung 84
7. Empfehlung für die Praxis 85
8. Literaturverzeichnis 87
Appendix 105
Artikel 1 105
Artikel 2 117
Artikel 3 129
Artikel 4 142
Artikel 5 162
Artikel 6 176

Abkürzungsverzeichnis

ACC	Accuracy
ANOVA	Analysis of Variance
APHV	Age Peak Height Velocity
AUC	Area Under [the] Curve
AUROC	Area Under [the] ROC-Curve
CV	Cross-validation
DA	Discriminant Analysis
FBC	Fuldaer Bewegungs-Check
KFM	Kohonen Feature Map
MANCOVA	Multivariate Analysis of Covariance
MCC	Matthews Correlation Coefficient
MLP	Multilayer Perceptron
OR	Odds Ratio
RIOC	Relative Improvement Over Chance
ROC	Receiver Operating Characteristic
SD	Standard Deviation
SRS	Specific Recommendation Score
TDE	Talent Development
TID	Talent Identification
U9	Group of children under 9 years old
VO_2max	Maximal Oxygen Consumption

Abbildungsverzeichnis

Abb. 1. Differentiated-Model-of-Giftedness-and-Talent (Gagné, 2004)……7

Abb. 2. Person-Umwelt-Konzept der Talententwicklung (Güllich & Krüger, 2013)……8

Abb. 3. Talentidentifikation und -förderung in Deutschland (mod. nach Güllich & Krüger, 2013)…….…… 10

Abb. 4. Entwicklung der persönlichen Bestleistung von zwei Weltrekordhalterinnen im Hochsprung (Güllich & Krüger, 2013, S. 634)…….. 19

Abb. 5. Spinnennetzdiagramm zum Merkmalsprofil der Sportarten Handball und Kanu/Kajak (Teunissen et al., 2021)…… 24

Abb. 6. Komplementäres Verhältnis zwischen Reliabilität und Validität (Ferrauti et al., 2020, S. 105)…….. 26

Abb. 7. Veränderung des Zusammenhangs zweier Variablen durch Vorselektion…….. 29

Abb. 8. Streudiagramm zur Verdeutlichung von einer möglichen Level-Einteilung im Jugendbereich der Sportart Tennis……...… 32

Abb. 9. Entwicklung der Wettkampferfolge von sechs britischen Goldmedaillengewinnern im Zeitraum bis zu ihrer ersten Goldmedaille (Güllich, 2019)……..…….. 34

Abb. 10. Vier-Felder-Streudiagramm der Talentidentifikation……. 35

Abb. 11. Positiver prädiktiver Wert (Prozentsatz „Geeigneter“) in Abhängigkeiten von der Validität des Prädiktors (Korrelation: $r = 0{,}55$ und $r = 0{,}35$), der Selektionsrate/-quote und der Grundrate (nach den Taylor-Russell-Tafeln von Taylor & Russell, 1939; Lienert & Raatz, 1998, S. 421)……36

Abb. 12. Diskriminanzanalyse zur Untersuchung der Gruppe Kajak/Kanu und achtzehn anderen Sportarten (Teunissen et al., 2021)……..39

Abb. 13. Beispiel einer möglichen Netzstruktur eines künstlichen Neuronalen Netzes (Backhaus et al., 2015, S. 299)…….. 40

Abb. 14. Grafische 4-Felder-Tafel zur Prognosegüte (Hohmann et al., 2014)….. 43

Abb. 15. 6-Felder-Tafel zur Darstellung der möglichen Zusammenhänge zwischen Potential und Leistung (Baker et al., 2018)……44

Abb. 16. Prognoseergebnisse von vier Prognose-/Klassifikationsmethoden zum Drop-Out-Verhalten junger Turnerinnen nach fünf Jahren (Pion et al., 2017)…… 45

Abb. 17. Sensitivität und Spezifität in Abhängigkeit verschiedener Cut-Off-Werte (Höner & Votteler, 2016, S. 274)…….. 47

Abb. 18. Beispiel einer "receiver operating characteristic"-Kurve (ROC Curve) und Precision-Recall-Kurve (PR Curve) in Abhängigkeit des jeweils gewählten „precision score"-Grenzwerts (Azevedo, 2020)…….. 54

Abb. 19. Erstes Modell zur Validierung von Talentidentifikation…….64

Abb. 20. Validierungskonzept mit möglichen Schritten und Kennwerten zur Bestimmung der prognostischen Validität von (Talent-)Prognosen……85

Tabellenverzeichnis

Tab. 1. *Vier-Felder-Tafel zum Talentbegriff (Ferrauti, 2020)*…… 6

Tab. 2. *Kriterien der Talentauswahl für D-Kader in verschiedenen olympischen Sportarten (Güllich & Krüger, 2013, S. 647)*…… 18

Tab. 3. *Talentgruppierung anhand der besten jemals erreichten Wettkampfleistung im Schwimmen (nach Pfeiffer & Hohmann, 2012, S. 347)*……31

Tab. 4. *Ansätze zur Prognosewerterstellung in der Talentidentifikation (Pfeiffer & Hohmann, 2012)*……... 40

Tab. 5. *2×2-Klassifikationstabelle*…….. 43

Tab. 6. *Prognosegüte zweier Klassifikationsmethoden (mod. nach Musa et al., 2019)*……. 48

Tab. 7. *Descriptive values of the receiver operating characteristic curves (Sieghartsleitner, Zuber, Zibung, Charbonnet & Conzelmann, 2019, S. 7)*……. 55

Tab. 8. *Sportmotorische Testaufgaben des Fuldaer Bewegungs-Checks*……59

Tab. 9. *Leistungslevel entsprechend der erreichten Wettkampferfolge*…….. 61

Tab. 10. *Sensitivitätswerte der sportartspezifischen Gewichtungen bei einer vorgegebenen Spezifität von 82,9%*..... 77

Formelverzeichnis

Formel 1. Berechnung des Prognosewertes…… 38

Formel 2. Sportartspezifische Gewichtung nach Höner und Votteler (2016)……. 38

Formel 3. Prognosewert mod. nach Höner und Votteler (2016)……...…. 38

Formel 4. Berechnung der Gesamttrefferquote…….. 45

Formel 5. Berechnung der Selektionsquote…… 45

Formel 6. Berechnung der Sensitivität……46

Formel 7. Berechnung der Spezifität…….. 46

Formel 8. Berechnung des positiven prädiktiven Wertes……... 47

Formel 9. Berechnung des negativen prädiktiven Wertes…….. 47

Formel 10. Berechnung der Zufallstrefferquote……50

Formel 11. Beispiel der Berechnung einer Zufallstrefferquote…… 50

Formel 12. Berechnung des Youden Index J……51

Formel 13. Berechnung des gewichteten Youden Index J'……...51

Formel 14. Berechnung des F_1-Scores……..51

Formel 15. Berechnung des Matthews-Korrelationskoeffizienten……... 51

Formel 16. Berechnung von Cohens Kappa κ…….. 52

Formel 17. Berechnung des RIOC-Wertes……... 52

Formel 18. Berechnung der maximalen Trefferquote……..…… 52

Formel 19. Berechnung des Brier-Wertes P……. 56

Formel 20. Berechnung des Likelihood-Quotienten LR^+……. 56

Formel 21. Berechnung des Likelihood-Quotienten LR^-……. 56

1. Einleitung und Problemstellung

Der Leistungssport hat viele Facetten. So können errungene Siege und Medaillen, bspw. bei sportlichen Großevents wie den Olympischen Spielen, das Prestige eines Landes erhöhen oder Sportler[1] sportliche Höchstleistungen buchstäblich in bare Münze verwandeln. Zudem gehört der sportliche Wettkampfgedanke zu einer kulturellen und gemeinschaftlichen Grundvorstellung des Sports in der Bevölkerung, was zur Folge hat, dass Menschen bei Sportevents Partei ergreifen, mitfiebern und auch selbst aktiv werden, um ihren Vorbildern nachzueifern oder nach eigenen Höchstleistungen zu streben. Die Förderung des Leistungssports ist somit in vielerlei Hinsicht bedeutsam und berechtigterweise ein zentraler Teil der Sportwelt.

Einer der Grundpfeiler des erfolgreichen Leistungssports stellt dabei die Suche und die Förderung von talentierten Nachwuchssportlern dar (De Bosscher et al., 2006). Beide Maßnahmen beginnen zumeist in jenem Alter, in dem in eine Sportart eingestiegen werden kann, und enden erst mit dem Erreichen der abschließenden Höchstleistung zum Abschluss eines *langfristigen Trainingsaufbaus*. Nicht selten erstreckt sich dieser Zeitraum über mehr als 10 Jahre (Hohmann et al., 2020). Gerade kleinere Länder und Vereine, die aus einem begrenzten Pool geeigneter Athleten (gleichzeitig) nur eine geringe Anzahl an Personen fördern können, haben mithilfe eines ausgereiften Talentkonzeptes die Chance, im allgemeinen Gerangel um sportliche Bestleistungen mithalten zu können (Vaeyens et al., 2009).

Hält man sich die wirtschaftlichen Erfolge einer erfolgreichen Nachwuchsarbeit vor Augen, ist es nicht verwunderlich, dass viele Institutionen und Vereine hohe Summen in die Talentidentifikation[2] (TID) und Talentförderung stecken (Brouwers et al., 2012). Neben dem unmittelbaren Erfolg, den ein talentierter Sportler dem Verein verspricht, besteht zudem die Möglichkeit, mithilfe einer erfolgreichen Nachwuchsarbeit einen Sportler langfristig an einen Verein zu binden. So wurde bspw. der siebenfache *Weltfußballer des Jahres*, Lionel Messi, bereits mit 14 Jahren in den *FC Barcelona* aufgenommen und durchlief dort alle Jugendmannschaften. Insgesamt blieb er dem Verein über 20 Jahre treu. Ein ähnliches Beispiel liefert der Fußballtrainer Josep „Pep“ Guardiola. Er gehörte ebenfalls der „Talentschmiede“ Barcelonas an und spielte insgesamt elf Jahre in der katalanischen Hauptstadt. Später trainierte er zunächst die zweite Mannschaft und gewann anschließend als Trainer zweimal mit seinem Stammverein die *UEFA-Champions League*.
Daher ist es nicht verwunderlich, dass Vereine einen besonderen Wert auf TID und Talentförderung legen und dafür zu außergewöhnlichen Maßnahmen greifen. So bot beispielsweise (bspw.) der niederländische Erstligist *VVV Venlo* einem durch ein YouTube-

[1] Aufgrund der besseren Lesbarkeit wird im Text das generische Maskulinum verwendet. Gemeint sind jedoch immer alle Geschlechter.
[2] Die Begriffe (Talent-)Identifikation und (Talent-)Identifizierung werden hier synonym verwendet. Gemeint ist der Vorgang des Erkennens einer Person (als Talent).

Video bekannt gewordenen erst 18 Monate alten Kleinkind (*Baerke van der Meij*) bereits einen 10-Jahres-Vertrag an (Baker et al., 2018).

Neben Vereinsmaßnahmen auf regionaler und nationaler Ebene nehmen TID und Talententwicklung (TDE) auch in nationalen Planungen eine immer größere Rolle ein. So könnte nicht zuletzt das gesteigerte Engagement von Spanien und Deutschland in TID und TDE zu den 2010 und 2014 erreichten Weltmeistertiteln im Fußball geführt haben (Höner et al., 2017). Ähnliche Hoffnungen hegt Großbritannien mit seinem etablierten TID- und TDE-Konzept – dem *Elite Player Performance Plan*.
Das gesteigerte Engagement von Ländern und Vereinen ist ebenfalls in der Wissenschaft zu spüren, wo die Publikationszahlen von wissenschaftlichen Beiträgen rundum TID und TDE in den letzten 20 Jahren stark angestiegen sind. Die meisten Studien stammen dabei aus Australien (8,8%) und Großbritannien (8,8%), gefolgt von Deutschland (5,9%) und den USA (4,8%; Baker et al., 2020).

Trotz dieser Vielzahl an Veröffentlichungen konnten bisher die Zusammenhänge von TID und TDE mit dem späteren Leistungserfolg in den verschiedenen Sportarten noch nicht umfassend geklärt werden. Dies liegt zum einen an dem bisher vorherrschenden Studiendesign: Dreiviertel (77%) aller Studien wurden nach Baker et al. (2020) als Querschnittstudien (68%) entwickelt oder decken nur einen geringen Untersuchungszeitraum (9%) ab. Tatsächlich für die Sportpraxis relevante Langzeitstudien mit einem möglichen retrospektiven Ansatz existieren dagegen vergleichsweise wenige (14%). Gerade diese wären jedoch notwendig, um die prognostische Validität von TID und TDE abschätzen zu können (Mostaert et al., 2020). Hinzu kommt, dass die meisten Studien (65%) nur eine kleine Stichprobe an Probanden betrachten und nicht über 100 Teilnehmer hinausgehen (Baker et al., 2020).
Zum anderen liegt der Fokus der meisten Artikel auf der Untersuchung der Inhaltsvalidität (content validity). Nur etwa 14 % aller TID- und TDE-Studien betrachten die prognostische Validität (S. J. Robertson et al., 2014). Sie ist jedoch mit einer der zuverlässigsten Parameter, um die notwendige Güte einer TID- und TDE-Studie nachzuweisen.

Eine weitere Lücke in der Studienlandschaft wird bei der Untersuchung des Altersbereichs der *mittleren Kindheit* (7. bis 10. Lebensjahr) deutlich. Üblicherweise beginnen Athleten, die im Erwachsenenalter Professionalität zeigen, bereits mit neun Jahren ihre erste Sportart (9,4 ± 2,5 Jahre: Güllich & Emrich, 2014; 9,1 ± 3,7 Jahre: Vaeyens et al., 2009). In Disziplinen, die zugleich mit einem frühen Hochleistungsalter verbunden sind, wie bspw. Turnen oder Eiskunstlauf, ist das Einstiegsalter noch geringer. Ein aktuelles Beispiel verkörpert die erst fünfzehnjährige Winterspiel-Olympiasiegerin des Jahres 2022 im Eiskunstlauf (Teamwettbewerb), *Kamila Walerjewna Walijewa*. Medienberichten zufolge stand sie bereits mit dreieinhalb Jahren auf dem Eis.

Neben diesen offensichtlichen Fällen eines frühen Sportbeginns gibt es auch Sportarten, die einen frühen Einstieg nicht erwarten lassen. So sind bspw. die Tennisspieler *Nadal*, *Federer* und *Djokovic* bereits über 30 Jahre alt. Trotz dieses hohen Leistungsalters wird Tennis dennoch als eine Sportart mit frühem Einstiegsalter bezeichnet (Faber, Bustin et

al., 2016). Studien von Li et al. (2020) zeigten, dass 75 % aller Top 300-Tennisspieler im Alter zwischen 3 und 7 Jahren mit dem Tennisspielen anfingen. Nur 4 % der Top 300-Tennisspieler begannen nach dem 10. Lebensjahr mit ihrer Sportart. Dementsprechend ist es nicht verwunderlich, dass auch die in Europa agierenden Talentakademien, Stützpunkte und Talentförderprogramme bereits Kinder unter elf und teilweise sogar unter neun Jahren aufnehmen (Güllich, 2019; Larkin & Reeves, 2018). Eine frühe TID ist der Modus Operandi im Leistungssport (Baker et al., 2018).

Obwohl frühe Talentmaßnahmen von der Sportpraxis aufgezeigt und durchgeführt werden, sind wissenschaftliche Talentstudien in diesem jungen Altersbereich bisher nur in geringem Maße in der Forschungslandschaft vertreten. Der überwiegende Anteil an Studien umfasst den Altersbereich der Erwachsenen (51,5%) oder des frühen und späten Jugendalters (37,5%; Baker et al., 2020). Nur 11% aller Studien beziehen jüngere Altersbereiche (3–11 Jahre) ein. Es existiert daher eine offensichtliche Diskrepanz zwischen Wissenschaft und Praxis im Feld der frühen TID.

Neben dem Mangel an relevanten Studien im frühen TID-Bereich gibt es zudem insgesamt noch kein klares Konzept zur Validierung von Talentprognosen. Je nach Autor wird die Validität der jeweiligen Talentprädiktoren, Talentvorhersage oder Talentklassifikation mithilfe unterschiedlicher Methoden überprüft und belegt.
So nutzen Bergkamp et al. (2019) und andere Autoren (Faber et al., 2015; Filipcic, 1999; Kramer et al., 2017) *Korrelationswerte,* um die prognostische Validität einzelner Diagnoseverfahren zu überprüfen. Sieghartsleitner et al. (2019) und Höner et al. (2017) verwendeten für eine ähnliche Fragestellung wiederum *Odds Ratios* (OR). Zudem sind auch *Mittelwertvergleiche* (Mostaert et al., 2020; Till et al., 2015), *Trainereinschätzungen* (Pion et al., 2021; Teunissen et al., 2021) und sogenannte *unechte Quasiprognosen* (Hohmann et al., 2015) in den Studien zu diesem Thema zu finden.
Ebenfalls werden für aufgestellte Talentprognosen und Klassifikationsergebnisse mitunter unterschiedliche Validitätsparameter genutzt. Um die Validität der Talentprognose einschätzen zu können, werden in den meisten Studien *Gesamttreffergenauigkeit* und *Sensitivität* verwendet (Pion et al., 2017; K. Robertson et al., 2018; Sieghartsleitner et al., 2017). Da diese Parameter jedoch bestimmte Fragestellung nicht gänzlich abdecken können, wurden zusätzliche Kennwerte eingeführt: Zu ihnen gehört bspw. die Fläche unter der *receiver operating characteristic* Kurve (ROC-Kurve; Faber et al., 2017; Till et al., 2016) oder der *Youden Index* (Sieghartsleitner, Zuber, Zibung & Conzelmann, 2019; Youden, 1950). Diese Vielfalt an Variablen zur Validitätsbestimmung – gepaart mit unterschiedlichen *Grund-* und *Selektionsraten*[3] – erschwert es, Ergebnisse zu vergleichen und ihren Nutzen für die Sportpraxis abzuleiten.
Das Ziel der vorliegenden Studienreihe war es deshalb, die prognostische Validität einer frühen TID mithilfe von (retrospektiven) Langzeitstudien zu überprüfen und die dazugehörigen Validierungsschritte gleichzeitig in ein einheitliches Konzept zur Validierung

[3] Die Begriffe Selektionsrate und Selektionsquote werden hier synonym verwendet.

von TID und TDE einzuordnen. Die jeweiligen Artikel betrachten dabei immer einzelne Aspekte der TID-Validierung und orientieren sich exemplarischen an bestimmten Sportarten. Insgesamt soll somit überprüft werden, ob die von einigen Autoren (Baker et al., 2018; Güllich, 2019; Güllich & Emrich, 2014) als fraglich angesehene, frühzeitige TID im mittleren Kindesalter bereits später erfolgreiche Sportler vorhersagen und, wenn ja, wie die Güte der TID bewertet werden kann.

2. Theoretischer Rahmen

2.1 Talent

2.1.1 Der Talentbegriff

Eine Person, die in gewissen Bereichen talentiert ist, wird bereichsspezifisch als *Talent* bezeichnet. Unter einem Talent kann demnach nicht nur die Ausprägung einer Disposition verstanden werden, sondern auch gleichzeitig die Person, die sie besitzt. Der Talentbegriff bezieht sich hierbei vorrangig auf eine *domainspezifische* (Singer, 1981) und *überdurchschnittliche Begabung* (Wirtz, 2017).
Schnabel et al. (2016) fassen ein Talent als eine Person auf, die auf einem Gebiet herausragende Leistungen vollbringt. So gilt in der Musik jemand als Talent, der fehlerfrei spielen, im Sport hingegen, der bspw. eine hohe Leistung erzielen kann. Hahn (1982) bezeichnet ein Talent als eine in eine bestimmte Richtung ausgeprägte und über ein durchschnittliches Maß hinausgehende, noch nicht entfaltete Begabung. In dieser Definition kommt die Erwartung von zukünftigem Erfolg als eine neue Perspektive hinzu (Cobley et al., 2021; Issurin, 2017). Ein Talent ist demnach auch mit einer Erwartungshaltung verbunden. Während im angloamerikanischen Raum der Talentbegriff auch noch im Erwachsenenalter zur Anwendung kommen kann, werden in der deutschen Sportpraxis und -wissenschaft typischerweise nur die Altersbereiche des Kinder- und Jugendalters auf den Talentbegriff bezogen (Güllich & Krüger, 2013). Zusammengefasst ist Talent „etwas, das man hat, das man ist, das man sein kann oder das man entwickeln kann“ (Mustafović et al., 2020, S. 38).

Definition: Talent
Bezogen auf den Spitzensport ist ein Talent „eine Person, die (a) aus retrospektiver Sicht in ihrer Sportlerkarriere bereits nachweislich Spitzenleistungen erbracht hat oder die (b) unter Berücksichtigung des bereits realisierten Trainings im Vergleich mit Referenzgruppen ähnlichen biologischen Entwicklungsstandes und ähnlicher Lebensgewohnheiten überdurchschnittlich sportlich leistungsfähig ist und bei der man unter Berücksichtigung personeninterner (endogener) Leistungsdispositionen und verfügbarer kontextueller (exogener) Förderbedingungen in prospektiver Hinsicht begründbar annimmt oder mathematisch-prognostisch ermittelt, dass sie in einem nachfolgenden Entwicklungsabschnitt sportliche Spitzenleistungen erreichen kann“ (Hohmann, 2009, S. 11).

Auch Hohmann greift bei seiner Definition für den Spitzensport die beiden Perspektiven „Talent sein“ und „Talent werden“ auf (Hohmann, 2009). Zudem erläutert er einen Ansatz, woran man Talente erkennen kann. Während Joch (1994) nur von „über dem Altersdurchschnitt entwicklungsfähigen Leistungsresultaten“ spricht, wird Hohmann genauer und berücksichtigt nicht nur das kalendarische, sondern auch das biologische Alter. Dies wird aufgrund der üblicherweise unterschiedlichen biologischen Entwicklungsverläufe – gerade im Jugendalter – einem nachhaltigen, prospektiven Ansatz gerechter. Beachtet werden muss jedoch, dass es im Nachwuchsleistungssport weniger darum geht, den besten Sportler zu finden, sondern vielmehr den geeignetsten (Schnabel et al., 2016). Der Eignungsgrad zur Bestimmung von Talenten ergebe sich hierbei aus den beiden Summanden "aktuelle Wettkampfleistung" und "künftige Entwicklungsmöglichkeiten" (Schnabel et al., 2016, S. 438). Man kann bei sportlichem Talent also nicht nur von einem reinen Personenmerkmal ausgehen, sondern sieht sich vielmehr einem Person-Umwelt-Merkmal gegenüber (Güllich, 2019).

Die Bestrebungen nach der Formulierung eines (konsensfähigen) Talentbegriffs sind im englischsprachigen Raum weit weniger fortgeschritten als bspw. im deutschsprachigen (Marconi, 2013). In Deutschland haben die Talentkonzeptionen der ehemals ost- und westdeutschen Forschungsinstitutionen innerhalb der Talentforschung auch in Bezug auf den Talentbegriff eine heute noch fortschreitende Entwicklung hervorgerufen (Gabler & Ruoff, 1979; Kupper, 1993; Senf, 1993; Willimczik, 1982). Hohmann (2001) prägte hierauf aufbauend im Sport anhand von verschiedenen definitorischen Schwerpunkten gleich vier Talentbegriffe mit unterschiedlicher Zielsetzung. Diese wurden von Ferrauti (2020) aufgegriffen und modifiziert (Tab. 1).

Tab. 1. *Vier-Felder-Tafel zum Talentbegriff (Ferrauti, 2020)*

	Eng	**Weit**
Statisch	Person mit aktuell überdurchschnittlichen *Leistungsresultaten* (Gabler & Ruoff, 1979)	Person mit aktuell überdurchschnittlichen Leistungsresultaten, die zudem über die *Bereitschaft* und die notwendigen sozialen und materiellen *Umweltbedingungen* verfügt, sportliche Spitzenleistungen zu vollbringen (Carl, 1989)
Dynamisch	Person mit überdurchschnittlicher *Leistungsfähigkeit* und überdurchschnittlichen *Leistungszuwachsraten* (Kupper, 1993; Siris, 1974)	Person mit überdurchschnittlicher Leistungsfähigkeit und überdurchschnittlichen Leistungszuwachsraten, von der man unter Berücksichtigung *personeninterner Merkmale* und *umweltbezogener Faktoren* für die *Zukunft* begründet annimmt, dass sie in den weiteren Entwicklungsabschnitten sportliche Spitzenleistungen vollbringt (Carl, 1989; Hohmann, 2009; Joch, 1994; Martin et al., 1999)

Grundsätzlich wird hier von einem engen und einem weiten Talentbegriff ausgegangen. Der enge Talentbegriff ist gekennzeichnet von frühzeitiger positiver Auffälligkeit aktueller (sportmotorischer) Leistungen bzw. Leistungszuwachsraten (Güllich, 2013). Damit geht ebenfalls die Erwartung einher, dass diejenigen Kinder und Jugendlichen mit überdurchschnittlichen Wettkampfleistungen auch zu einem späteren Zeitpunkt höchste Spitzenleistungen erreichen (Güllich & Krüger, 2013). Im weiten Talentbegriff werden neben den sportmotorischen Leistungen außerdem Persönlichkeitsmerkmale (z.B. die Leistungsbereitschaft), das Training und das soziale und materielle Umfeld miteinbezogen. Hier besteht die Auffassung, dass diejenigen Sportler das höchste Potential für Spitzenleistungen verkörpern, die neben einem hohen individuellen Leistungspotential auch über günstige personelle und Bedingungen der Umwelt verfügen. Von einem statischen Talentbegriff wird ausgegangen, wenn eine einmalige Erhebung der Talentprädiktoren zu einem bestimmten Zeitpunkt zugrunde liegt. Werden mehrmalige Erhebungen durchgeführt, sodass ein Leistungsverlauf entsteht, aus dem man die entsprechenden Leitungszuwachsraten ableiten kann, spricht man von einem dynamischen Talentbegriff (Güllich & Krüger, 2013).

Heutzutage geht man häufig nur noch von einem weiten und dynamischen Talentverständnis aus, da eine Person, die einmalig die Voraussetzungen eines Talents erfüllt hat, zumeist nicht für alle Zeit auch als solches bezeichnet wird (Baker et al., 2020; Till & Baker, 2020). Man nimmt demnach an, dass zukünftiges Potential veränderbar, also plastisch ist (Güllich & Krüger, 2013). Büch (2005) hingegen sieht ein Talent als eine Person an, der das Potential für Wettkampferfolge und höchste Leistungen bereits innewohnt. Dieses Potential müsse man mit verschiedenen Ressourcen (Training, Förderung, Familie, Motivation, etc.) nur noch freilegen. Eine Übersicht dieser möglichen Ressourcen wird in Talentmodellen dargestellt.

2.1.2 Talentmodelle

Die Vorstellungen von Talentmodellen sind sehr vielfältig (Spies et al., 2022). Exemplarisch soll hier das allgemein bekannte *Differentiated-Model-of-Giftedness-and-Talent* von Gagné (2004) herausgestellt werden (Abb. 1). Es verdeutlich den Zusammenhang zwischen natürlichen Fähigkeiten, verschiedenen internen und externen Bedingungen, einem Entwicklungsprozess und der schlussendlichen Ausprägungen des Talents.

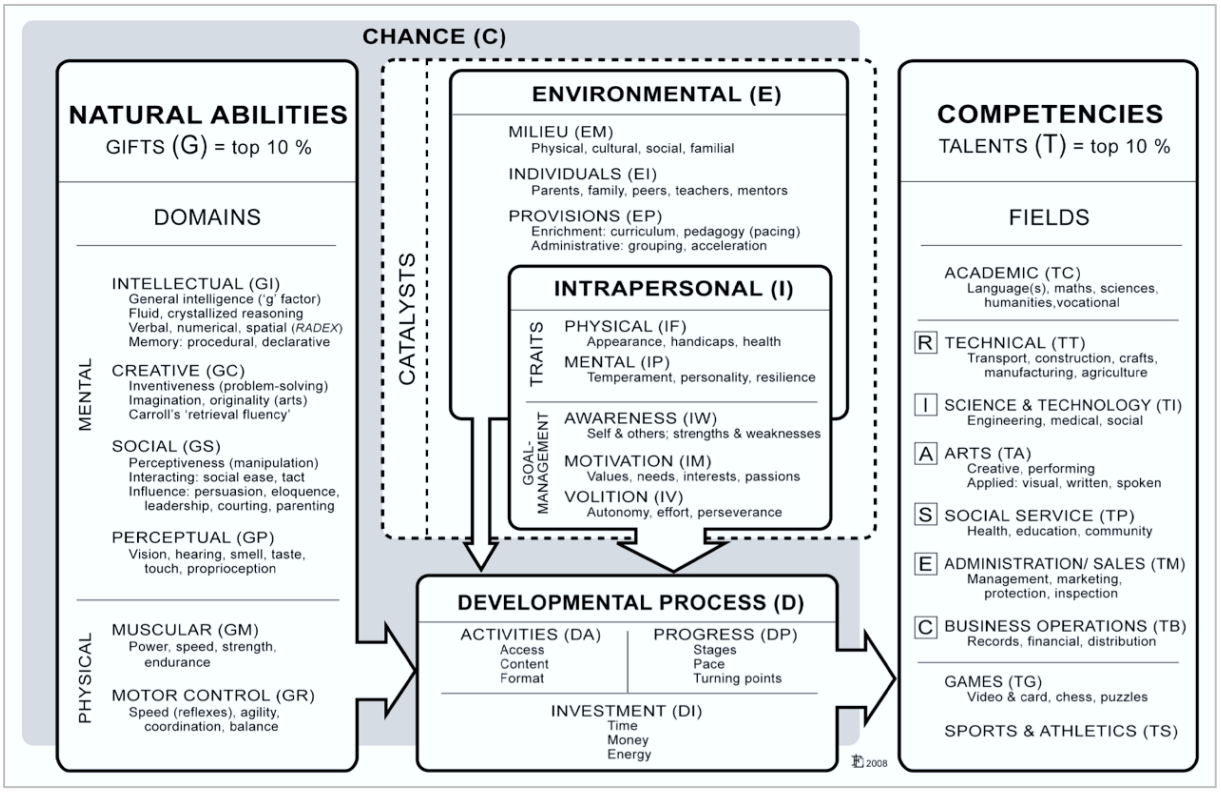

Abb. 1. Differentiated-Model-of-Giftedness-and-Talent (Gagné, 2004)

Gagné (2004) spricht in seinem Modell von den *natürlichen Fähigkeiten* (aptitudes) als Potential, das in jeder Person stecke. Er sieht jedoch nur die besten 10% aller Sportler als begabt an. Das besondere Potential für späteren Höchstleistungen zeige sich in verschiedenen Domänen. Neben den konditionellen und koordinativen Fähigkeiten werden auch intellektuelle und soziale Punkte genannt (vgl. enger Talentbegriff). Sie sind jedoch nicht zwangsläufig rein genetisch bedingt, sondern entwickeln sich im Laufe des Lebens – im jungen Alter mehr als im Erwachsenenalter.
Um diese Ausgangspunkte in eine Höchstleistung zu überführen, bedarf es eines Entwicklungsprozesses. Dieser wird wiederum bedingt von verschiedenen *Catalysts* bzw. *Moderatoren* (vgl. weiter Talentbegriff). Erst durch sie wird ein Potential sichtbar, und das Talent – die herausragende Beherrschung von systematisch entwickelten Fähigkeiten – zeigt sich. Als Moderatoren sind hier sowohl Umwelteinflusse, wie bspw. das soziale Milieu, Trainer oder die elterliche Unterstützung (Fahrdienste, etc.), als auch intrapersonelle Faktoren, wie bspw. Motivation oder Volition, zu sehen. Bleibt einer der

beiden Kernpunkte (*Potential* und *Moderator*) aus, so kann eine begabte Person ihre Begabung nur in seltenen Fällen voll entfalten.
Der Entwicklungsprozess selbst verläuft über verschiedene Stufen vom *Novizen* bis zum *Experten* (Stages) und ist abhängig von Zeit (Trainingsumfang und -häufigkeit), Energie und Geld, das man in die Entwicklung investiert. Auch das Lernumfeld (Stützpunkte, Vereine, etc.) spielt eine besondere Rolle.
Viele Faktoren auf dem Weg zum professionellen Sportler, wie der Körper, in den eine Person hineingeboren wird, ihr familiärer Hintergrund oder die Verletzung eines anderen Sportlers, dessen (Stamm)Position man einnehmen kann, sind vorgegeben oder dem Zufall geschuldet. Glück und Pech sind daher ebenfalls ein Teil des Modells und Wegbegleiter vieler Topathleten. (Gagné, 2004)

Ein inhaltlich ähnliches Modell erstellten Güllich und Krüger (2013) in Anlehnung an Heller et al. (2005). Auch in ihrem Modell (Abb. 2) ist das jeweils zugrundeliegende Potential des Sportlers Ausgangspunkt für spätere Spitzenleistungen.

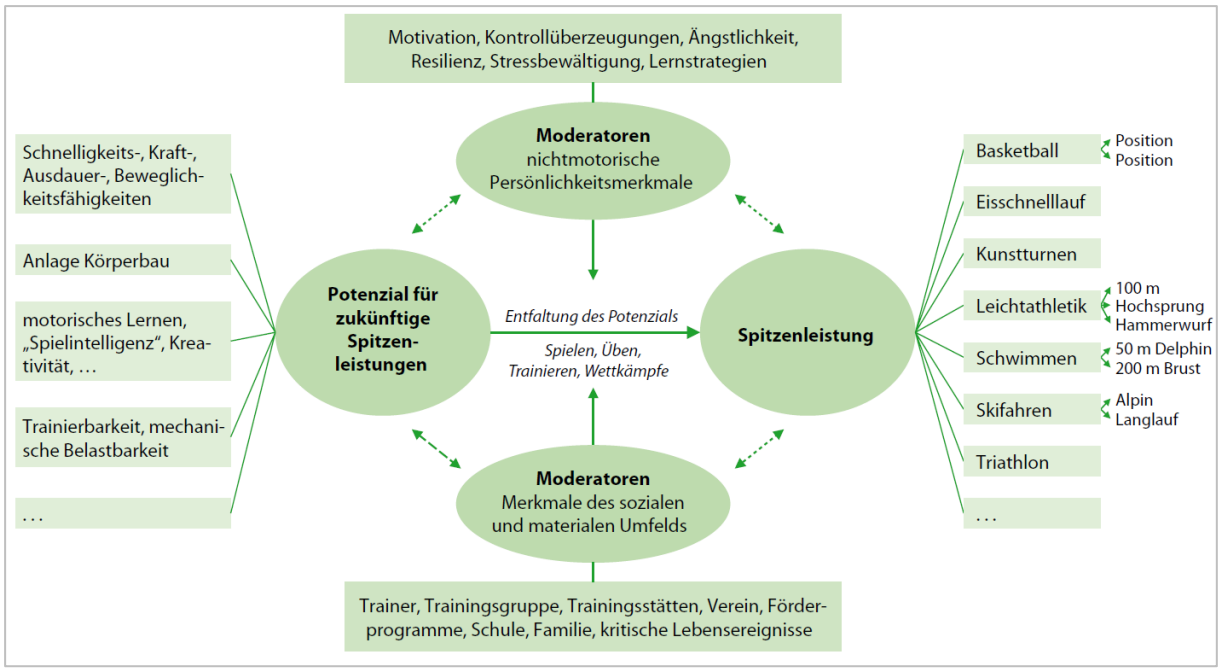

Abb. 2. Person-Umwelt-Konzept der Talententwicklung (Güllich & Krüger, 2013)

Das Potential einer Person setzt sich hierbei unter anderem aus den konditionellen und koordinativen Fähigkeiten und anthropometrischen Eigenschaften zusammen. Wie im Modell von Gagné (2004) wird die Entfaltung dieses Potentials bedingt von verschiedenen Moderatoren. Zu ihnen gehören Merkmale des sozialen und materialen Umfelds, wie bspw. die familiäre Situation, die Güte der Förderprogramme oder die Qualität der Trainer – und nicht motorische Persönlichkeitsmerkmale, wie bspw. Motivation, Volition, Copingstrategien oder Mnemotechniken. Beide Merkmale sind sowohl mit dem Potential (große, kräftige Körpertypen könnten z.B. weniger ängstlich sein) als auch mit der schon erreichten Leistung verbunden (gute Leistungen erleichtern z.B. die Aufnahme in Förderprogramme) und wirken zentral auf die Entfaltung des Potentials ein. Das Erreichen der späteren sportartspezifischen Spitzenleistung ist hierbei abhängig von

den geleisteten Trainings- und Übungsstunden sowie von den Erfahrungen in Wettkämpfen und Turnieren. Anders als Gagné (2004) nennen Güllich und Krüger (2013) keine konkreten Zahlen, ab welchem Referenzbereich (z.B. die besten 10% der Altersgruppe) ein Talent als solches gesehen wird. Auch der Glücksfaktor spielt in ihrem Modell keine explizite Rolle.

2.1.3 Talentidentifikation und Talententwicklung

Die Suche der Trainer nach talentierten Sportlern beginnt (teils unterbewusst) bereits, wenn Athleten in ihre Sportart einsteigen und mehr oder weniger systematisch Sport treiben (Hohmann, 2009). So berichtete der australische Schwimmtrainer und Forscher Brian Blanksby (1980), dass er bereits in der Schule die erste Talentauswahl[4] vorgenommen habe (Güllich, 2019). Er wählte verschiedene Schüler aus, die beim Gehen die Füße nach außen stellten und ließ sie sich auf das Brustschwimmen, die Paradedisziplin der Schule konzentrieren.
Solche Auswahlsituationen kennen viele Sportler. Um in teilweise stark heterogenen Gruppen differenzieren zu können und jedem Athleten den für ihn optimalen Trainingsplan zu erstellen, müssen Trainer die Fähigkeiten ihrer Sportler kennen und nach möglichem Potential bewerten. Hierbei wird schon früh nach etwaigen Talenten geschaut, um diese adäquat fördern zu können.

In der Talentforschung haben sich mehrere Begriffe etabliert, die unter die zwei Überbegriffe *Talentidentifikation* (im weitesten Sinne; teilweise auch Talentdiagnose genannt; talent identification [TID]) und *Talentförderung* (teilweise auch talent promotion genannt) fallen (Marconi, 2013; Williams et al., 2020). Die Talentidentifikation umfasst hierbei den Prozess und alle darin enthaltenen Maßnahmen zum Erkennen und Einordnen von Personen, die das Potential haben, später Spitzenleistungen zu vollbringen (Mohamed et al., 2009; Sarmento et al., 2018; Williams & Reilly, 2000). Unter Talentförderung versteht man die „Gesamtheit der Strukturen und Maßnahmen, denen die beabsichtigte Funktion zugrunde liegt, durch gezielte Interventionen Verhaltensweisen zu stärken, denen positive Effekte auf die Erhöhung der Wahrscheinlichkeit langfristiger Erfolge im Spitzensport zugeschrieben werden“ (Emrich et al., 2005, S. 63). William und Reilly (2000) bezeichnen die Talentförderung als die Unterstützung eines Sportlers mit einem optimalen (Lern-)Umfeld, um diesem die Möglichkeit zu geben, sein Potential schneller oder sogar voll auszuschöpfen.

[4] Auswahl der Talente, die in ein Förderprogramm aufgenommen werden sollen (Güllich & Krüger, 2013).

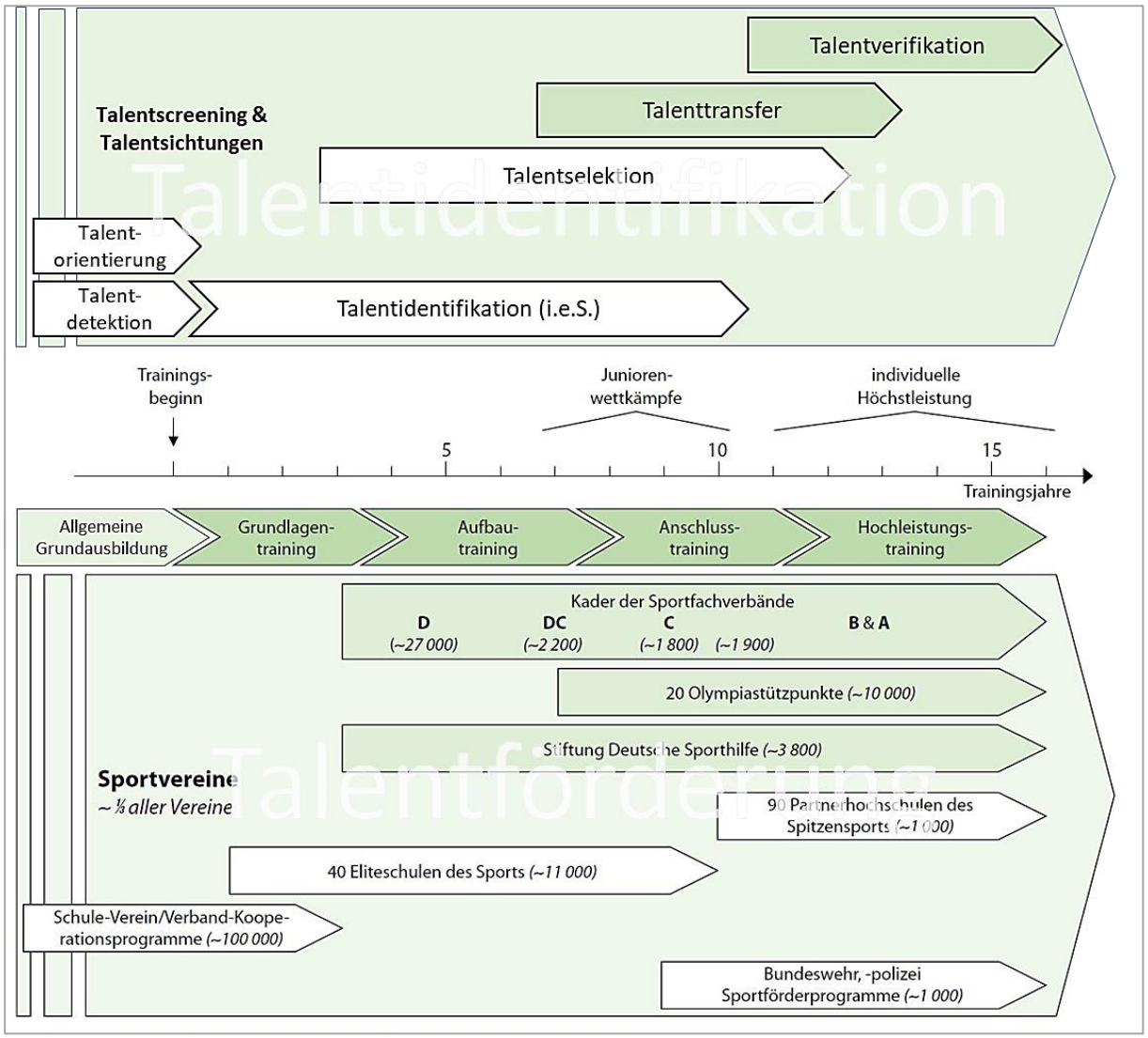

Abb. 3. Talentidentifikation und -förderung in Deutschland (mod. nach Güllich & Krüger, 2013)

Grundsätzlich orientieren sich die Unterpunkte der beiden Teilbereiche *Talentförderung* und *Talentidentifikation* (teilweise auch Talentdiagnose genannt) im deutschsprachigen Raum an der Struktur des langfristigen Leistungsaufbaus (Long-Term-Athletic Development, Pechtl et al., 1993). Diese Strukturierung (Abb. 3) besteht neben der allgemeinen Grundausbildung aus den aufeinander aufbauenden und immer spezialisierter werdenden Trainingsetappen des Grundlagen-, Aufbau-, Anschluss- und Hochleistungstrainings. Im Bereich der Talentförderung nehmen die Vereine über all diese Etappen hinweg eine wichtige Position ein. Ihr Trainings- und Wettkampfbetrieb, ihre Trainer, Trainingslager, finanzielle Unterstützung und ihr Equipment sind in Deutschland ein zentraler Teil des Fördersystems junger Talente.
Einen zweiten wichtigen Pfeiler der Talentförderung bilden die Kader der Sportfachverbände. Für einen Kader (allg. eine Sportlergruppierung) werden vom Landes- oder Spitzenverband diejenigen Sportler ausgewählt, die entweder schon besondere Leistungen erbracht haben oder von denen zukünftige Spitzenleistungen erwartet werden. In einem

Kader sollten sich dementsprechend vorrangig die Talente eines Jahrgangs befinden. Der Kaderstatus, den jeder Kaderteilnehmer erhält, ermöglicht dann die Teilnahme an bestimmten Fördermaßnahmen des Fachverbands (z.B. Stützpunkttraining, Lehrgänge, Teilnahmen an Meisterschaften/Wettkämpfen oder finanzielle Unterstützung) oder anderer Förderorganisatoren (Olympiastützpunkte, Stiftung Deutsche Sozialhilfe, Partnerhochschulen des Spitzensports, Sportförderprogramme der Bundeswehr oder -polizei; ausführliche Erklärung der einzelnen Fördermaßnahmen bei Güllich & Krüger, 2013).

Ebenso wie die Talentförderung lässt sich auch die Talentidentifikation je nach Trainingsetappe und Zielsetzung in verschiedene Unterpunkte untergliedern, die in der Mehrheit auf Talentscreeningverfahren bzw. leistungsdiagnostischen Verfahren oder Sichtungen aufbauen. Teilweise werden für die Bezeichnungen der Unterpunkte mehrere Begriffe synonym verwendet. Die nachfolgenden Begriffserklärungen beziehen sich auf Pion (2015, S. 21-22):

Talentdetektion
(auch Talententdeckung oder Talentsuche genannt) bezieht sich auf die Entdeckung potenzieller Leistungsträger in einer heterogenen Population junger Menschen, die derzeit noch nicht die Zielsportart ausüben (Vaeyens et al., 2008). Die Talentdetektion ist zumeist der erste Schritt im allgemeinen Talentidentifikationsprozess.

Talentorientierung
Die Talentorientierung (teilweise auch Sportartorientierung) steht im Zusammenhang mit der Talentdetektion und zielt darauf ab, Kinder zu motivieren, eine Sportart zu wählen, die ihren individuellen Talentmerkmalen entspricht. Während die Talentdetektion verstärkt auf Leistungsunterschiede zwischen den Athleten einer Sportart achtet, geht es in der Talentorientierung um die Wahl/Empfehlung von für das jeweilige gezeigte Talentprofil geeigneten Sportarten. Im Fokus der Talentorientierung steht somit die Zuweisung von Sportlern in eine für sie passende Sportart.

Talentidentifikation (i.e.S.)
(auch Talenterkennung genannt) bezieht sich auf den Prozess der Erkennung junger Sportler in einer „homogenen" Gruppe mit dem Potential, spätere Höchstleistungen zu erreichen (Williams & Reilly, 2000). Gesucht werden also Personen mit hohem Potential, die bereits Vorerfahrung in der Zielsportart sammeln konnten.

Talentselektion
(auch Talentauswahl genannt) umfasst den fortlaufenden Prozess der Identifizierung von Spielern in verschiedenen Stadien (z.B. an Kaderschwellen), die sich durch ihr Leistungsniveau für die Aufnahme in eine bestimmte Mannschaft oder Gruppierung qualifizieren (Pion, 2015). Sie wird auch als „sehr kurzfristige Talentdetektion" bezeichnet, da es darum geht, die für eine bestimmte Aufgabe am besten geeigneten Sportler zu einem bestimmten Zeitpunkt zu bestimmen (Williams & Reilly, 2000).

Talenttransfer
ist die Möglichkeit, die Hochleistungssportlern geboten wird, ihre sportlichen Fähigkeiten auf eine andere Sportart zu übertragen. Dieser Wandel in den Systemen zur Talentidentifizierung und -entwicklung soll die Wahrscheinlichkeit erhöhen, Athleten zu identifizieren, die eine höhere Kompetenz erreichen können, indem Reifungsprobleme in der Jugend minimiert, Zeiträume für die Talententwicklung verkürzt und die Rendite der bereits in diese älteren Athleten getätigten Entwicklungsinvestitionen maximiert werden (Gulbin & Ackland, 2009).

Talentverifikation
ist die Validierung der Entscheidung bzw. Prognose, die im Rahmen der Talentidentifizierung getroffen wurde. Hierbei erhofft man sich, dass zuvor getätigte Talentvorhersagen im Kindes- und Jugendalter sich nun im angehenden Erwachsenenalter bestätigen.

Die einzelnen Phasen des allgemeinen Talentidentifikationsprozesses verlaufen kontinuierlich und bauen im Optimalfall aufeinander auf (Abb. 3). So könnte in jungen Jahren, bereits bevor ein Athlet mit der Sportart beginnt, eine *Talentdetektion* und *-orientierung* geschehen. Dadurch wäre zum einen gewährleistet, dass ein Sportler später die für ihn geeignete Sportart ausüben würde und sein mögliches Potential optimal ausschöpfen könnte (Pion, 2015). Zum anderen würde er die Möglichkeit erhalten, frühzeitig Erfahrungen in Turnieren und Wettkämpfen zu sammeln und eine optimale nachhaltige *Talentförderung* zu erfahren.

Die Talentförderung beruht dabei unter anderem auf dem in den Maßnahmen der Talentdetektion (*Talentscreening* bzw. *Talentdiagnostik* und *Talentsichtung*) festgestellten Potential des Sportlers. Ist dieses für eine angestrebte Förderung geeignet (*Talentauswahl*), wird der Athlet „rekrutiert" (*Talentrekrutierung*) und kann an Fördermaßnahmen (Talentförderung) teilnehmen (Güllich & Krüger, 2013). Übt ein Sportler bereits die Zielsportart aus, spricht man nicht mehr von Detektion, sondern von Identifikation (i.e.S.). An den Übergängen der verschiedenen Leistungsgruppen (z.B. Kaderübergänge) geschieht dann eine *Talentselektion*. Sie verkörpert die Idee, in bereits trainierten, homogenen Leistungsgruppen weiter zu differenzieren. In dieser Etappe ist auch der Wechsel einer Sportart teilweise noch möglich und sogar ratsam und wird mithilfe des *Talenttransfers* bestimmt. Häufig betrifft er Sportler, denen man das notwendige Potential der Zielsportart nicht prognostizieren kann und ihre sportliche Zukunft in anderen, meist ähnlichen Sportarten höher einschätzt als in ihrer gegenwärtigen Sportart. Beispielsweise gelang es mithilfe des australischen Talenttransfer-Programm des *Australian Institute of Sport* erfolgreich, Sportler der Sportarten Leichtathletik, Rettungsschwimmen und Wasserski in die olympische Wintersportart Skeleton zu überführen (Bullock et al., 2009). Ein anderes Beispiel für einen erfolgreichen Talenttransfer verkörpert der Hochspringer *Donald Thomas*, der im Jahr 2007 mit 2,35m die Goldmedaille gewann. Er war erst im Studium auf Anraten eines Leichtathletik-Trainers vom Basketball zum Hochsprung gewechselt.

Nach Daten von Güllich und Emrich (2014) wechselten 44% der Medaillengewinner Olympischer Spiele oder Weltmeisterschaften in ihrer Karriere die Sportart. Für 52% dieser Sportler (Güllich & Cobley, 2019) geschah der Sportartwechsel nach dem 15. Lebensjahr.
Im Optimalfall bewahrheitet sich die in der Talentidentifikation oder späteren Talentselektion und des Talenttransfers getätigte Talentprognose und ein Athlet kann seine zuvor vorhergesagte Finalleistung in der Talentverifikation bestätigen.
Die einzelnen Teilprozesse der allgemeinen Talentidentifikation entsprechen daher nachfolgender Reihung: Talentdetektion und -orientierung → Talentidentifikation (i.e.S.) → Talentselektion → Talenttransfer → Talentverifikation

2.2 Gütekriterien

Die Prognosen der Talentidentifikation bauen häufig auf einer Leistungsdiagnostik auf (Bös & Schlenker, 2016; Faber et al., 2017; Faber et al., 2018; Fett et al., 2017; Pion, 2015; Pion et al., 2017). „Unter Leistungsdiagnostik im engeren Sinne wird die Messung konditioneller und technomotorischer Leistungskomponenten in einer weitgehend standardisierten Testumgebung unter Einhaltung der Testgütekriterien (Lienert & Raatz, 1998) verstanden“ (Ferrauti, 2020, S. 98). Hierbei können leistungsdiagnostische Maßnahmen sowohl unter Laborbedingungen als auch unter realen sportpraktischen Umgebungsbedingungen (sogenannten *Feldtests*) durchgeführt werden (Ferrauti, 2020). In den Talentstudien der deutschlandweit verbreiteten *Bewegungs-Checks* (Hohmann et al., 2015; Niessner et al., 2020) werden vermehrt letztgenannte Untersuchungen durchgeführt (Bergkamp et al., 2019). *Labortests* sind mitunter aufwändiger in der Durchführung als Feldtests und benötigen teilweise standortgebundene Messapparaturen. Sie sind in einer groß angelegten, dezentral ausgeführten Screening-Kampagne mit teilweise mehreren tausend Teilnehmern schwierig umzusetzen, weshalb Talentidentifikations-Kampagnen überwiegend auf sportmotorischen (Feld-)Tests aufbauen (Pion, 2015; Ulbricht et al., 2016). Testungen, die auf Sportplätzen oder in Sport- und Schwimmhallen stattfinden können, wird eine größer Nähe zur realen sportpraktischen Situation des Sportlers zugesprochen, was sich wiederum auf die Validität – die Gültigkeit der Testergebnisse – positiv auswirken kann.
Neben der Validität werden in der klassischen Testtheorie (Lienert & Raatz, 1998) auch die beiden Gütekriterien Reliabilität und Objektivität zu den Hauptgütekriterien gezählt. Eine Einhaltung dieser drei Hauptgütekriterien und eine standardisierte Testumgebung ist eine zwingende Voraussetzung für die Wertigkeit und Aussagekraft der einzelnen Testaufgaben (Ferrauti, 2020).

Objektivität
Darunter wird der Grad verstanden, in dem die Ergebnisse eines Tests unabhängig vom Untersucher sind (Lienert & Raatz, 1998). Eine hohe Objektivität liegt demnach vor, wenn verschieden Testleiter bei ein und demselben Probanden zum gleichen Ergebnis kommen (Kubinger, 1996). Dies schließt die drei Bereiche Testdurchführung, -auswertung und -interpretation mit ein.

Reliabilität
Unter Reliabilität versteht man den Grad der Zuverlässigkeit (formale Exaktheit), mit dem ein Test ein bestimmtes Merkmal misst, wobei es zunächst gleichgültig ist, ob der Test dieses Merkmal auch zu messen beansprucht (dies wäre ein Frage der Validität; Lienert & Raatz, 1998). Ergebnisse sollten demnach bei einem Test mit hoher Reliabilität zuverlässig, weitgehend reproduzierbar und ohne nennenswerte Messfehler sein (Ferrauti, 2020).

Validität
Die Validität oder auch Gültigkeit gibt den Grad der Genauigkeit an, mit dem ein Test jenes Merkmal, das er messen soll, tatsächlich misst (Lienert & Raatz, 1998). Ein völlig valides Testergebnis würde dementsprechend einen fehlerfreien und unmittelbaren Rückschluss auf den Ausprägungsgrad des zumessenden Merkmals geben. „Während eine ansprechende Reliabilität eine *notwendige*, aber nicht *hinreichende* Bedingung für einen brauchbaren Test ist, stellt die Validität (‚Gültigkeit') ohne Frage das wichtigste Gütekriterium dar – gleichzeitig ist es dasjenige, welches […] am schwierigsten zu prüfen ist" (Kubinger, 1996, S. 40).

Neben diesen drei Hauptgütekriterien existieren weitere Nebengütekriterien, wie bspw. die Normierung, Ökonomie, Nützlichkeit oder Akzeptanz. In bestimmten Situationen kann diesen Nebengütekriterien eine größere Bedeutung zugesprochen werden als den Hauptgütekriterien. So gibt es eine Reihe von Diagnostiken/Tests, wie z.B. die Muskelbiopsie zur Bestimmung des Muskelfaserspektrums, die bezogen auf die Hauptgütekriterien theoretisch einen hohen Grad der Genauigkeit aufweist, jedoch aufgrund von ökonomischen Nachteilen (hohe Kosten, benötigt Fachpersonal, Equipment, etc.) und ethischen Bedenken in der Praxis vergleichsweise selten durchgeführt wird.

Die Validität wird üblicherweise als wichtigstes (Haupt-)Güterkriterium angesehen, da sie etwas über den eigentlichen Zweck der Testaufgabe – die gültige Messung eines Merkmals – aussagt. Validität ist jedoch nicht gleich Validität. Insgesamt lassen sich drei unterschiedliche Aspekte der Validität erkennen, darunter die *inhaltliche Validität* (content validity), die *Konstruktvalidität* und die *Kriteriumsvalidität* (criterion vilidity). Man spricht von hoher inhaltlicher Validität eines Tests, wenn der Test selbst, „quasi *definitionsgemäß,* das optimale Kriterium des interessierenden Merkmals" darstellt (Kubinger, 1996, S. 41). So kann bspw. bei einem Test zur Messung der Körpergröße mittels Maßband von einer hohen inhaltlichen Validität ausgegangen werden, da hier Messwert und Merkmal identisch sind. In anderen Bereichen ist die Verknüpfung

weniger trivial. In solchen Fällen wird die inhaltliche Validität häufig durch ein Expertenrating bestimmt (Lienert & Raatz, 1998; S. J. Robertson et al., 2014). Bezogen auf die Talentidentifikation ist dies bspw. der Fall, wenn Trainer die für eine Talentprognose herangezogene Testauswahl (Diagnostik) und Merkmalsgewichtung (Statistik) bewerten (Johansson & Fahlén, 2017; Murr et al., 2018; Ochs, 2019; Teunissen et al., 2021). Auch die inhaltliche Validität des *Deutschen Motorik-Tests 6–18* konnte mithilfe eines Expertenratings ermittelt werden. Die Aussagekraft der Testbatterie wurde hierbei von den Experten mit einem Mittelwert von 2,1 auf einer Notenskala von eins bis fünf bewertet (Bös & Schlenker, 2011).

Die Konstruktvalidität gibt Aufschluss über die Beziehung der Testleistungen zu einem bestimmten Konstrukt mehrerer Dimensionen (Ferrauti, 2020). Ein Konstrukt besteht zumeist aus bestimmten Beschreibungsmerkmalen (z.B. Zuständen, Instanzen, Eigenschaften oder Fähigkeiten), die nicht direkt beobachtbar bzw. in eindeutiger Weise erfassbar sind, sondern eher einen theoretischen Charakter haben. Typische Beispiele für mögliche Konstrukte sind Angst und Intelligenz (Lienert & Raatz, 1998). Beide Merkmale lassen sich nur indirekt und zumeist über eine Vielzahl verschiedener Testaufgaben (bspw. für Intelligenz durch mathematisches, räumliches, sprachliches und logisches Denken) in Gänze erfassen. Eine Möglichkeit, die Konstruktvalidität eines Tests zu prüfen, gelingt mittels einer Faktorenanalyse (Filipcic, 1999; Kubinger, 1996). Soll ein Test validiert werden, dann spricht man von einer gegebenen Konstruktvalidität, wenn der zu validierende Test bezogen auf die Faktorenanalyse im selben Faktor wie andere konstruktnahe Tests lädt und gleichzeitig diejenige Faktoren mit konstruktfernen Tests nicht lädt (Kubinger, 1996, S. 45). Auch die Konstruktvalidität ganzer Testbatterien kann mithilfe der Faktorenanalyse überprüft werden. Hierbei „sollte bei vollständiger Varianzextraktion jedes Item nur auf einem Faktor laden“ (Bös & Schlenker, 2011, S. 347).

Der letzte Aspekt von Validität ist die Kriteriumsvalidität. Sie kommt dann zum Tragen, wenn es eine Vergleichsmöglichkeit/Außenkriterium zu dem zu validierenden Test gibt, an dem man den Grad der Genauigkeit statistisch bestimmen kann. Die Aussagekraft des Vergleichswertes wird umso zuverlässiger, je valider das Außenkriterium ist. Der sogenannte *Goldstandard* ist im Sport ein von den Gütekriterien sehr hoch einzuschätzendes Außenkriterium (S. J. Robertson et al., 2014). Im Sport wird bspw. die VO_2max als Goldstandard für die Ausdauerleistung gesehen (Smirmaul et al., 2013).

Die Kriteriumsvalidität lässt sich in zwei weitere Teilbereiche unterteilen: die *Übereinstimmungsvalidität* (concurrent validity) und die *prognostische Validität* (predictive validity). Von Übereinstimmungsvalidität wird dann gesprochen, wenn das Außenkriterium zeitgleich mit dem Messwert vorliegt bzw. erhoben wurde. Prognostischen Validität liegt dann vor, wenn das (Außen-)Kriterium erst später, sozusagen in Zukunft, erhoben wird oder werden kann (Lienert & Raatz, 1998). Dies ist zumeist auch für die Talentprognosen der TID der Fall, da sich diese Prognosen der zukünftigen Leistungsfähigkeit einzelner Sportler erst nach einigen Jahren bestätigen können. Das Außenkriterium ist hierbei üblicherweise die Finalleistung/Höchstleistung (Hohmann & Seidel, 2005).

Häufig wird die Kriteriumsvalidität mithilfe von Korrelationsanalysen bestimmt (Faber et al., 2015; Filipcic, 1999). Als statistischer Kennwert der Korrelation von gemessenem Merkmal und (Außen-)Kriterium dient der Korrelationskoeffizienten *r* (Faber et al., 2014). Die Interpretation des Betrags des Kennwerts *r* erfolgt nach folgendem Schema (J. Cohen, 1992; Ulbricht et al., 2016): *trivial* (0 – 0,1), *small* (0,1 – 0,3), *moderate* (0,3 – 0,5) und *large* (0,5 – 0,7).

Neben den Korrelationsanalysen existieren noch weitere Ansätze (Sensitivität, Odds Ratios, etc.), die prognostische Validität von Talentprognosen zu bestimmen (Höner et al., 2017; Höner et al., 2019; Kondric & Filipcic, 2009).

3. Validierungsschritte prognostischer Validität

Bezogen auf die Talentidentifikation spricht man üblicherweise von hoher Validität, wenn der Grad an Genauigkeit der Talentprognosen darüber, wie hoch die spätere Leistung eines Sportlers sein wird, große Aussagekraft besitzt und viele der späteren erfolgreichen oder auch nicht-erfolgreichen Sportler frühzeitig korrekt klassifiziert bzw. prognostiziert worden sind.
Die Talentidentifikation besteht hierbei aus einer Reihe von Teilschritten, an deren Ende erst die Talentprognose steht. Jeder dieser Teilschritte trägt einen Teil zur Gesamtvalidität einer Talentidentifikation bei und lässt sich mitunter auch einzeln validieren.
Folgt man den Aussagen von Bergkamp et al. (2019), besteht jede Talentidentifikation mindesten aus den vier großen Kernbereichen: *Criterion*, *Predictors*, *Range* und *Utility*.
Das Kriterium (Criterion) wird üblicherweise mithilfe von *Leistungsleveln*, die (später) erreicht werden sollen oder auf denen sich Sportler bereits befinden, dargestellt. Hierbei gibt es jedoch keine einheitliche Auslegung, sodass in der Literatur ganz unterschiedliche Levelintervalle (Stufen), Professionalisierungsgrade und auch Bezeichnungen genannt werden (Deprez et al., 2014; Figueiredo et al., 2009; Höner & Votteler, 2016; Zuber et al., 2015, 2016).
Der Prädiktor verkörpert das schon angesprochene Merkmal, das es zu messen oder zu beobachten gilt. Üblicherweise sind dies im Sport verschiedene Leistungsdispositionen wie bspw. sportmotorische Kraft, Schnelligkeit oder Ausdauer (Hohmann et al., 2020). Aber auch Persönlichkeitsmerkmale oder Umweltfaktoren können als Prädiktoren für Talente gelten (vgl. Gagné, 2004; Sieghartsleitner, 2020; Zuber et al., 2015).
Unter der Reichweite (Range) verstehen Bergkamp et al. (2019) die Auswahl der (meist homogenen) Stichprobengruppen. Diese sind häufig aus bestimmten Vereinen oder Kadergruppen zusammengestellt (Dobbin et al., 2019). Eine auf solche Art erhaltene Stichprobe birgt die Gefahr, dass die Probanden schon eine gewisse Vorselektion bspw. durch Trainer oder Vereinsphilosophie erfahren haben und sich die Ergebnisse dadurch nicht verallgemeinern lassen (vgl. externe Validität).
Der vierte Kernbereich ist die Nützlichkeit (Utility). Sie ergibt sich, wenn ein zufälliges Prognoseergebnis mit dem in der Talentidentifikation erreichten Prognoseergebnis (welches wiederum indirekt bedingt wird durch die Selektionsrate) verglichen wird. Nur wenn zwischen Ergebnissen und dem Zufall ein deutlicher Unterschied erkennbar ist,

kann die Talentidentifikation einen Mehrwert für den Leistungssport bieten (Marx & Lenhard, 2010; Schorer et al., 2017).

Bergkamp et al. (2019) lassen in ihren Überlegungen zwei wichtige Teilschritte/Kernbereiche in der Talentprognosegenese aus. Sie gehen weder auf die angewandten (statistischen) *Berechnungsmethoden* zur Erstellung der Talentprognose noch auf die gewählten *Selektionsraten* ein. Letztere geben für stufenweise Talentvorhersagen (z.B. internationaler Erfolg, nationaler Erfolg, regionaler Erfolg) die Anzahl der empfohlenen Sportler pro Stufe an und orientieren sich stark an der in der Praxis vorherrschenden Förderkapazität (z.B. an den Plätzen im Bundeskader, der Fördergruppe oder der Mannschaft). Beide Bereiche haben einen großen Einfluss auf die Talentprognose und sollten daher bei der Bestimmung der Gesamtvalidität mit beachtet werden (Kubinger, 1996; Lienert & Raatz, 1998). Als weiteren Teilschritt bzw. Kernbereich kommt zudem die *Prognosegüte* als Kennwert zur Validitätsbestimmung hinzu.

Insgesamt ergeben sich so sieben Validierungsschritte von TID, welche auf den aufeinander aufbauenden Kernbereichen fußen.

Validierungs- und Kernbereiche von Talentidentifikation:

1. Auswahl der Prädiktoren (Predictor)
2. Auswahl der Stichprobe (Range)
3. Bestimmung des Kriteriums (Criterion)
4. Bestimmung der Selektionsrate (Selection Ratio)
5. Bestimmung der Prognosemethode (Method)

→ ***Erstellung der Talentprognose***

6. Ergebnis der Prognosegüte (Prediction)
7. Ergebnis der Nützlichkeit (Utility)

Das Ergebnis der Prognosegüte umfasst unter den Validierungsschritten der Talentidentifikation den zumeist wertbaren Kern. So wird in vielen Studien in diesem Schritt bspw. der erhaltene Korrelationswert zwischen Prädiktor und Kriterium bestimmt (Faber et al., 2015; Filipcic et al., 2010; Vandorpe et al., 2012) oder die Sensitivität einer Klassifikationsmethode erhoben (Allen et al., 2015; Kovacs, 2006; Till et al., 2016). Diese Werte dienen dem von Bergkamp et al. (2019) bereits genannten letzten Validierungsschritt/Kernbereich der Überprüfung der Nützlichkeit und der abschließenden Validierung der Talentidentifikation.

3.1 Auswahl der Prädiktoren (Predictor)

Howe et al. (1898) weisen einem Talent fünf (Grund-)Eigenschaften zu: (a) es hat seinen Ursprung in genetischen Strukturen; (b) sein volles Potential ist möglicherweise noch nicht in einem frühen Stadium erkennbar, aber es gibt einige Anzeichen, die es geschulten Personen ermöglichen, das Vorhandensein von Talent zu erkennen; (c) diese frühen

Anzeichen von Talent bieten eine Grundlage für eine Talentprognose; (d) nur eine Minderheit ist talentiert; (e) Talente sind relativ bereichsspezifisch.
Bei der zweiten Eigenschaft sprechen die Autoren von gewissen Anzeichen, die es ermöglichen, Talente frühzeitig zu erkennen. Diese Anzeichen werden allgemein als *Prädiktoren* bezeichnet, welche sich wiederum in gewissen *Merkmalen* (Kraft, Ausdauer, etc.) widerspiegeln. Die jeweilige *Merkmalsausprägung* gibt dann die „Höhe" des Merkmals an. Auf den Sport bezogen, könnte dies z.B. die Körpergröße sein und die entsprechende Merkmalsausprägung die Größe in Zentimetern. Betrachtet man mehrere Merkmale eines Athleten gemeinsam, spricht man von einem *Merkmalsprofil* (Sarmento et al., 2018).

Als Prädiktoren können alle möglichen, auf den Sportler bezogene Faktoren eingerechnet werden. Typischerweise orientieren sich Prädiktoren jedoch an den Determinanten der Leistungsfähigkeit (Gagné, 2004), darunter bspw. sportmotorische Fähigkeiten, technische Fertigkeiten oder kognitive und motivationale Faktoren (Faber, Bustin et al., 2016; Fernandez-Fernández et al., 2014; Koopmann et al., 2020). So konnten bspw. die Untersuchungen von Davids und Baker (2007) zeigen, dass stark intrinsisch motivierte, leistungssportorientierte Individualsportler mit hoher Wahrscheinlichkeit das höchste Leistungsniveau erreichen – im Unterschied zu eher versagensängstlichen Athleten. Auch Mosoi (2013) konnte nachweisen, dass ein signifikanter Zusammenhang zwischen intrinsischer Motivation und der nationalen Ranglistenposition von Sportlern besteht.
Möglich sind neben den bereits genannten Merkmalen ebenfalls das soziale oder materielle Umfeld, schon erbrachte Leistungen, (Wettkampf-)Erfolge oder auch schlicht das „Bauchgefühl" einiger Trainer (Sieghartsleitner et al., 2017).

Tab. 2. *Kriterien der Talentauswahl für D-Kader in verschiedenen olympischen Sportarten (Güllich & Krüger, 2013, S. 647)*

Kriterien	**Anteil der Kaderplätze**
Trainerbeurteilung	34%
Wettkampfleistung/-erfolg	83%
Sportmotorische Testleistung	42%
Leistungsentwicklung	4%

Die D-Kaderplätze olympischer Sportarten wurden bspw. zu 83% nach den schon gezeigten Wettkampfleistungen/-erfolgen vergeben (Tab. 2). Einen Kaderplatz aufgrund seiner sportmotorische Testleistungen (42%) oder aufgrund einer Trainerbeurteilung (34%) zu erhalten, traf auf deutlich weniger Sportler zu, außerdem spielte die Leistungsentwicklung mit einem Anteil von nur 4% für die Aufnahme in die D-Kader kaum eine Rolle. Dieser Umstand ist insoweit erstaunlich, da das Verständnis eines dynamischweiten Talentbegriffs nicht nur die einmalige Bestandsaufnahme der Prädiktoren, sondern auch die Betrachtung des Entwicklungsprozesses fordert. Nach Abbott et al. (2005) sei dieses Vorgehen der Talentidentifikation anhand von „snap-shot performance variables" vergleichbar mit der Analyse der Technik eines Baseballschlags mithilfe nur

eines einzelnen Fotos (Abbott et al., 2005, S. 69). Ein Video besitze dagegen eine durchaus höhere Aussagekraft über das Potential des Sportlers.

Um jedoch die für TID notwendigen Daten eines Athleten über einen längeren Zeitraum oder optimalerweise begleitend über seine ganze Karriere zu erheben, benötigt es ein längsschnittliches Studiendesign über mehrere Jahre. Dies ist mit einem großen zeitlichen und zumeist auch finanziellen Aufwand verbunden, weshalb sich die bisherigen Validitätsstudien zur Talentidentifikation – mit wenigen Ausnahmen (Güllich, 2019; Hohmann, 2009; Leyhr et al., 2020) – zumeist auf einzelne Querschnitte oder nur kürzere Zeiträume der Leistungserhebung beschränken und kaum Entwicklungsverläufe oder -vorgänge betrachten (Bergkamp et al., 2019; Johnston et al., 2018). Die wenigen Studien zu Entwicklungsverläufen von späteren Spitzensportlern zeigen zudem teilweise keine signifikanten Befunde. So konnten bspw. Leyhr et al. (2020) in einer Längsschnittstudie über vier Jahre (U12–U15) keine signifikanten Entwicklungsunterschiede zwischen insgesamt 737 professionellen und nicht-professionellen Fußballern feststellen. Dieses Ergebnis scheint nicht verwunderlich, da die Entwicklungsverläufe von Sportler zu Sportler stark unterschiedlich sind und nur in den wenigsten Fällen strikt linear verlaufen (Güllich & Krüger, 2013).

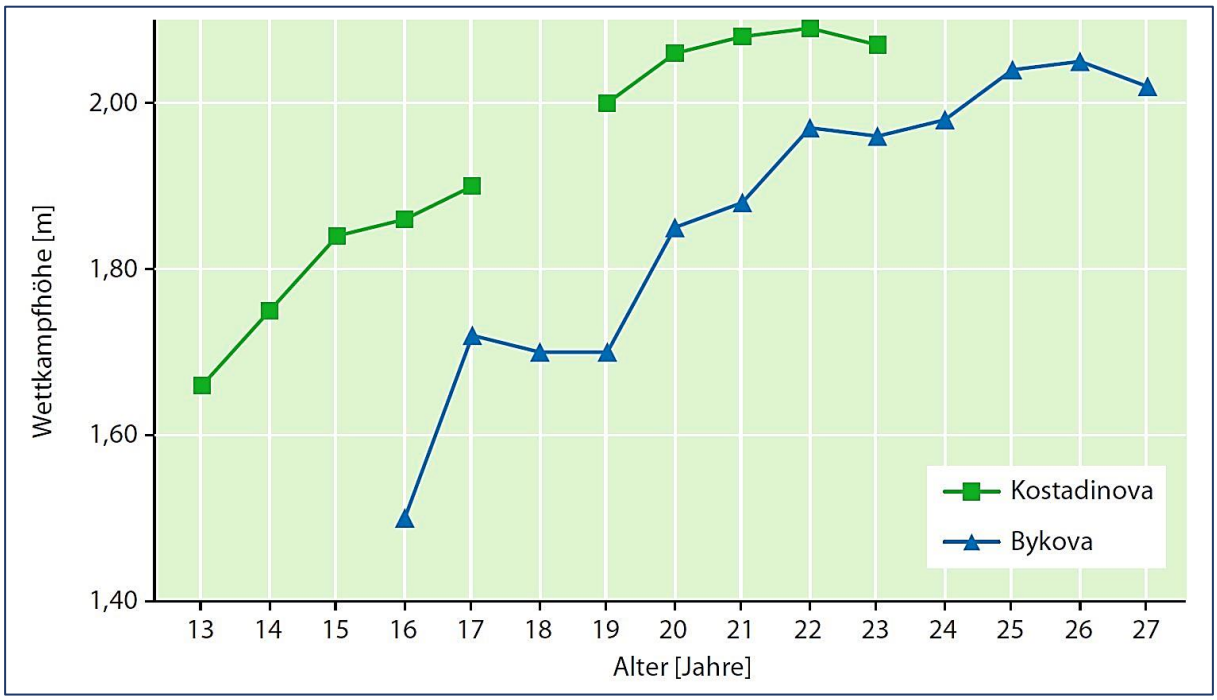

Abb. 4. Entwicklung der persönlichen Bestleistung von zwei Weltrekordhalterinnen im Hochsprung (Güllich & Krüger, 2013, S. 634)

Vergleicht man bspw. die erreichten, persönlichen Jahresbestleistungen der beiden späteren Weltrekordhalterinnen im Hochsprung (Abb. 4), sieht man zwei unterschiedliche Leistungsverläufe. Während die Leistung von *Stefka Kostadinova* (Weltrekord 1987: 2,09m) vergleichsweise linear ansteigt, weisen die Leistungen von *Tamara Bykova* (Weltrekord 1984: 2,05m) immer wieder größere Leistungssprünge auf. Diese Sprünge

im Leistungszuwachs könnten nach Ansicht einiger Autoren (Güllich & Krüger, 2013; Hohmann & Seidel, 2005; Schnabel et al., 2016) ebenfalls auf die spätere Leistungsfähigkeit schließen lassen.

Passend dazu stellte Thiess 1979 vier Diagnosekriterien zur Talenterkennung vor. Neben dem aktuellen Leistungsniveau, dem Grad des ausgeschöpften Entwicklungspotentials (Utilisation[5]) und der Belastungsverträglichkeit nannte er auch das Tempo der Leistungserhöhung als zentrales Merkmal der Talenterkennung (Hohmann, 2009; Thiess, 1979).

Trotz seiner theoretischen Fundierung wird das Tempo der Leistungssteigerung in den wenigsten Studien zur Talentidentifikation miteinbezogen (Bergkamp et al., 2019). Eine große Anzahl an Studien konzentrierte sich stattdessen auf die einmalige Erhebung von sogenannten generischen Tests, also jenen Merkmalen, die in sehr vielen Sportarten auftreten (Pion, 2015). Typischerweise werden hierzu die sportmotorischen Testaufgaben der konditionellen und koordinativen Fähigkeiten gezählt, die mehr oder weniger in jeder Sportart benötigt werden (Grosser et al., 2008). Die meisten Studien konnten jedoch nur geringe bis moderate Effektstärken für die Einflüsse solche Parameter auf die Leistungserfolge feststellen (Bergkamp et al., 2019; Gonaus & Müller, 2012).
Trotz dieser mäßigen Erfolge werden in sehr vielen Studien (Tomkinson et al., 2017; Ulbricht et al., 2016; Vandorpe et al., 2012) sportmotorische Tests eingesetzt, da diese erstens zumeist einfach auswertbar sind und zweitens in genau dieser Form auch in den Zielsportarten auftreten (vgl. inhaltliche Validität).

In vielen Fällen der Talentidentifikation ist jedoch noch nicht ganz klar, welche Leistungsdispositionen ein Sportler bereits früh aufweisen muss und welche er sich noch später im Laufe der Karriere aneignen kann. In diesem Zusammenhang steht die Nature-Nurture-Debatte, die der Frage nachgeht, welchen Einfluss genetische Veranlagung, Trainierbarkeit und Umwelteinwirkungen auf die sportliche Karriere haben. Da noch nicht gänzlich geklärt werden konnte, in welchem Maß die späteren Erfolge von einzelnen Faktoren abhängen, wird von einigen Autoren ein ganzheitlicher bzw. holistischer Analyseansatz favorisiert (Sieghartsleitner, 2020; Zuber et al., 2016). Hierbei versucht man, möglichst viele Merkmale eines Sportlers und seiner Umwelt (und die Interaktion dieser Merkmale untereinander) zu erheben (Sieghartsleitner, 2020). Dieses Vorgehen hat den Vorteil, dass für eine Talentprognose zunächst alle, möglicherweise wichtigen Faktoren miteinbezogen werden und eine hohe Aufklärung der Varianz erreicht wird. Eine damit verbundene, in den Merkmalen „breit aufgestellte“ Talentidentifikation ermöglicht es auch, jene Prädiktoren als solche zu erkennen, die auf den ersten Blick als für eine Talentprognose irrelevant erscheinen, aber dennoch einen Mehrwert für die

[5] „Die Utilisationshypothese besagt nun, dass bei vergleichbarer Leistungsstärke talentiertere Sportler die zur Verfügung stehenden Ressourcen in geringerem Maße in Anspruch nehmen. Für diese Annahme existieren im Nachwuchstraining eine Reihe praktischer Belege. So herrscht beispielsweise Einigkeit darüber, dass es im Hinblick auf die langfristige Leistungsentwicklung positiv ist, wenn die Ressourcennutzung dem Minimaxprinzip folgt: Die höchstmögliche Leistung soll bei möglichst geringem Ressourceneinsatz erbracht werden.“ (Hohmann et al., 2020, S. 302)

Prognoseerstellung besitzen. Als Beispiele könnten die Wurfkraft im Fußball oder die Gleichgewichtsfähigkeit im Tennis genannt werden – beides Fähigkeiten, deren Zusammenhang mit der gewählten Sportart auf den ersten Blick fraglich erscheint. Jedoch können Ausprägungen gerade dieser „Randmerkmale" einem Sportler in einer bereits homogenen Kadergruppe den entscheidenden Leistungsvorteil ermöglichen (Baker et al., 2018).

Die Suche nach geeigneten Prädiktoren ist damit ein Kernbereich der erfolgreichen Talentidentifikation. Güllich (2013) stellte vor diesem Hintergrund drei Kriterien[6] auf, die jedes Diagnosekriterium bzw. jeder Prädiktor erfüllen sollte, um für eine zielgerichtete Talentprognose in Betracht zu kommen:

1. Relevanz
2. Empirische Erfassbarkeit
3. Differenzielle Entwicklungsstabilität

Ein Merkmal ist dann relevant für die Talentidentifikation, wenn es eine bedeutsame Einflussgröße für die spätere Wettkampfleistung vorweist. Dies muss nicht immer direkt geschehen (z.B. durch Körpergröße oder Maximalkraft) – auch Merkmale, die indirekt auf die Wettkampfleistung wirken und bspw. das Training (z.B. Erholungsfähigkeit, Belastbarkeit, etc.) oder auch andere leistungsrelevante Einflussgrößen bedingen (z.B. das Verkraften von Rückschlägen), werden zu den leistungsrelevanten Faktoren gezählt (Güllich, 2013, S. 632). Nur wenn diese *Relevanz* für ein Merkmal vorliegt, lohnt es sich auch, dieses für die Talentidentifikation zu erheben.
Ob eine Erhebungsvariable ein relevantes Merkmal darstellt, lässt sich an drei Voraussetzungen erkennen (Faber, Bustin et al., 2016). Das Merkmal sollte erstens eine Unterscheidung zwischen Spitzensportlern und Nicht-Spitzensportlern aufzeigen, zweitens zukünftige Leistungen vorhersagen können und drittens eine Empfehlung für die Talententwicklung ermöglichen.
Viele Querschnittsstudien rund um das Thema Talentidentifikation konzentrieren sich stark auf die erste Voraussetzung (Filipcic et al., 2010; Meckel et al., 2015; Vaeyens et al., 2006) und untersuchen, welche Tests am besten zwischen professionellen Sportlern und Amateuren unterscheiden.
Ob ein Test auch für eine Prognose über zukünftige Leistungen geeignet ist, lässt sich anhand eines längsschnittlichen Designs über mehrere Jahre bestimmen. Eine der größeren Studien mit mehr als 14.000 Probanden, die sich mit diesem Thema der prognostischen Validität von Diagnosemerkmalen beschäftigt, wurde bspw. von Höner et al. (2017) für die Sportart Fußball vorgestellt. In dieser Studie wurden in der U12 erhobene Testergebnisse mit der nach 8 bis 10 Jahren erreichten Spielleistung verglichen.

[6] Güllich (2013) stellt insgesamt vier Kriterien auf. Das letzte Kriterium, die Erfüllung der Gütekriterien, bezieht sich jedoch streng genommen nicht mehr auf das Merkmal an sich, sondern schon auf die Diagnose-/Erhebungsmethode des Merkmals.

Die dritte von Faber et al. (2016) genannte Voraussetzung bzgl. der Empfehlung für die Talententwicklung wird dann erfüllt, wenn der Test für ein Profil des Spielers genutzt werden kann und eine Beobachtung der Leistungsveränderung zulässt.
Ergänzt werden können die drei Voraussetzungen, die Prädiktoren bzw. Tests erfüllen müssen, außerdem um den Punkt „Test diskriminiert zwischen verschiedenen Sportarten“, da im Falle einer Sportartorientierung weniger die Unterschiede zwischen verschieden erfolgreichen Sportlern im Fokus stehen, sondern verstärkt die Unterschiede zwischen Sportarten betrachtet werden müssen.

Relevante Merkmale sind für die Talentidentifikation jedoch nur dann von Bedeutung, wenn diese zum Zeitpunkt der Erhebung bereits vorliegen und durch Diagnoseverfahren erfasst werden können (*Empirische Erfassbarkeit*; Güllich, 2013). Einige vielversprechenden Merkmale müssen sich zunächst erst entwickeln oder erreicht werden und stehen dementsprechend für eine frühe Talentidentifikation und -detektion nicht zur Verfügung. Ein typisches Beispiel stellen bestimmte Techniken dar, die erst zu einem späteren Zeitpunkt im Laufe des Trainings erlernt werden, oder Wettkampferfolge in Turnieren, die erst ab dem Juniorenalter angeboten werden.
Teilweise sind aber auch Merkmale nur unzureichend messbar, wie bspw. die Fähigkeit, im Wettkampf noch einmal „über sich hinauszuwachsen“ (Güllich, 2013, S. 632) oder die eigentlich notwendige Messmethodik ist aus verschiedenen Gründen nicht verfügbar. Dies ist z.B. bei medizinischen Eingriffen der Fall, die keine breite Akzeptanz in der Bevölkerung oder bei Sportlern erfahren und/oder teilweise aus ethischen Gründen nicht eingesetzt werden, wie bspw. bestimmte medizinische Eingriffe bei Kindern, Kleinkindern oder Säuglingen.

Unter dem dritten Kriterium für ein Diagnosemerkmal, der *differenziellen Entwicklungsstabilität*, versteht Güllich (2013) die Stabilität des Merkmals in Bezug auf eine soziale Bezugsnorm. Dies bedeutet, dass ein Sportler mit einer sehr hohen Merkmalsausprägung zum Diagnosezeitpunkt auch später noch eine vergleichsweise hohe Ausprägung in diesen Merkmalen aufweisen sollte. Dieser Punkt trifft in besonderem Maße auf stark hereditätsabhängige Merkmale zu, wie man dies bspw. der Sprintfähigkeit nachsagt (Schnabel et al., 2016). Die Konstanz des Prädiktors ist aber zumeist ungleich der Konstanz seines Einflusses auf das Kriterium (Hohmann, 2009, S. 28). Die Entwicklungsstabilität eines Merkmals muss daher nicht zwingend gegeben sein, um ein Merkmal erfolgreich für eine Talentidentifikation nutzen zu können.
Viele Merkmale dienen auch als „Mediatoren“ für andere Merkmale, welche sich ohne diese Mediatoren deutlich schlechter entwickeln lassen. Ein einfaches Beispiel hierfür ist die Beweglichkeit, die in jungen Jahren das Erlernen bestimmter wettkampfrelevanter Techniken erleichtern kann, jedoch später hierfür keine Entwicklungsstabilität mehr aufweisen muss. Ebenfalls denkbar ist es, dass sich bestimmte, später instabile Fähigkeiten nur auf die Leistung in einem frühen Stadium auswirken und dort dazu führen, dass der Sportler für einen frühen Leistungskader nominiert wird und eine bessere Förderung erhält. Jene Förderung wirkt sich dann positiv auf die spätere Wettkampfleistung aus. Diese Auffassung eines Diagnosemerkmals entspricht einem dynamisch weiten

Talentbegriff, bei dem sich Talent auch später noch entwickeln kann (Baker et al., 2018; Till & Baker, 2020) und von einer breiten polysportiven Grundausbildung profitiert (LaPrade et al., 2016). Das Kriterium der Entwicklungsstabilität kann daher eher als ein „hinreichendes“, aber nicht „notwendiges“ Kriterium angesehen werden.

Die Auswahl solcher Prädiktoren, die die genannten Kriterien erfüllen, kann sich jedoch mitunter als (methodisch) schwierig herausstellen (Hohmann, 2009), weil häufig nicht ganz klar ist, welche Prädiktoren für eine Talentdiagnose am geeignetsten sind. Gabler und Ruoff (1979) schlagen daher ein Modell zur Gewinnung von aussagekräftigen Diagnosekriterien aus vier gekoppelten Phasen vor:

Gewinnung aussagekräftiger Prädiktoren

1. Suche nach potenziellen Prädiktoren
2. Methodenentwicklung zur Prädiktorerfassung
3. Prüfung des Zusammenhangs zwischen Prädiktor und Kriterium
4. Bewährung des zuvor geprüften Prädiktors

Potenzielle Prädiktoren lassen sich über mehrere Wege finden. Eine Möglichkeit ist es, die Prädiktoren direkt aus der Zielsportart abzuleiten. In vielen Sportarten ist das dafür notwendige Anforderungsprofil des Leistungssportbereichs bereits bekannt (Drid et al., 2015; Franchini et al., 2011; Roetert et al., 1992). Werden bestimmte Situationen im Wettkampf sehr häufig vorgefunden und ist für diese Situationen das vorherrschende Merkmal bekannt, kann wiederum dieses Merkmal für die Talentidentifikation herangezogen werden. Als Beispiele lassen sich Sprünge im Basketball oder Volleyball anführen. In vielen Situationen hat ein Angreifer in diesen Sportarten einen Vorteil, wenn er möglichst hoch springen kann. Daher liegt die Vermutung nahe, dass die Sprungkraft ein entscheidendes Merkmal der Talenterkennung ist.

In anderen Sportarten sind die zugehörigen Merkmale nicht immer so einfach zu erkennen. Bestätigt werden kann in solchen Fällen eine bestehende Merkmalsvermutung durch eine Experteneinschätzung. Hierbei geben üblicherweise Trainer, Sportler oder auch Wissenschaftler eine Einschätzung zur Bedeutung des Merkmals für die gewählte Sportart (Robertson et al., 2018). Mithilfe von Interviews von fünf Nationaltrainern der Junioren-Fußballmannschaften (U16–U18) konnten bspw. Jokuschies et al. (2017) eine Aufstellung vornehmen, welche Kategorien am häufigsten auf ein Talent hinweisen. Die Kategorie "Personality" (bestehend aus: Motivation, Concetration, Aggriveness, Coping, Responsibility) war mit 19 Nennungen am häufigsten vertreten. Die anderen Bereiche kamen nur auf 1 bis 3 Nennungen, darunter auch die "motor abilities" (3 Nennungen) und die "technique" (2 Nennungen).
Neben einzelnen spezifischen Merkmalen lassen sich auf diese Weise aber auch ganze Merkmalsprofile erstellen (Pion et al., 2021; Teunissen et al., 2021). In Untersuchungen von Teunissen et al. (2021) zum Talenttransfer zeigte sich, dass nach Meinung von 17

ausgewählten Trainern die Sportart Handball vor allem von der Wurffähigkeit und der Agilität bestimmt wird (Abb. 5). Dementsprechend sollten Merkmale dieser beiden Fähigkeiten auch für eine Talentidentifikation herangezogen werden.

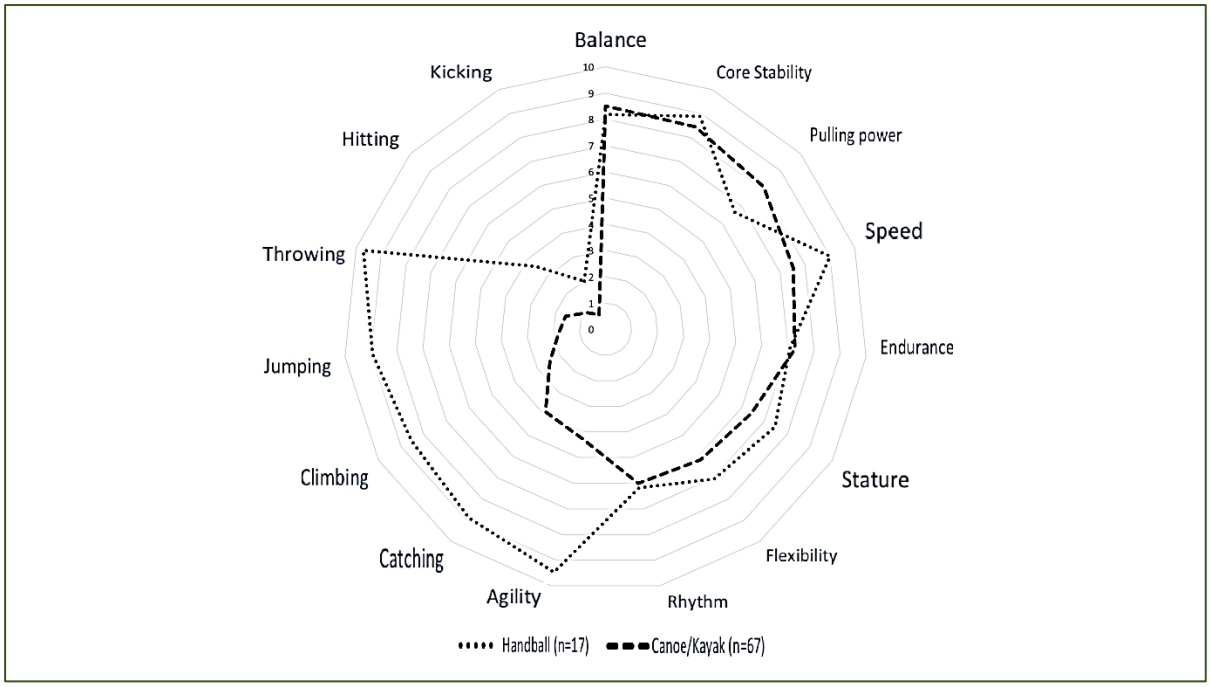

Abb. 5. Spinnennetzdiagramm zum Merkmalsprofil der Sportarten Handball und Kanu/Kajak (Teunissen et al., 2021)

Der Vorteil solcher Erhebungen ist, dass neben den relevanten Merkmalen auch ihr (notwendiger) Ausprägungsgrad ermittelt werden kann, der wiederum zur Erstellung einer Merkmalsgewichtung und zur Berechnung von Talentprognosen genutzt werden kann (Kap. 3.6). Häufig werden Merkmalsprofile professioneller Sportler erhoben und eins zu eins für die Talentprognose junger Sportler verwendet. Diesem Vorgehen liegt die Annahme zugrunde, dass Merkmalsprofile zum einen relativ stabil bleiben und zum anderen auch bei jungen Sportlern das gleiche Sportartprofil vorliegen muss, das Erwachsene aufweisen.

Trotz der möglichen Bedenken, die sich bei diesem Vorgehen auftun können, ist dieses Vorgehen gerade für Sportarten, in denen noch keine pro- oder retrospektive Betrachtung der Leistungsmerkmale vorliegt, eine einfache und praktikable Methode, um Merkmale, Ausprägungen, aber auch Prognosen zu erhalten. Eine auf diese Art erhaltene Prognose wird als *Pseudo*- oder auch unechte Prognose (vgl. unechte „Quasi"-Prognose; Hohmann et al., 2015) bezeichnet. Nach Schnabel et al. (2016) werden demnach für eine Pseudo-Prognose zunächst die Leistungsanforderungen von der Sport-Weltspitze abgeleitet und die Relation der Merkmale des Leistungsprofils untereinander und zur Wettkampfleistung bestimmt (vgl. Merkmalsprofil). Danach erfolgt eine Individualisierung dieser Anforderungen, um den Einfluss von Alterskategorien (z.B. kalendarisches, biologisches oder Trainingsalter) herauszurechnen. Die somit erhaltenen, altersbereinigten „Erwartungsprofile" können nun für eine Talentprognose verwendet werden. Eine Validierung dieser Methode kann bereits bei der Profilbestimmung der professionellen

Athleten erfolgen. Sollten sich Leistungssportler und Amateure nicht schon anhand ihrer Merkmalsprofile korrekt bestimmen und ihrer Leistungsgruppe zuordnen lassen, scheint das Vorgehen auch im Jugendbereich nicht zuverlässig. Untersuchungen von Hohmann et al. (2015) zeigen für solche „Quasi"-Prognosen je nach Sportart trotz einer geringen Stichprobengröße hohe Erfolgsaussichten – 89,3% der Prognosen sind korrekt. Zur Unterscheidung von Leistungsgruppen (Kaderathleten) scheinen solche Sportartprofile und zugehörige Gewichtungen daher geeignet zu sein. Fraglich ist noch, ob dieses Vorgehen auch in einer frühen Talentidentifikation anwendbar ist. Geht man von der Annahme aus, dass sich die Profile von Nachwuchssportlern und Leistungssportlern deutlich unterscheiden, kann wohl nur eine retrospektive Studie früherer Leistungsprofile und evtl. auch der Leistungsverläufe heute erfolgreicher Leistungssportler Aufschluss über mögliche Prädiktoren bzw. vielversprechende Nachwuchs-Merkmalsprofile geben.
Eine weitere, wenngleich triviale Methode, Prädiktoren zu erhalten, stellt der Vorgang „Trial-and-Error" dar. Hierbei werden beliebige Merkmale als mögliche Prädiktoren erprobt und ihr Einfluss auf die (spätere) Wettkampfleistung nachfolgend (retrospektiv) bestimmt.

Der zweite Schritt der Gewinnung aussagekräftiger Prädiktoren umfasst die Methodenentwicklung, um die zuvor ausgewählten Merkmale zu erheben. Bei der Suche nach einer geeigneten Testauswahl steht man jedoch häufig dem *Generalitäts-Spezifitäts-Dilemma* gegenüber (Hohmann et al., 2015). Es besagt, dass spezifische Merkmale einer Sportart auch immer ein spezifisches Messinstrument benötigen, um adäquat abgebildet werden zu können. Werden die Messverfahren jedoch zunehmend spezifischer, so neigen sie dazu, nur noch eingeschränkte Aussagen über eigentlich ähnliche Merkmale, Situationen oder Sportarten zuzulassen – dies ist vor allem ein Problem der Talent- bzw. Sportartorientierung. So kann bspw. ein Torschusstest im Fußball unter Belastungsdruck und Gegnereinwirkung die Situation eines Stürmers darstellen, in der 85. Minute noch ein Tor schießen zu wollen, der Test deckt jedoch weder Freistöße noch Kopfbälle ab – ganz zu schweigen von Sportarten, die keine Schuss-Vorgänge beinhalten. Ein zu generelles Messverfahren hingegen mag die tatsächlichen, zumeist spezifischen Situationen einer Sportart nicht mehr wirklich abzubilden. Will man dieses Problem umgehen und in Testungen allen spezifischen Situationen einer Sportart gerecht werden, wird die gesamte Testbatterie schnell sehr groß und ist damit nicht mehr ökonomisch vertretbar. „Ein weitgehend akzeptierter Lösungsvorschlag besteht darin, in den frühen juvenilen Stadien eher eine allgemeine motorische Begabung zu testen und dann schrittweise auf sportartgruppen- und später auch sportartspezifische Verfahren umzustellen" (Hohmann et al., 2015, S. 10). Ein mit dem Generalitäts-Spezifitäts-Dilemma verknüpftes Problem ergibt sich auch hinsichtlich der Gütekriterien, da die beiden Hauptgütekriterien Validität und Reliabilität teilweise in einem komplementären Verhältnis zueinander stehen (Abb. 6; Ferrauti, 2020).

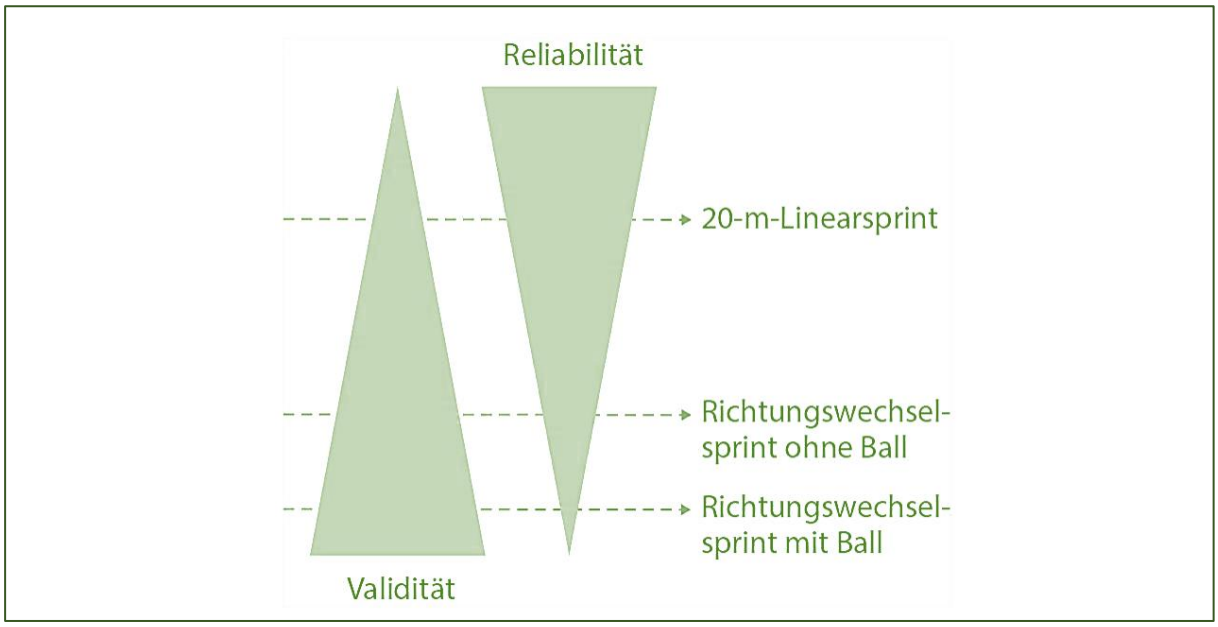

Abb. 6. Komplementäres Verhältnis zwischen Reliabilität und Validität (Ferrauti et al., 2020, S. 105)

Entsprechend dieses Verhältnisses weisen unspezifische Tests wie bspw. der 20m-Sprint eine hohe Reliabilität auf, sind jedoch in Bezug auf das eigentliche Talentkriterium eher als schwach einzustufen (Ferrauti et al., 2020). Ganz anders verhält es sich mit einem spezifischen Test zu Richtungswechselsprints mit Ball. Die Validität dieses Fußballtests kann hier höher eingeschätzt werden, die Reliabilität sinkt jedoch ab (Ferrauti et al., 2020). Ein Beispiel für einen Test mit akzeptabler Reliabilität und vermeintlich schwacher Validität ist der *Loughborough Soccer Passing Test*, der häufig für die Erhebung der Passqualität im Fußball genutzt wurde (Ali et al., 2007; Huijgen et al., 2013). Untersuchungen von Serpiello et al. (2017) zeigen einen nur geringen Zusammenhang dieses Tests mit der eigentlich implizierten Passqualität während eines Spiels (N = 24; U18; Pearson-Korrelation: $0{,}3 \leq r \leq 0{,}47$). Eine hohe Validität scheint daher fraglich (Bergkamp et al., 2019). Trotz der angesprochenen Problematik sollten Messverfahren insgesamt in einem hohen Grad den Gütekriterien entsprechen, um für eine Talentidentifikation geeignet zu sein (vgl. Kapitel 2.2).

Als dritte Phase zur Gewinnung aussagekräftiger Prädiktoren wird die Prüfung des Zusammenhangs von Prädiktor und Kriterium genannt (Gabler & Ruoff, 1979). Zu den statistischen Methoden zur Prüfung dieser Zusammenhänge bzw. der Kriteriumsvalidität werden die Korrelations- und die Regressionsanalyse gezählt. Untersuchungen von Kramer et al. (2017) zeigten bspw. mithilfe einer Pearson-Korrelation für 92 erfolgreiche Tennisspieler der U13, dass für männliche Probanden der Prädiktor „Kraft im oberen Körperbereich“ ($r = -0{,}5$) signifikant mit der bisher erreichten Ranglistenposition (Kriterium) zusammenhängt. Eine anschließende Regressionsanalyse wies nach, dass dieser Prädiktor bereits 25% der Varianz aufklären konnte. Für Mädchen wurde ein signifikanter Zusammenhang mit der Ranglistenplatzierung für die Prädiktoren „age peak

height velocity“ (APHV; $r = 0{,}383$) und „Kraft im unteren Körperbereich“ ($r = -0{,}323$) nachgewiesen. Die durch APHV aufgeklärte Varianz lag im Bereich von etwa 15%. APHV korrelierte auch drei Jahre später (U16) immer noch signifikant mit der Ranglistenposition ($r = -0{,}355$; $R^2 = 0{,}126$). „Man beachte, dass $r = \pm 0{,}4$ auf eine relativ hohe prädiktive Validität hindeutet, insbesondere in Anbetracht der Komplexität der Vorhersage eines Leistungsergebnisses bei jungen Spielern, das mehrere Jahre in der Zukunft vom Testzeitpunkt liegt“ (Bergkamp et al., 2019, S. 1329). Auch andere Autoren (Faber et al., 2014; Faber et al., 2015; Filipcic, 1999; Filipcic et al., 2010; Girard & Millet, 2009; Ulbricht et al., 2016) konnten mittels einer Korrelationsanalyse signifikante Zusammenhänge einzelner Tests und der Kriteriumsleistung nachweisen.

Anhand von Regressionsanalysen wurde ebenfalls in mehreren Studien eine Verbindung von Prädiktoren und einem Kriterium gezeigt (Kondric & Filipcic, 2009; Leyhr et al., 2020; Meckel et al., 2015; Sieghartsleitner et al., 2017). Höner et al. (2017) betrachteten bspw. in einer prospektiven Kohortenstudie den Zusammenhang der zwei latenten Leistungsfaktoren „speed ability“ und „technical skills“ der U12 (Prädiktoren) mit dem ca. neun Jahre später erreichten Erfolg im Fußball (Kriterium). Die mithilfe einer logistischen Regression berechneten *Odds Ratios* (OR; übersetzt Chancenverhältnis) von $OR_{speed\ abilty} = 4{,}58$ und $OR_{technical\ skills} = 6{,}72$ zeigten einen signifikanten Einfluss der beiden Faktorvariablen auf die spätere Spielleistung. Stieg die Faktorvariable „speed ability“ um eine Standardabweichung (SD = 0,09) an, so erhöhte sich die Chance für das Erreichen des späteren Spitzenniveaus um den Faktor 1,15 ($= 4{,}58^{0{,}09}$) und für die Faktorvariable „techical skill“ sogar um 2,59 ($= 6{,}72^{0{,}5}$).

Einige Studien (Höner et al., 2019; Mostaert et al., 2020; Till et al., 2015) prüften ebenfalls anhand einer *analysis of variance* (ANOVA), ob und gegebenenfalls wie sich die Testergebnisse in verschiedenen Leistungsgruppen unterscheiden. Tests mit hoher „Trennschärfe“ (vgl. Effektgröße) können dann wiederum in der Talentidentifikation genutzt werden. Höner et al. (2019) untersuchten bspw. mithilfe von ANOVAs signifikante Mittelwertunterschiede zwischen den motorischen Leistungen der Teilnehmer der U12 und ihrer Auswahlstufe (Kriterium) der U17. Für alle getesteten Variablen konnten signifikante ($p \leq 0{,}5$) Ergebnisse gefunden werden. Während sich für Sprint, Beweglichkeit und Ballkontrolle nur kleine Effektgrößen ergaben ($0{,}04 \leq \eta^2 \leq 0{,}05$), zeigten sich für Dribbling ($\eta^2 = 0{,}10$) und den Gesamtscore ($\eta^2 = 0{,}12$) mittlere Effektgrößen (Höner et al., 2019).

Generell gilt für die meisten Studien, dass häufig nur kleine bis mittlere Effekte für einzelne Tests nachgewiesen werden können (Bergkamp et al., 2019). Murr et al. (2018) analysierten in einer Review-Studie (mit insgesamt neun Einzelstudien) die Effektgrößen einzelner Tests im Fußball. Im Mittel lagen die Effektgrößen der Tests der untersuchten Studien (mit einer Ausnahme: $d_{Repeated\ Sprint\ Ability} = 1{,}21$) zwischen $d = 0{,}23$ (kleiner Effekt; Prädiktor Körpergewicht) und $d = 0{,}57$ (mittlerer Effekt; Prädiktor Sprint <20m). Kleine Effektgrößen einzelner Prädiktoren sprechen jedoch nicht zwangsweise gleich gegen eine erfolgreiche Talentidentifikation, da die letztendlich

wichtige Talentprognose zumeist nicht nur auf der Grundlage eines, sondern auf der Gesamtheit aller relevanten Prädiktoren erstellt wird.

Die vierte und letzte Phase des Modells zur Gewinnung von aussagekräftigen Diagnosekriterien (Gabler & Ruoff, 1979) umfasst die Bewährung des zuvor geprüften Prädiktors. In diesem Verfahren sollen die Prädiktoren erneut anhand anderer Stichproben und Zeitpunkte überprüft werden, um die Aufnahme von zufällig stichproben- oder epochenspezifischen Merkmalen zu verhindern (Hohmann, 2009). Dieser Schritt mag auf den ersten Blick als unnötige Nachkontrolle der eigentlich schon bewiesenen Befunde erscheinen, jedoch zeigen mehrere Studien an ein und demselben Prädiktor teilweise unterschiedliche Ergebnisse. In der bereits angesprochenen Review-Studien von Murr et al. (2018) wurde sichtbar, dass die Effektstärken des Prädiktors Körpergewicht für unterschiedliche Altersstufen (U10–U15) stark schwankten und signifikante Werte von $-0{,}38 \leq d \leq 1{,}23$ annahmen. Auch die Effektstärken der anderen untersuchten Prädiktoren wiesen ähnliche Schwankungen auf (Murr et al., 2018). Aus diesem Grund sind Ergebnisse vorerst immer in Abhängigkeit zur gewählten Stichprobe zu sehen.

3.2 Auswahl der Stichprobe (Range)

Die Auswahl der Stichprobe stellt eine entscheidende Einflussgröße in der Talentidentifikation dar, da sich die gefundenen Ergebnisse bspw. von Prädiktor und Kriterium zunächst vorrangig nur auf diese Probanden beziehen. Nach Bergkamp et al. (2019) könnte daher in manchen Studien der Eindruck entstehen, dass bestimmte Prädiktoren für eine allgemeine Talentidentifikation nicht geeignet seien, obwohl diese in anderen Studien mit anderen Probandengruppen signifikante Einflüsse auf die Wettkampfleistung zeigen würden.
Das Problem einer verfälschten Verallgemeinerung (vgl. externe Validität) von Befunden tritt häufiger für relativ homogene Gruppen der Talentselektion auf, da die Probanden hier oft eine Vorselektion erfahren haben. So könnten bspw. die Sportler für Kadergruppen anhand bestimmter Merkmale ausgewählt worden sein, die in den Augen der Landes- und Bundestrainer für wichtig erachtet werden, und es ist dementsprechend nicht verwunderlich, wenn gerade diese Merkmale in einer Talentselektion aufgrund der geringen Varianz nicht mehr signifikant auffallen (Abb. 7). Die gleichen Merkmale könnten in einer heterogeneren Gruppe noch einen signifikanten Einfluss oder auch einfach größere Effektstärken der Prädiktoren auf das Kriterium zeigen (Bergkamp et al., 2019; Williams & Reilly, 2000).

In Abbildung 7 sind die Streudiagramme der Werte eines Prädiktors und eines Kriteriums zu sehen. Aufgezeigt werden sowohl die ursprüngliche Verteilung als auch die geänderte Verteilung einer bereits nach diesem Prädiktor vorselektierten Gruppe. Die ursprünglichen Korrelationswerte von $r \approx 0{,}62$ sinken im rechten Beispiel auf $r \approx 0{,}29$ ab – die Ergebnisse werden „schwächer“. Dieses Problem kann bspw. dann eintreten, wenn als Stichprobe nur Sportler gewählt werden, die bereits für eine Kader-Sichtung eingeladen wurden. Eine statistische Bestimmung der Prädiktoren für eine Prognose

erfolgreicher Kadersportler ist dann mitunter aufgrund der Vorselektion der Stichprobe nicht mehr zutreffend.

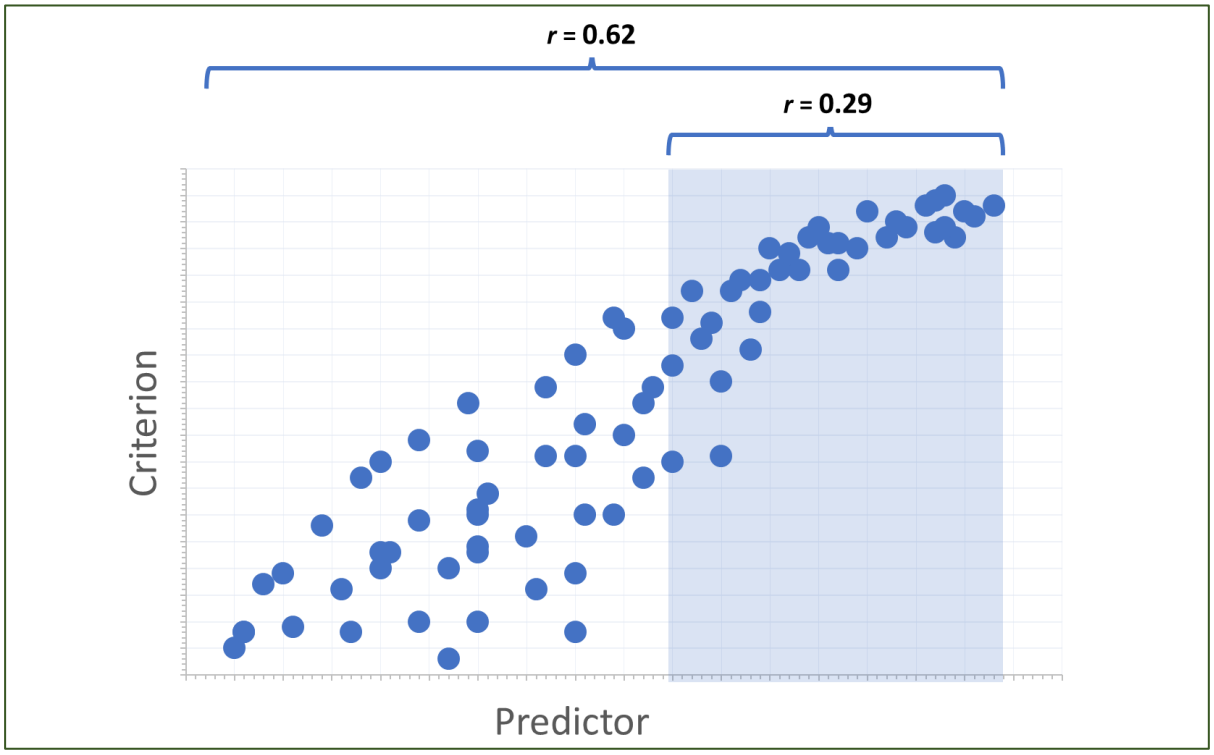

Abb. 7. Veränderung des Zusammenhangs zweier Variablen durch Vorselektion

Neben der bereits erläuterten Vorselektion der Probandengruppen kann auch die Trainingshistorie der getesteten Stichprobe für bestimmte Ergebnisse verantwortlich gemacht werden. Werden mehrere Teilnehmer vom gleichen Trainer, der gleichen Vereinsphilosophie, der gleichen Trainingsroutine oder auch nur von den gleichen sozialen Kontakten (im Verein, Freundeskreis, Internat, Kader, etc.) beeinflusst, kann sich dies wiederum auf die Talentidentifikation auswirken und Ergebnisse verzerren.
Die Schwierigkeiten im Umgang mit vorselektierten Gruppen lassen sich nur schwer umgehen, da gerade das Erreichen dieser (Leistungs-)Gruppen in der Talentidentifikation als Kriterium herangezogen wird. Bergkamp et al. (2019) empfehlen daher den Einsatz von Korrekturformeln, die die Varianzunterschiede der verschiedenen Merkmale der unterschiedlichen Gruppen (häufig Grundgesamtheit und Kriteriumsgruppe) korrigieren würden (Hunter & Schmidt, 2007; Sackett & Yang, 2000).
Auch die reine Stichprobengröße hat einen Einfluss auf das Ergebnis, da sich mit zunehmender Probandenanzahl auch kleinere Effekte noch signifikant nachweisen lassen (Rey, 2017). Wird die Stichprobe bspw. durch eine Betrachtung von nur einzelnen Gruppen z.B. höherer Leistungslevel künstlich verkleinert, so können auch vormals signifikante Ergebnisse nun insignifikant werden. Die Ergebnisse von Analysen solch kleiner Stichproben lassen sich dann auch nur schwer verallgemeinern bzw. auf andere Gruppen übertragen.

3.3 Bestimmung des Kriteriums (Criterion)

Ziel von Talentidentifikationsstudien ist es, die Ausprägungen eines gewählten Kriteriums möglichst genau vorhersagen zu können. Eine studienübergreifende Definition des Kriteriums ist jedoch bisher nicht zu erkennen. Häufig wird als Kriterium die in der (späteren) Karriere eines Sportlers erreichte persönliche Bestleistung, das erreichte Leistungslevel (*performance levels*: Ranglistenposition, Kaderplatz, Liga, Nationalmannschaft, etc.) oder die Wettkampfleistung herangezogen (Hohmann, 2009; Höner et al., 2019). Dementsprechend sind auch die Bezeichnungen und Einordnungen eines Kriteriums sehr verschieden. Man findet in der Literatur bspw. die Ausdrücke „elite, sub-elite, and non-elite level; professional, semi-professional, or non-professional level; first team or reserves; elite, club level, or dropouts; national or regional level; selected and non-selected players; and nationally drafted or non-drafted players" (Bergkamp et al., 2019, S. 1319). Hinzu kommen Begriffe wie final selection, elite cadets, high division, senior, successful, survivor oder top world (Johnston et al., 2018). In vielen der verwendeten Ausdrücke, z.B. selected und non-selected oder national und regional, wird der Versuch sichtbar, eine klare Grenze zwischen zwei Gruppen zu ziehen. Die verwendete Kriteriumsvariable liegt somit teilweise dichotom/binär (nur zwei Ausprägungen) vor. Dieses Vorgehen ist nachvollziehbar, da in der Talentidentifikation in aller Regel der Wunsch nach einer klaren Entscheidung darüber besteht, einen Probanden in eine Förderung aufzunehmen oder nicht.
In der Realität sind feste Grenzen jedoch selten. Zum einen kann nicht angenommen werden, dass alle Personen eines Leistungslevels die gleiche Leistung zeigen, zum anderen sind die Übergänge der Gruppierungen üblicherweise fließend. Mit starren Grenzen erhält man somit nur noch limitierte Informationen über die Unterschiede zwischen den einzelnen Individuen (Bergkamp et al., 2019; Höner & Votteler, 2016). Auf der individuellen Ebenen müssten, so Ackermann (2014), prognostische Aussagen „wahrscheinlichkeitstheoretisch" sein, was durch feste Grenzen ganz oder teilweise verhindert würde (Höner & Votteler, 2016). Eine verwendete Dichotomisierung wirkt sich auch auf die statistischen Berechnungen aus. So sinkt bei dichotom vorliegenden Variablen der berechnete Korrelationskoeffizient dieser Variablen im Vergleich zu metrisch vorliegenden Variablen (mitunter sogar deutlich) ab (Hunter & Schmidt, 2007). In der Praxis ist dieses Vorgehen der Talentidentifikation, die Kriteriumsvariablen anhand fester Gruppen zu unterscheiden, jedoch weit verbreitet, da sich Sensitivität und Spezifität der erstellten Talentprognose vergleichsweise einfach bestimmen lassen (Kap. 3.6).

Hat man sich trotz der genannten Vorbehalte für eine Aufteilung des Kriteriums in Gruppen bzw. Leistungslevel entschieden, stellt sich ein neues Problem. Zum einen gibt es im Sport nicht immer klar ersichtliche Gruppen, anhand derer man bestimmte Leistungslevel ableiten kann, zum anderen können womöglich gleich mehrere Gruppen für ein Leistungslevel in Betracht gezogen werden (siehe Tab. 3; Hohmann, 2009; Pfeiffer &

Hohmann, 2012). So wird in *mks*-Sportarten[7] wie bspw. der Leichtathletik die Leistung üblicherweise in metrischen Messwerten erfasst, was die Frage aufwirft, ab welchem Messwert eine Gruppe beginnt oder endet. Häufig sind Leistungen und eine dazu passende Gruppenzugehörigkeit nicht immer offensichtlich. Bei den Deutschen Leichtathletikmeisterschaften 2016 wäre man bspw. mit einer Sprintzeit von 10,48 Sekunden über 100m (Herren) auf dem zweiten Platz gelandet. Bei den Meisterschaften 2019 hätte diese Zeit jedoch nicht für den Einzug ins Finale gereicht (<10,38 Sek.). Eine passende Auswahl geeigneter Leistungslevel ist daher häufig nicht trivial.

Tab. 3. *Talentgruppierung anhand der besten jemals erreichten Wettkampfleistung im Schwimmen (nach Pfeiffer & Hohmann, 2012, S. 347)*

Final competition performance	Talent group [Criterion]
1. Medallist at European/World Championship or Olympic Games 2. Final at European/World Championship or Olympic Games 3. Medallist at German or Junior-European Championships 4. Final at German or at Junior-European Championships	3
5. Participation at German or Junior-European Championships or medallist at Junior German Championship 6. Final at Junior German Championship or medallist at federal state Championship 7. Participation Junior German Championship or final at federal state Championship	2
8. Participation at federal state Championship or medallist at Junior federal state Championship 9. Participation at Junior federal state Championship 10. Competition on county level or below	1

Eine Möglichkeit, die Gruppierungen der Leistungslevel zu gestalten, wäre eine Orientierung an den gängigen Prozentwerten der Literatur für Talente und eine Festsetzung der Leistungslevel entsprechend dieser Prozentangaben. Wählt man bspw. das Talentmodell von Gagné (2004) als Referenz, so müssten die besten 10% der Sportler als Talente angesehen werden (vgl. Kap. 2.1.2). Dieser Wert ist jedoch nicht einheitlich in der Literatur zu finden. Hohmann et al. (2020) berichten von wesentlich geringeren Prozentwerten, die für das höchste Leistungslevel gewählt werden. So werden je nach Autor teilweise zwei (Selektion der besten 2,28%; Ljach, 1998), drei (Selektion der besten 0,123%; Kovar, 1981) oder sogar vier (Selektion der besten 0,0034%; Matsudo, 1996) Standardabweichungen über dem Niveau von Normalsportlern gefordert (Hohmann et al., 2020). Auch Kunz (2007) merkt an, dass letztendlich nur 0,3% aller registrierten Fußballer für professionelle Vereine spielen. In vielen Talentstudien wird dennoch von diesen geringen Zahlen abgewichen, sodass sich Werte von 15% (Gonaus & Müller,

[7] *mks*-Sportarten sind Sportarten, die sich durch „klare" Messwerte auszeichnen. Die Abkürzung mks steht für Meter, Kilogramm und Sekunde, da die Messeinheiten in diesen Sportarten zumeist die Leistung beschreiben oder diese darstellen. In der Sportart Leichtathletik (als typisches Beispiel) wird die Leistung, bspw. im Weitsprung, an diesen festen Größen bewertet.

2012), 14,2% (Höner et al., 2019), 12,1% (Till et al., 2015), 10,7% (Le Gall et al., 2010), 6,2% (Leyhr et al., 2020) oder 0,9% (Höner & Votteler, 2016) finden. Dies liegt teilweise daran, dass das gewählte Kriterium noch im Jugendalter liegt und hier die prozentualen Anteile der Sportler der höchsten Leistungsklassen aufgrund der häufig noch nicht voll ausgebauten Ligenstruktur noch größer sind als im Erwachsenenalter. Teilweise werden diese höheren Prozentwerte aber auch in Abhängigkeit der Stichprobe gewählt – zu kleine Leistungsgruppen in bereits schon kleinen Stichproben könnten die Ergebnisse verzerren – oder diese orientieren sich an den in der Praxis vorkommenden natürlichen Gruppierungen (Kadergruppen, Wettkampfgruppen, etc.).

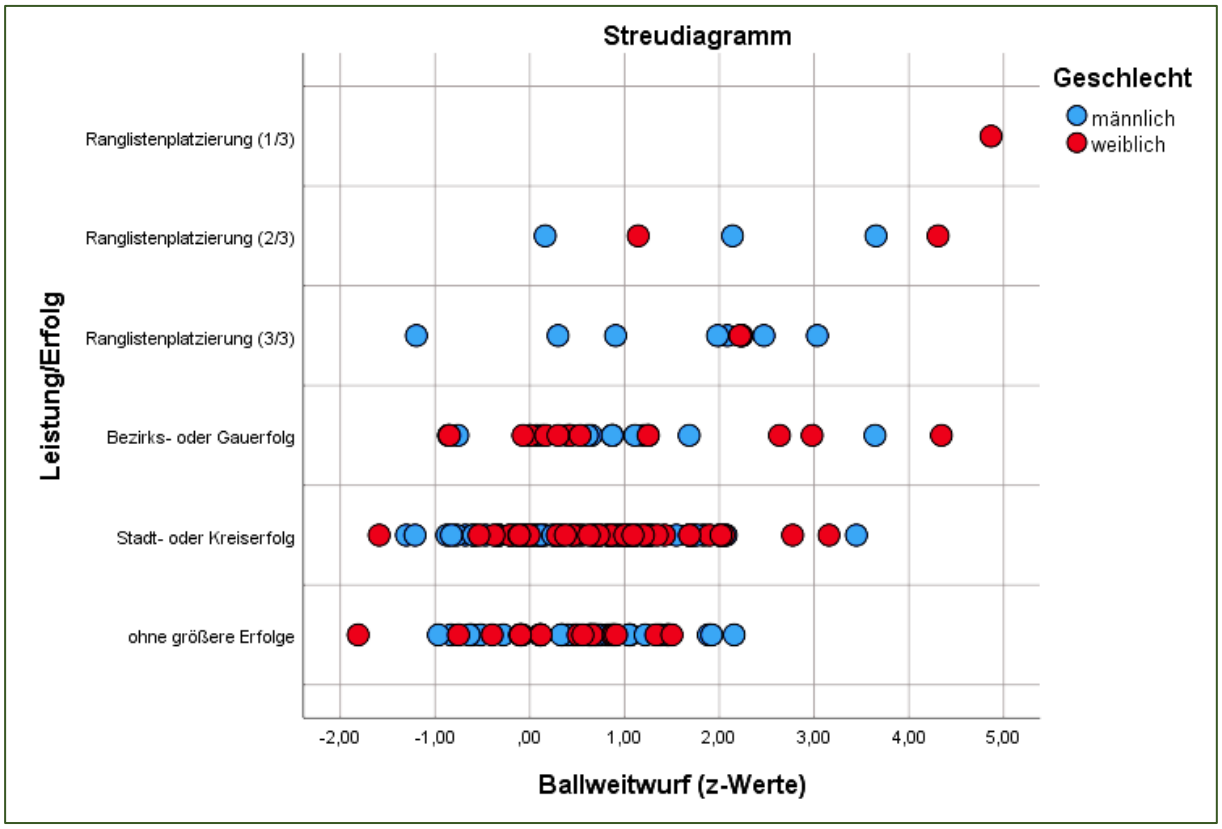

Abb. 8. Streudiagramm zur Verdeutlichung einer möglichen Level-Einteilungen im Jugendbereich der Sportart Tennis

Die Auswahl eines festen Prozentwertes scheint daher allgemein wenig sinnvoll, steigt doch mit jeder Gruppierung, die in einen Leistungslevel miteinbezogen wird, die Anzahl der Teilnehmer immer nur sprunghaft an. Vergleicht man bspw. die Gruppenzusammensetzungen der in Abbildung 8 dargestellten Sportler, fällt auf, dass in der höchsten Leistungsgruppe nur eine Person vertreten ist (0,6%), während in der darauffolgenden weitere fünf (summiert 3,4%), dann zehn (summiert 9,2%) und in der vierten Gruppierung bereits 26 (summiert 24,1%) der insgesamt 174 Sportler zu finden sind. Eine Auswahl von exakt 5% oder 15% der besten Sportler wäre demnach anhand dieser Leistungsgruppierung nicht möglich. Als grobe Orientierung werden im Jugendbereich dennoch häufig die von Gagné vorgestellten besten 10% der Sportler für das höchste

Leistungslevel herangezogen. Dieser Wert findet sich in vielen Studien zu Talentidentifikation wieder (Höner et al., 2019; Leyhr et al., 2020; Till et al., 2015).
Der Anteil an Personen der interessierenden Grundgesamtheit, die in der Lage sind, eine zufriedenstellende Leistung oder ein zufriedenstellendes Leistungslevel im Kriterium zu erreichen (d.h. der Anteil der Personen, die erfolgreich sind [wenn keine Auswahl stattfindet]), wird auch *Grundrate* (base rate; Bergkamp et al., 2019; Taylor & Russell, 1939) genannt. In einem dichotom vorliegenden Kriterium ist somit die Grundrate gleich dem Anteil an Personen des höchsten Leistungslevels an der Gesamtstichprobe – üblicherweise die Prozentzahl der Leistungssportler.

Ob ein gewähltes Kriterium und die damit verbundene Einteilung der Stichprobe in gewählte Leistungsgruppen für eine nachfolgende Analyse geeignet scheint, kann mithilfe des Mittelwertes der einzelnen Gruppen überprüft werden. Sollten bspw. durch einen *t*-Test oder eine ANOVA keine signifikanten Gruppenunterschiede der Leistungsgruppen anhand verschiedener Prädiktoren erkennbar sein, so müsste (neben der Prädiktorauswahl) wohl auch die Einteilung des Kriteriums überdacht werden.

Neben der Auswahl der passenden Leistungsgruppierungen für ein Kriterium kann allerdings auch die Gruppierungsvariable selbst schon fehlerbehaftet sein. So dient eine Gruppierung nach Wettkampferfolgen im Jugendbereich nicht zwangsweise als Kriterium für das Erwachsenenalter. Wie das Beispiel der Wettkampferfolge von sechs späteren britischen Goldmedaillengewinnern zeigt, schwankten die zuvor erreichten Wettkampferfolge mitunter stark (Abb. 9). Würde man bspw. die maximale jemals erreichte Leistung zum Ende des sportartspezifischen Jugendalters (Ordinate) als Kriterium nehmen, würden nur drei (die gepunkteten Fälle) der sechs späteren Spitzensportler in die Kriteriumsvariable mit aufgenommen. Wenn möglich sollte sich daher das Kriterium auf die abschließende Spitzensportleistung des sportartspezifischen Höchstleistungsalters beziehen. Hohmann (2009) empfiehlt deshalb Langzeitstudien für die Talentidentifikation.

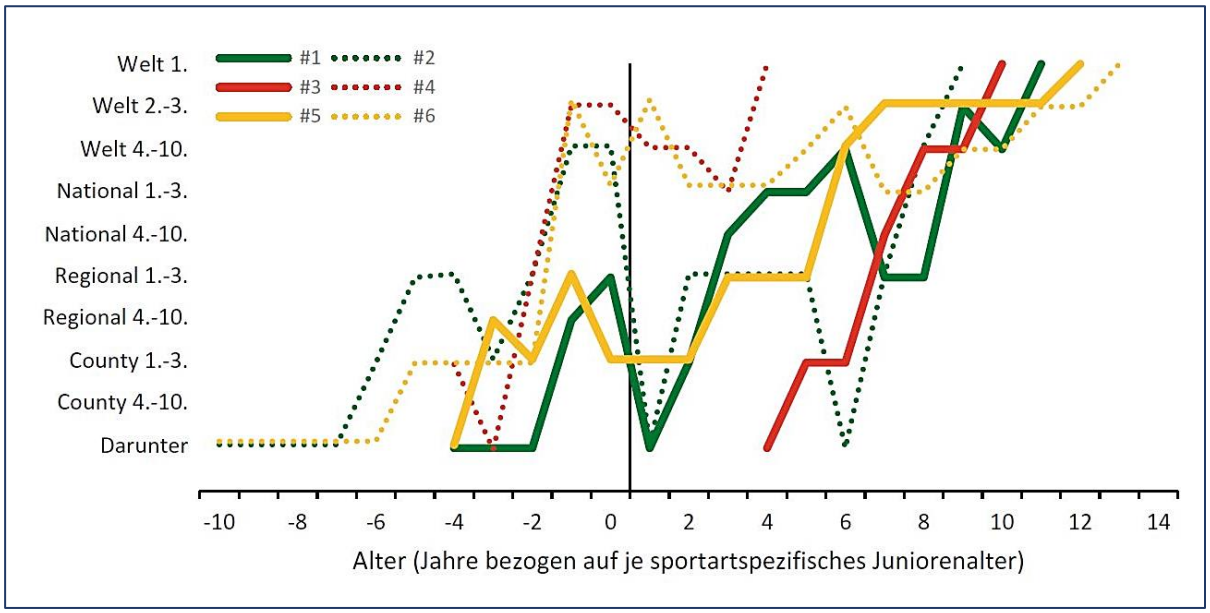

Abb. 9. Entwicklung der Wettkampferfolge von sechs britischen Goldmedaillengewinnern im Zeitraum bis zu ihrer ersten Goldmedaille (die Ordinate markiert das Ende des sportartspezifischen Jugendalters; Güllich, 2019)

3.4 Bestimmung der Selektionsrate (Selection Ratio)

Die Selektionsrate (selection ratio) oder auch Empfehlungsquote gibt den Anteil der Spieler in der interessierenden Grundgesamtheit an, der für eine positive Prognose bzgl. des Kriteriums ausgewählt wird – sozusagen wie viele Empfehlungen im Vergleich zur Gesamtempfehlungszahl für das Erreichen des entsprechenden Leistungslevels ausgesprochen werden. Nimmt man bspw. das Erreichen einer Kaderplatzierung (dichotom) als Kriterium, dann entspricht die Selektionsrate der Prozentzahl an ausgesprochenen Kaderplatzempfehlungen und die Grundrate der Anzahl an tatsächlich vorhandenen Kaderplätzen (Prävalenz). An diesem Beispiel wird bereits deutlich, dass Selektionsrate und Grundrate nicht immer gleich groß sein müssen. Teilweise macht es sogar Sinn, die Selektionsrate größer als die Grundrate zu wählen, denn man geht davon aus, dass nicht alle Vorhersagen exakt zutreffen. Durch solche „Überempfehlungen“ können demnach einerseits eine größere Anzahl der später tatsächlich erfolgreichen Sportler durch die Prognosen korrekt vorhergesagt werden – die Sensitivität wird gesteigert –, andererseits sinkt aber auch die Spezifität, die Anzahl der korrekt vorhergesagten nicht-erfolgreichen Sportler (vgl. Kap. 3.7).

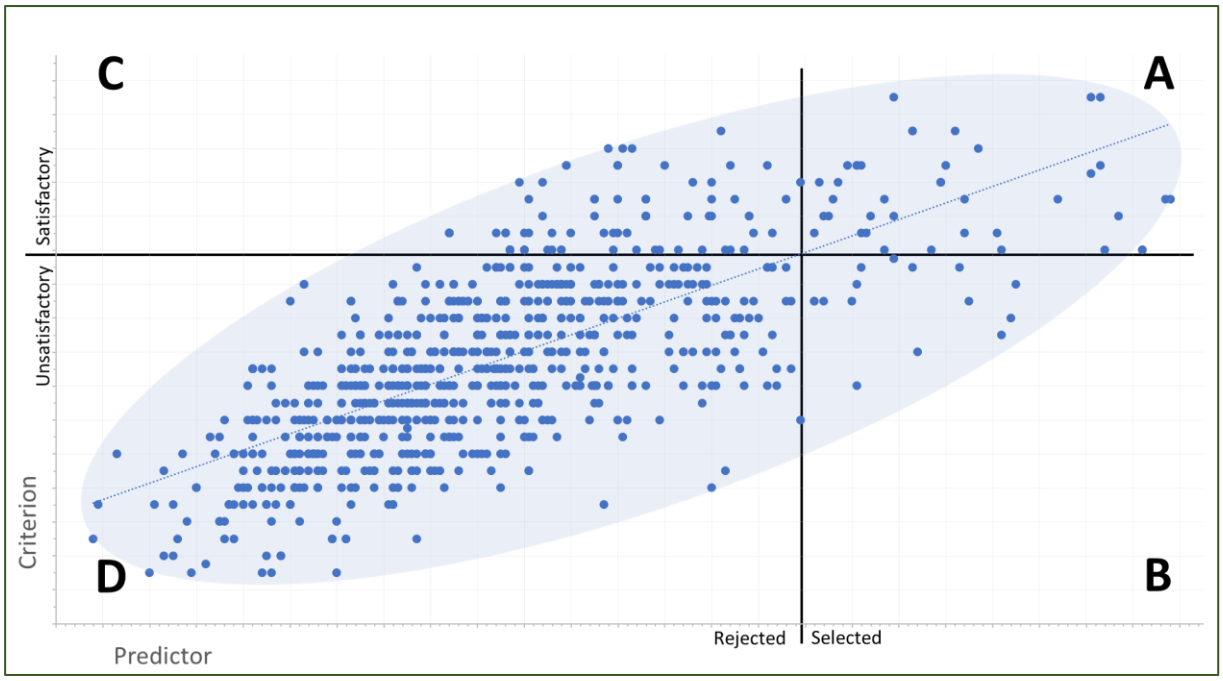

Abb. 10. Vier-Felder-Streudiagramm der Talentidentifikation

Abbildung 10 zeigt eine Vier-Felder-Grafik, die den Zusammenhang von Prädiktor und Kriterium verdeutlicht. Im Beispiel wird deutlich, dass der Prädiktor mit dem Kriterium positiv korreliert Die mit *A* bezeichneten Fälle stellen hierbei die erwarteten „Höchstleister" dar, die mit *D* bezeichneten Fälle die erwarteten „Niedrigleister" (Bezeichnungen nach Hohmann et al., 2020). Bezogen auf die Gesamtstichprobe bildet folglich der Anteil der Bereiche *A* und *C* die Grundrate ab und der Anteil der Bereich *A* und *B* die Selektionsrate. Die Ordinate markiert dabei den (Trenn-)Wert der Prädiktorwerte, ab denen eine positive Höchstleister-Prognose gestellt wird. Die Abszisse verkörpert jenen (Trenn-)Wert der Kriteriumsvariablen, ab dem eine Person als Höchstleister angesehen wird.

Die Auswahl der Selektionsrate (vgl. Ordinate) hat einen entscheidenden Einfluss auf die spätere Sensitivität und Spezifität der Talentprognosen und sollte daher mit Bedacht gewählt werden. Insgesamt können hier verschieden Überlegungen im Raum stehen: Zunächst könnte die Selektionsrate genauso hoch gewählt werden wie die Grundrate (Prävalenz). Dieses Vorgehen kann bei einem „sehr guten" Prädiktor von Vorteil sein, der eine hohe Korrelation mit dem Kriterium aufweist, da so keine oder nur sehr wenige Über- oder Unterempfehlungen entstehen (es werden mehr/weniger Empfehlungen ausgesprochen als Spitzensportler beobachtet wurden) und sowohl Sensitivität als auch Spezifität der Talentprognose maximale Werte erzielen können (vgl. Kap. 3.7).

Eine zweite Möglichkeit wäre es, die Selektionsrate von den Fördermaßnahmen abhängig zu machen. Reichen bspw. die finanziellen Mittel eines Vereins aus, um 15% der talentiertesten Kinder entsprechend zu fördern, kann es Sinn machen, so viele Kinder auch als spätere Höchstleister zu empfehlen (Selektionsrate = 15%).

Eine weitere Alternative wäre es, den Schnittpunkt von Bezugslinie (zumeist die Diagonale) und Grundrate-Linie (Abszisse) als Ansatzpunkt für die Selektionsrate zu

wählen (siehe Abb. 10). Dieses Vorgehen der Findung einer Selektionsrate berücksichtigt verstärkt die eigentliche Eignung der empfohlenen Sportler.
Als mögliche statistische Verfahren, im Nachgang einer Prognoseberechnung retrospektiv eine geeignete Selektionsrate zu ermitteln, kommen ROC-Analysen (Köster et al., 2021; Mandrekar, 2010; Unal, 2017) und die Berechnung von Youdens *J* bzw. der Youden-Index (Hedderich & Sachs, 2016) zum Einsatz (Kap. 3.7).

Das bereits angesprochene Problem von Überempfehlungen (es werden mehr Höchstleisterempfehlungen ausgesprochen, als es tatsächliche Höchstleister gibt) und Unterempfehlungen (weniger Höchstleisterempfehlungen als tatsächliche Höchstleister), wird durch eine Abweichung der Selektionsrate von der Grundrate verursacht. Taylor und Russell haben die Auswirkungen der verschiedenen Selektionsraten auf den positiven prädiktiven Wert (den Anteil an korrekten Höchstleisterempfehlungen unter allen Höchstleisterempfehlungen) bereits 1939 untersucht (Abb. 11).

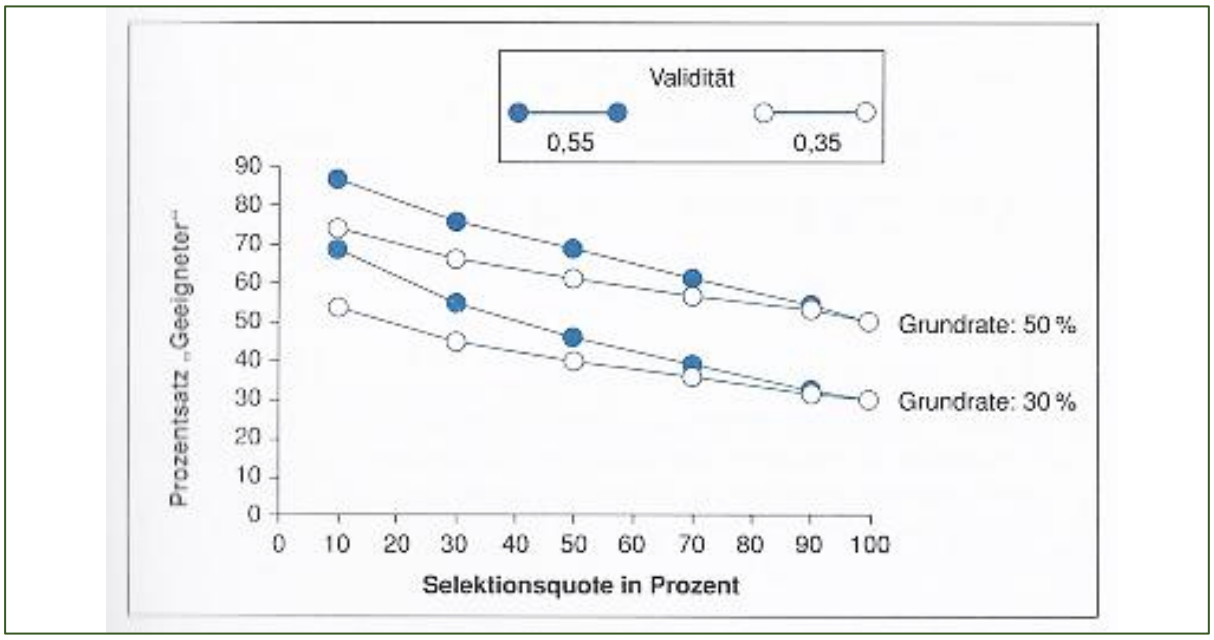

Abb. 11. Positiver prädiktiver Wert (Prozentsatz „Geeigneter") in Abhängigkeiten von der Validität des Prädiktors (Korrelation: $r = 0{,}55$ und $r = 0{,}35$), der Selektionsrate/-quote und der Grundrate (nach den Taylor-Russell-Tafeln von Taylor & Russell, 1939; Lienert & Raatz, 1998, S. 421)

Es zeigt sich, dass der positive prädiktive Wert (Prozentsatz „Geeigneter") mit steigender Selektionsrate abnimmt. Würde man bspw. bei einer Grundrate von 50% – d.h. in der Stichprobe erfüllt jede zweite Person das gesuchte Leistungslevel (Kaderplatzierung, Wettkampferfolg, etc.) – und einem moderaten Prädiktor (Korrelation: $r = 0{,}35$) eine Selektionsquote von 10% wählen, wären 74% der aufgestellten Höchstleisterempfehlungen korrekt. Bei einer Selektionsquote von 30% wären es nur noch 66% und bei einer Selektionsrate von 50 % nur noch 61% korrekte Höchstleisterempfehlungen (siehe Taylor-Russell-Tafeln). Spräche man jedem Probanden eine Höchstleisterempfehlung aus (Selektionsrate von 100%), so hätte man zu 50% korrekt getippt, da in der

Population 50% der Personen (spätere) Höchstleister sind (vgl. Grundrate von 50%) – dies entspricht folglich einem zufälligen Ergebnis (Taylor & Russell, 1939). Wird von einer Talentidentifikation eine hohe Treffsicherheit ihrer Höchstleisterprognosen gefordert, so sollte demnach die Selektionsrate vergleichsweise klein gewählt werden.

In bestimmten Situationen kann es Sinn machen, von einer festen Selektionsquote abzusehen und stattdessen zunächst einen geeigneten *Cut-Off-Wert* zu suchen. Dies ist dann der Fall, wenn die Grundrate unbekannt ist oder die Anzahl und Art der Fördermaßnahmen noch nicht feststehen. Der Cut-off-Wert gibt dabei an, ab welchem in einer Prognosemethode erzielten *Eignungswert* (siehe Kap. 3.5) ein Proband einer bestimmten Leistungsgruppe zugeordnet werden soll. Die prozentuale Anzahl der Empfehlungen (Selektionsquote) ist somit nicht im Vorhinein festgesetzt, sondern wird in Abhängigkeit von der jeweiligen „Eignung" der Probanden bestimmt. Wird bspw. der Cut-off-Wert eines statistischen Prognosewertes (Intervall: [0;1]) auf 0,9 gesetzt, bedeutet dies, dass alle Personen mit einem Prognosewert von 0,9 und größer als spätere Talente vorhergesagt werden. Die genaue Anzahl der Vorhersagen wird somit erst nach der Berechnung der individuellen Eignung der Probanden deutlich und kann mitunter je nach Prognosemethode stark von der tatsächlich gegebenen Grundrate abweichen. Dieses Vorgehen ist dann vielversprechend, wenn explorativ vorgegangen wird und bspw. der Bedarf an Kaderplätzen erst ermittelt werden soll. Cut-Off-Wert und Selektionsquote hängen hierbei zwar zusammen, können jedoch je nach Verteilung der Eignungswerte innerhalb der Stichprobe deutlich voneinander abweichen. Erreichen bspw. nur drei von 100 Sportlern in einer Prognose einen Eignungswert von mindestens 0,9 (Cut-off-Wert), so liegt die Selektionsrate nur bei 3% und nicht bei einem zumeist erwarteten Anteil von 10%.

3.5 Bestimmung der Prognosemethode (Methods)

Die Talentprognose beruht meistens auf einem Prognose- oder Eignungswert, der bspw. die Wahrscheinlichkeit der für das Kriterium gewählten Gruppenzugehörigkeiten angibt (Singer, 1981). Anhand dieses Prognosewertes und der gewählten Selektionsrate wird dann die vorhergesagte Gruppenzugehörigkeit bestimmt. Wird bspw. für eine Kaderplatznominierung eine Anzahl von zehn Personen gesucht, würde man diejenigen Personen mit den zehn höchsten Prognosewerten für die Kaderplatznominierung empfehlen (Talentempfehlung).

Für die Berechnung eines Prognosewertes ergeben sich mehrere Möglichkeiten. Zum einen kann der Prädiktorwert bzw. die Merkmalsausprägung selbst als Prognosewert angesehen werden (z.B. bei Trainermeinungen/-einschätzungen). Entsprechend der gewählten Selektionsrate können demnach die Personen mit den höchsten Ausprägungen im Prädiktor als Talente empfohlen werden. Eine gängige Methode ist es, die Prädiktoren dabei zunächst zu standardisieren (*z*-Werte, Prozentrang, etc.), um sie als alters- und geschlechtsbereinigte Werte vorliegen zu haben (Hohmann, 2009).

Zum anderen können Prädiktoren verrechnet oder mit anderen Werten z.B. linear kombiniert werden. Für einen einzelnen Prädiktor kann dies durch eine Multiplikation mit einem Vorfaktor a_1 (z.B. einen Regressionskoeffizienten) und Addition mit einer Konstanten a_0 geschehen. Für mehrere Prädiktoren p_1, p_2, ..., p_n lässt sich analog verfahren (Formel 1):

$$Prognosewert = a_0 + \sum_{k=1}^{n} a_k\, p_k \tag{1}$$

In der Formel stellt a_0 die Konstante und die Menge der a_k die verschiedenen Vorfaktoren (z.B. Regressionskoeffizienten) der Prädiktoren p_k dar. Neben klassischen Methoden, wie die Bildung von Summe und Mittelwert, können so je nach Studie und Prädiktoren ganz verschiedene Berechnungsweisen, sogenannte Gewichtungen, entstehen (Hohmann et al., 2020). Höner und Votteler (2016) untersuchten bspw. die Eignung von über 22.000 Kindern der U12 mithilfe einer fußballspezifischen Gewichtung aus Sprint, Agilität, Dribbling, Ballkontrolle und Schussfähigkeit (siehe auch Leyhr et al., 2020; Formel 2 & 3). Solche Gewichtungen können durch die Meinung von Experten zustande kommen, aus Sportartprofilen abgeleitet (Abb. 5) oder anhand von statistischen Analyseverfahren berechnet werden (Hohmann et al., 2015).

$$Score = \frac{10.000}{(17{,}29 * Sprint) + (9{,}43 * Agility) + (4{,}11 * Dribbling) + (2{,}41 * Ball\ Control) + Shooting} \tag{2}$$

$$\Rightarrow Prognosewert := \frac{10.000}{Score} = 17{,}29 * Sprint + 9{,}43 * Agility + 4{,}11 * Dribbling + 2{,}41 * Ball\ Control + Shooting \tag{3}$$

Neben den aus Trainer- und Expertenmeinungen entstehenden linearen Gewichtungen werden weitere Klassifikationsmethoden verwendet, um die Zugehörigkeit eines Probanden zu verschiedenen Leistungsgruppen zu prüfen. Eine lineare Berechnungsweise des Prognosewertes ergibt sich bspw. anhand einer linearen Diskriminanzanalyse (Singer, 1981). Teunissen et al. (2021) untersuchten bspw. in ihrer Studie zum Talenttransfer, ob sich die Gruppe der Kanu- und Kajakfahrer von der Gruppe anderer Sportler unterscheidet. Das Streudiagramm (Abb. 12) zeigt die Koeffizienten der kanonischen Diskriminanzfunktionen auf den Achsen. Die beiden Gruppen sind innerhalb des zweidimensionalen Raums sichtbar getrennt, was darauf hinweist, dass die Funktionen zwischen Kanu/Kajak und 18 anderen Sportarten diskriminieren (Teunissen et al., 2021). Anhand der Lage der einzelnen Diskriminanzwerte in Abhängigkeit vom jeweiligen Gruppenzentroiden lassen sich die gesuchten Prognosewerte berechnen. Auch in Studien anderer Autoren wurden Diskriminanzanalysen im Rahmen der Talentidentifikation erfolgreich eingesetzt (Dugdale et al., 2019; Mostaert et al., 2021).

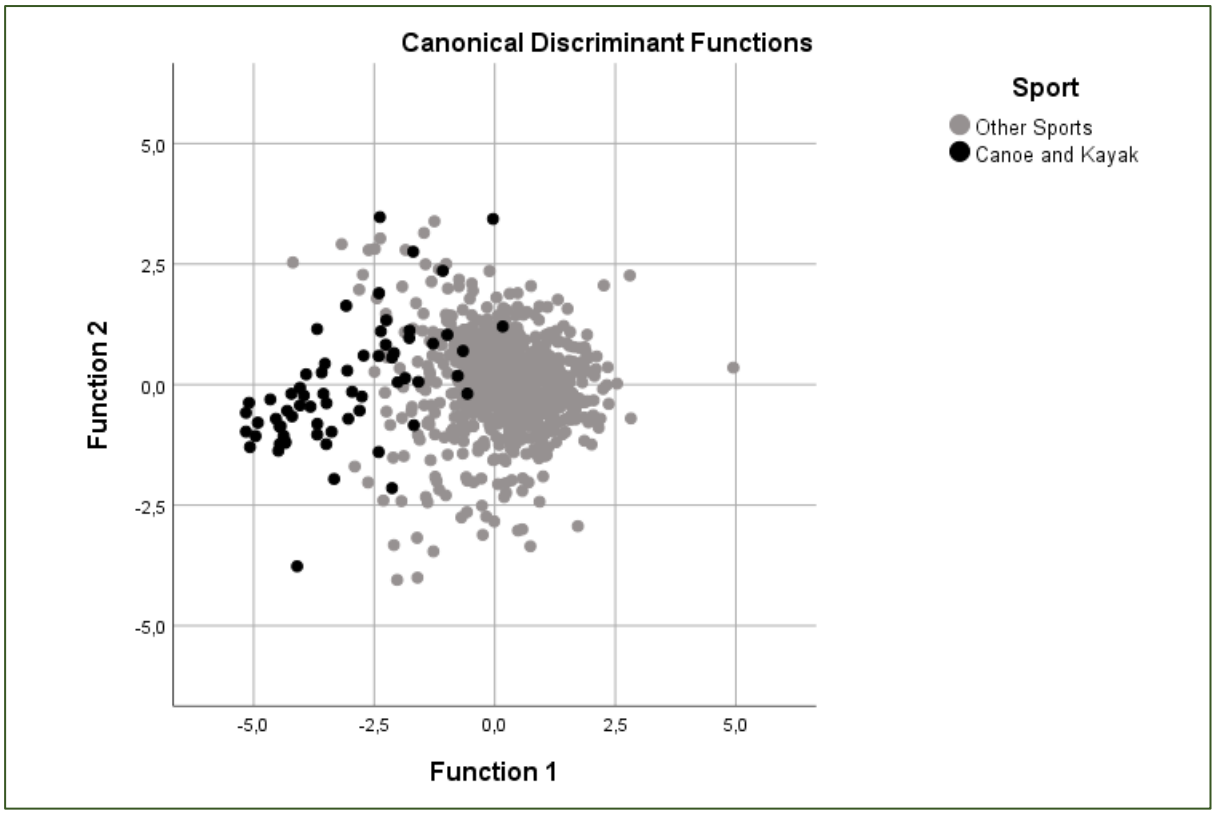

Abb. 12. Diskriminanzanalyse zur Untersuchung der Gruppe Kajak/Kanu und achtzehn anderen Sportarten (Teunissen et al., 2021)

Auch Regressionsanalysen werden zur Erstellung eines Prognosewertes verwendet (Forsman et al., 2016; S. Robertson et al., 2015; Sieghartsleitner et al., 2017). Hierbei wird ebenfalls versucht, die abhängige Variable (hier: die Kriteriumsvariable) durch eine oder mehrere unabhängige Variablen (hier: die Prädiktoren) zu erklären. Die dadurch erhaltene Gleichung kann wiederum genutzt werden, um einen Prognosewert zu bestimmen. Neben der linearen Regressionsanalyse kommt vor allem bei dichotom vorliegenden Kriteriumsvariablen die binomiale logistische Regression (Logit-Modell) zum Einsatz (Fahrmeir et al., 2009). In einer Langzeitstudie von Faber et al. (2017) wurde bspw. mittels einer linearen und einer logistischen Regression an 1.191 Tischtennisspielern geprüft, ob die im Alter von 7 bis 10 Jahren erhobenen Testwerte der Testaufgaben „sprint“, „agility“, „speed while dribbling“ und „ball throwing“ die spätere Tischtennisleistung der U13, U15 und U18 vorhersagen können.

Neben den drei beschriebenen Ansätzen gibt es noch eine Vielzahl anderer (Klassifikations-)Methoden, die es ermöglichen, Prognosewerte zu berechnen (Robertson, et al., 2015). Zu den bekannteren Methoden zählen *Support Vector Machines, Nearest Neighbour Klassifikation, Decision Trees* und *Fuzzy*-Klassifikationen (Giles et al., 2020). Pfeiffer und Hohmann (2012) ergänzen diese Aufzählung zudem noch um *künstliche neuronale Netze* (Tab. 4).

Tab. 4. *Ansätze zur Prognosewerterstellung in der Talentidentifikation (Pfeiffer & Hohmann, 2012)*

	Linear methods	**Nonlinear methods**
Static approach	Regression analyses, disciminant analyses	Neural networks
Dynamic approach	Dynamic testing, regression analyses of time series (e.g., cross-correlation)	Differential equation systems, computer simulations

Aufgrund ihrer Flexibilität und ihres großen Einsatzbereichs werden künstliche neuronale Netze in den letzten Jahren verstärkt in der Talentforschung eingesetzt (Mann et al., 2017; Musa, Taha et al., 2019). Sie sind angelehnt an ihre biologischen Vorbilder, die Nervensysteme, und bauen auf einer Netzstruktur mit teilweise in mehreren Schichten (*layer*) liegenden künstlichen Neuronen auf. Anhand von existierenden Trainingsdaten lernen neuronale Netze meist nach der Methode „Versuch und Irrtum“, über mehrere Iterationsdurchläufe/Trainingsepochen – je nach Lernverfahren auch rekurrent – von den Input- auf die Output-Daten zu schließen (Abb. 13). Die interne Struktur des Netzes (z.B. die Anzahl der Neuronen im Hiddenlayer) kann sich daher im Vergleich zu weniger komplexen Klassifikationsverfahren, in denen die Struktur von vornherein bekannt ist, je nach Datenlage und/oder Startvektor verändern.

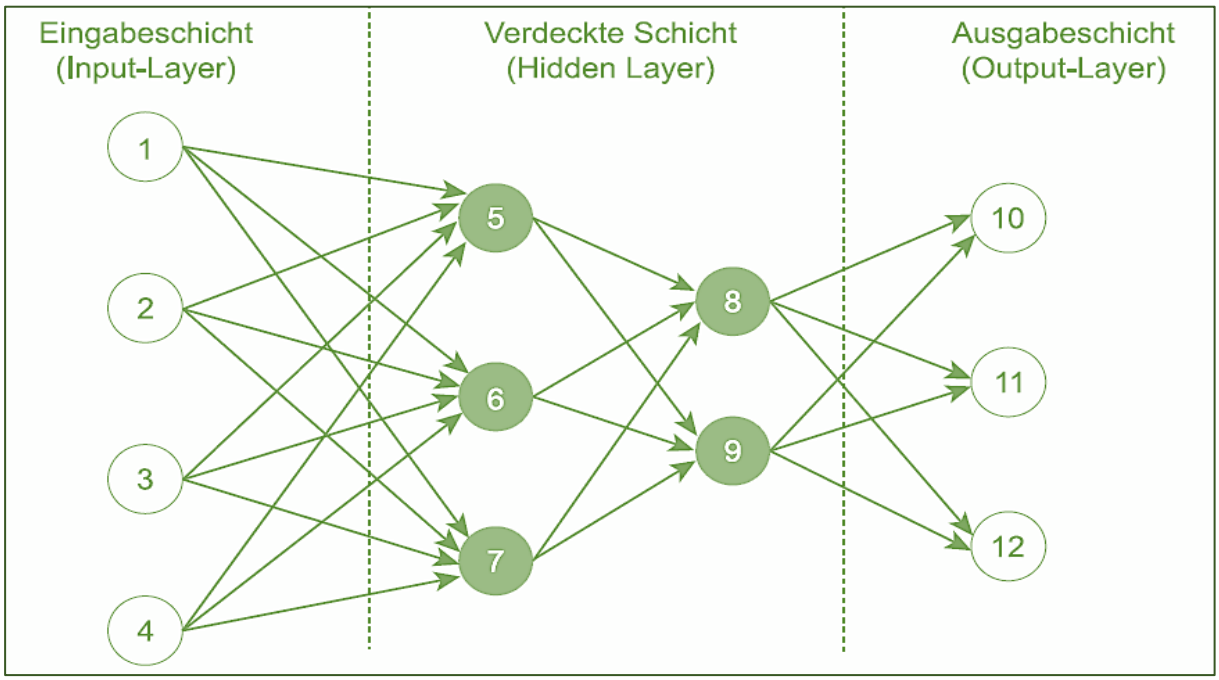

Abb. 13. Beispiel einer möglichen Netzstruktur eines künstlichen neuronalen Netzes (Backhaus et al., 2015, S. 299)

Durch diese Methode können künstliche neuronale Netze im Gegensatz zu den meisten linearen Verfahren auch sehr komplexe Zusammenhänge bestimmen. So ist bspw. ein neuronales Netz in der Lage, sowohl einen agilen, kleinen Stürmer als auch einen großen, kopfballstarken Verteidiger korrekt als Fußballer zu erkennen. Einzelne stärkere

Merkmalsausprägungen können im Klassifikationsverfahren demnach schwächere Merkmalsausprägungen kompensieren (Pion et al., 2017). In weniger komplexen Klassifikationsverfahren würde evtl. nur einer der beiden Fußballertypen oder sogar nur eine Mischform beider Typen als Talent erkannt werden.

Die Güte eines künstlichen neuronalen Netzwerks hängt jedoch stark von der Anzahl und Qualität der vorhandenen Trainingsdaten ab. Eine Bestimmung von talentierten Sportlern kann mittels eines neuronalen Netzes daher auch erst dann geschehen, wenn schon Daten (Kriteriums- und Prädiktor-Daten) von bereits erfolgreichen Sportlern bekannt sind – der Zugang ist folglich überwiegend retrospektiv. Zudem kann es je nach Trainingsdatenlage und Netzvorgaben im Training zu einem *Overfitting* (Überanpassung) kommen. Dies ist unter anderem dann der Fall, wenn für die Berechnungen zu viele Trainingsschritte (oder auch Variablen) angesetzt wurden und es bspw. zum Auswendiglernen einzelner Fälle kommt. Dadurch können neue, noch unbekannte Fälle nicht mehr korrekt klassifiziert werden und die Modellgüte sinkt. Neben einem Datensatz aus reinen Trainingsdaten kommen daher üblicherweise auch ein Validierungssatz (*Dev Set*) und ein Testsatz/Teststichprobe zum Einsatz (Barron et al., 2018). Anhand der Ergebnisse der Validierungs- und Teststichproben lässt sich die Gefahr des Overfittings einschätzen. Um Overfitting zu vermeiden und auch die Fehlerrate der Prognosen zu vermindern, gibt es mehrere Stellschrauben, an denen man (händisch) Korrekturen ansetzen kann (Anzahl der vorgegebenen Hiddenlayer, Eingrenzung der Neuronenanzahl, Anzahl der Trainingsepochen, Größe des Fehlers, Ausklammern von Datenpunkten [dropout]). Diese Modifikationen machen wiederum jede Netzlösung in Abhängigkeit der Trainingsdaten individuell und sind für Außenstehende nicht immer direkt nachvollziehbar.

In der Talentforschung werden unterschiedliche nicht-lineare Prozeduren und Netzlösungen verwendet, darunter *Mulitlayer Perceptrons* (Barron et al., 2018; Musa, Majeed et al., 2019), *Kohonen Feature Maps* (Hohmann & Seidel, 2005; Pion et al., 2017) und *Radial Base Functions* (Heazlewood et al., 2016). Silva et al. (2007) versuchten bspw. mithilfe eines *Multilayer Perceptrons* (MLP), anhand kinanthropometrischer Daten (bioelektrische Impedanzanalyse, Hautfaltendickenmessung, Anthropometrie, etc.), funktioneller (Kraft und Flexibilität) und sportartspezifischer (Hydrodynamik, hydrostatische und bioenergetische Eigenschaften) Fähigkeiten und qualitative Bewertung der Schwimmtechnik, die Schwimmzeiten/-leistung auf 400m und 200m von 65 Schwimmern und 73 Schwimmerinnen auf nationalem Level vorherzusagen. Der Unterschied zwischen den tatsächlich erreichten und den statistisch vorhergesagten Zeiten war hierbei gering (Silva et al., 2007). Die Ergebnisse beruhten allerdings nur auf einem Trainings- (80%) und einem Validierungsdatensatz (20%). Ein weiterer unabhängiger Testdatensatz fehlte.

→ Erstellung der Talentprognose

Mithilfe des durch die gewählte Methode erhaltenen Prognosewertes wird anhand der (zuvor) bestimmten Selektionsrate oder dem Cut-off-Wert eine Auswahl der als Talent vorhergesagten Personen getroffen.

Der in Klassifikationsverfahren oder in Gewichtungen erhaltene Prognosewert stellt den Grad der Zuordnung einer Person zu einer der vorher festgelegten Gruppen (abhängig vom Kriterium) dar. Dementsprechend können üblicherweise Personen mit einem hohen Prognose- bzw. Gewichtungswert statistisch gesehen eher der jeweils gewählten Leistungsgruppe zugeordnet werden als Personen mit einem niedrigeren Prognosewert. Die genaue Anzahl der nun als für eine Leistungsgruppe vorhergesagten bzw. empfohlenen Personen hängt wiederum von der Selektionsrate oder dem Cut-off-Wert ab. Läge bspw. die Selektionsrate bei 10%, so würden von 100 Probanden diejenigen zehn Personen mit den höchsten Prognosewerten als Talent empfohlen. Bei einem Cut-off-Wert von 0,7 würden alle Probanden mit einem Prognosewert größer 0,7 als später erfolgreiche Sportler vorhergesagt.

3.6 Prognosegüte (Prediction)

Die Prognosegüte über den Wert einer Vorhersage lässt sich an statistischen Proberechnungen mithilfe einer Teststichprobe schätzen oder in einer Langzeitstudie anhand des tatsächlich erreichten Kriteriums (retrospektiv) einordnen. Hierbei ist von Interesse, von welchen und wie vielen Sportlern das prognostizierte Kriterium mithilfe der gewählten Prognosemethode korrekt vorhergesagt wurde. Korrekte und falsche Prognosen lassen sich mit einer 4-Felder-Tafel verdeutlichen (Tab. 5). Anhand dieser Tafeln können die verschiedene Güteparameter der Talentprognose bestimmt werden.

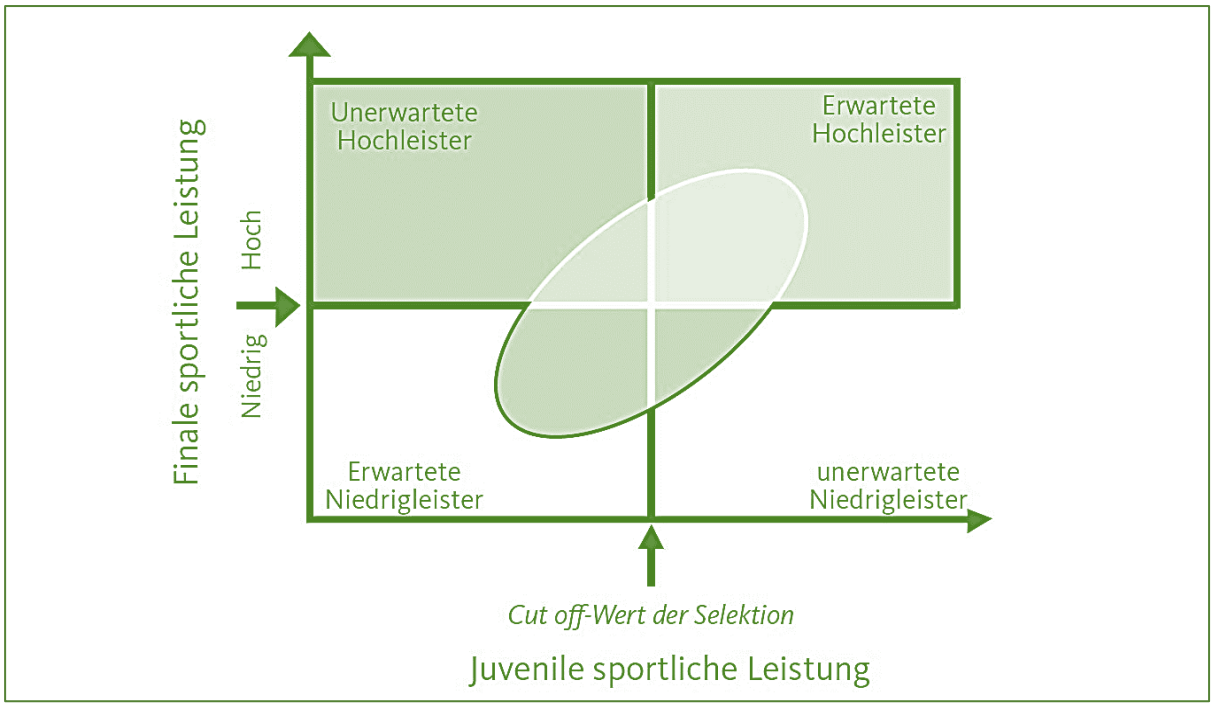

Abb. 14. Grafische 4-Felder-Tafel zur Prognosegüte (Hohmann et al., 2014)

Hohmann et al. (2014) kombinieren in ihrer Abbildung (Abb. 14) die bereits in Abbildung 10 verdeutlichte Streudiagrammdarstellung der Verteilung zwischen Kriterium und Prädiktor mit einer klassischen 2×2-Klassifikationstabelle (teilweise auch *Confusion Matrix* genannt, Tab. 5). Je nach Wissenschaftsbereich (Medizin, Psychologie, Talentforschung, etc.) werden die vier Tabellenbereiche „erwartete Niedrigleister", „unerwartete Niedrigleister" (Typ-I-Fehler), „unerwartete Hochleister" (Typ-II-Fehler) und „erwartete Hochleister" teilweise unterschiedlich angeordnet (Schorer et al., 2017). So können je nach Darstellung z.B. die „erwarteten Hochleister" auch oben links oder unten rechts zu finden sein. Bei nicht dichotom vorliegenden Variablen kann zudem von einer 2×2-Tabelle abgewichen werden. Baker et al. (2018) nutzen bspw. eine 3×3-Tabelle zur Darstellung eines in drei „Leveln" vorliegenden Kriteriums (Abb. 15).

Tab. 5. *2×2-Klassifikationstabelle*

		Predicted/Recommended		
		Low Potential	High Potential	∑
Observed/ Existing	High Performance	C	A	A+C
	Low Performance	D	B	B+D
	∑	C+D	A+B	A+B+C+D = *N*

Legend: A = true positives (TP), B = false positives (FP), C = false negatives (FN), D = true negatives (TN)

Im oberen rechten Feld (A) des vorliegenden dichotomen Beispiels (Abb. 14 bzw. Tab. 5) ist die Anzahl der korrekt prognostizierten Personen zu finden, die das Kriterium erfüllen – sie sind demnach die erwarteten Hochleister oder auch „true positives“. Das linke obere Feld (C) beschreibt die Anzahl der Personen, die das Kriterium erfüllen, jedoch nicht mithilfe der Prognosemethode vorhergesagt bzw. bestimmt werden konnten – sie stellen die unerwarteten Hochleister oder auch „false negatives“ dar.

In der unteren Zeile findet sich links (D) die Anzahl der korrekt vorhergesagten Personen mit nicht erfülltem Kriterium. Dies sind die „erwarteten Niedrigleister“ oder auch „true negatives“. In der rechten Spalte (B) findet sich die Anzahl derjenigen Personen, denen das Kriterium fälschlicherweise prognostiziert/klassifiziert wurde. Es handelt sich dabei um die Menge der „unerwarteten Niedrigleister“ oder auch „false positives“. Die Summe A + C entspricht folglich der Anzahl an Personen der Stichprobe, die das Kriterium erfüllen (in medizinischen Belangen z.B. die Anzahl betroffener bzw. erkrankter Personen). Der relative Anteil dieser Summe zur Gesamtstichprobe entspricht somit der Grundrate (base rate). Die Summe B + D gibt die Anzahl an Personen an, die das Kriterium nicht erfüllen. Summiert man A und B, erhält man die Anzahl der positiven Prognosen. Die relative Anzahl an positiven Prognosen zur Gesamtstichprobe ergibt wiederum die Selektionsrate *S* (Formel 5; diese ist gegebenenfalls abhängig von einem gewählten Cut-off-Wert, sofern dieser verwendet wurde). Die Anzahl der negativen Prognosen geht aus der Summe C + D hervor. Die Summe aller vier Felder (A+B+C+D) ergibt die Gesamtstichprobenanzahl *N*.

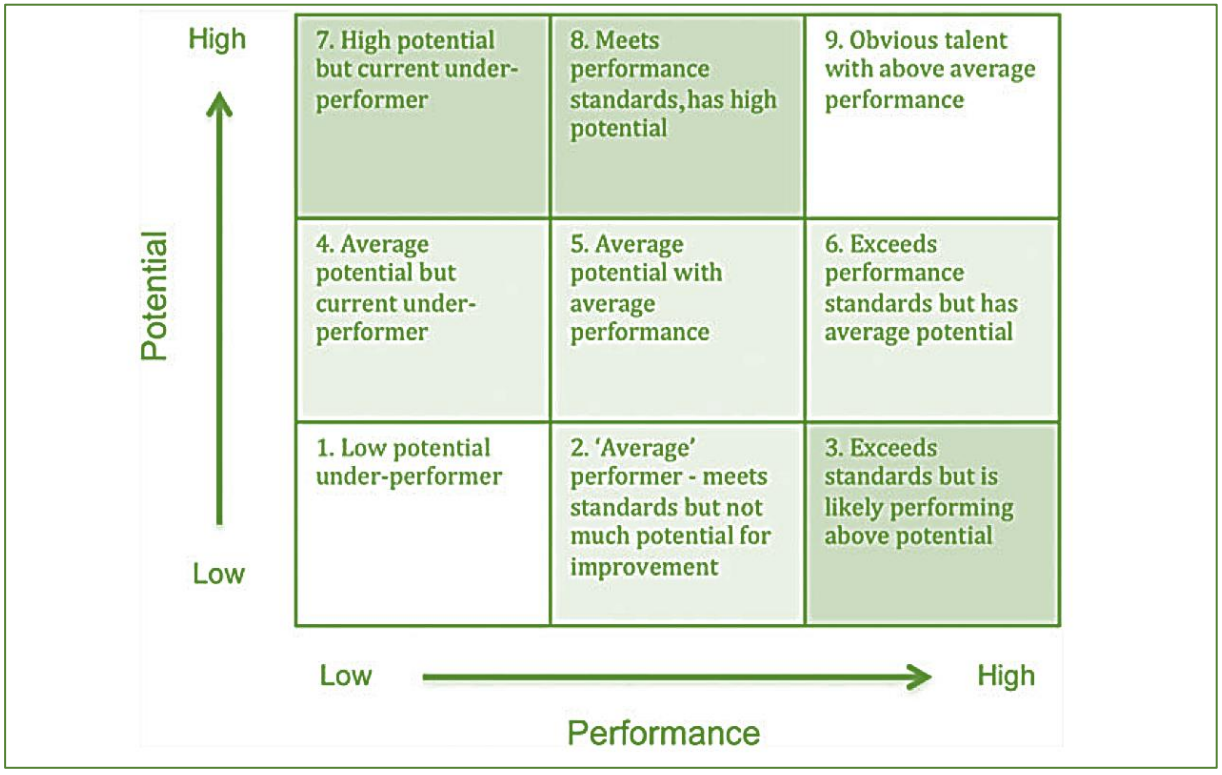

Abb. 15. 6-Felder-Tafel zur Darstellung der möglichen Zusammenhänge zwischen Potential und Leistung (Baker et al., 2018)

Die Prognosegüte kann nun anhand der 2×2-Tabelle (Tab. 5) bestimmt werden (Marx & Lenhard, 2010; Musa, Majeed et al., 2019). In der Praxis finden sich als mögliche Kennzahlen hierbei häufig die *Gesamttrefferquote*, die *Sensitivität* und die *Spezifität*. Die Gesamttrefferquote (Hit-Rate oder auch Accuracy [ACC]) leitet sich aus dem relativen Anteil aller korrekt prognostizierten bzw. klassifizierten Kriteriumsvorhersagen (true positves [A] + true negatives [D]; Formel 4) ab. Sie gibt demnach die Prozentzahl der korrekten Prognosen an.

$$Hit\ Rate = \frac{A + D}{A + B + C + D} \tag{4}$$

$$Selection\ Rate := S = \frac{A + B}{A + B + C + D} \tag{5}$$

Die Gesamttrefferquote schwankt in verschiedenen Studien mitunter stark. So konnten bspw. in Studien von Sieghartsleitner et al. (2019) zu Talentprognosen von 117 Juniorenfußballspielern mittels Trainermeinungen, motorischer Tests, eines multidimensionalen Modells (aus sportmotorischen, familiären, psychologischen und biologischen Variablen) und Kombinationen dieser Ansätze nach fünf Jahren Gesamttrefferquoten von durchschnittlich 77,8% (je nach Methode 71–88%; Tab. 7) erreicht werden. Am besten schnitt erwartungsgemäß das holistische Modell, das alle Einzelmodelle kombiniert, mit 88% korrekten Vorhersagen ab.
Pion et al. (2017) untersuchten retrospektiv mittels verschiedener Prognosemethoden (Trainermeinung [Coaches], Diskriminanzanalyse [DA], Kohonenkarte [KFM] und Multilayer Perceptron [MLP]) die Dropout-Rate nach fünf Jahren von 243 jungen Turnerinnen. Die Gesamttrefferquote der vier Methoden lag durchschnittlich bei 70,6%. Die Trainermeinung erzielte hierbei mit 51,9% korrekten Vorhersagen den schlechtesten Wert der vier Klassifikationsmethoden, während die Klassifikationsergebnisse des Multilayer Perceptrons zu 80,2% zutrafen (Abb. 16).
Nach Güllich und Cobley (2019) liegen die durchschnittlichen Gesamttrefferquoten der meisten Stichproben im Bereich von bis zu 70% (vgl. Dugdale et al., 2019; Heazlewood et al., 2016; Robertson et al., 2015).

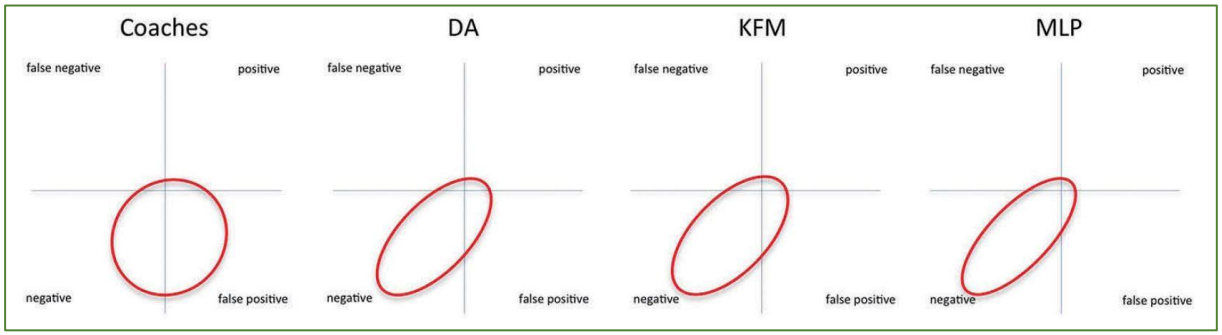

Abb. 16. Prognoseergebnisse von vier Prognose-/Klassifikationsmethoden zum Drop-Out-Verhalten junger Turnerinnen nach fünf Jahren (Pion et al., 2017)

Die beiden Kennzahlen Sensitivität (Recall) und Spezifität geben jeweils das Verhältnis von den durch die jeweilige Methode korrekt vorhergesagten Kriterium-Zuordnungen in ihrer jeweiligen tatsächlich beobachteten Populationsgruppe an (Davis & Goadrich, 2006; Hedderich & Sachs, 2016). Die Sensitivität verdeutlicht hierbei das Verhältnis von „erwarteten Hochleistern“ (true positives, A) unter den tatsächlich später beobachteten Hochleistern (positives, A + C; Formel 6). Unter der Spezifität wird analog das Verhältnis der „erwarteten Niedrigleister“ (true negatives, D) unter den später tatsächlich beobachteten Niedrigleistern (negatives, B + D) verstanden (Musa, Majeed et al., 2019; Formel 7).

$$Sensitivity = \frac{A}{A + C} \tag{6}$$

$$Specificity = \frac{D}{B + D} \tag{7}$$

Aus diagnostischer Sicht verdeutlich demnach die Sensitivität die Anzahl der wahrhaft positiv Getesteten unter den betroffenen bzw. erkrankten Personen und die Spezifität die Anzahl der wahrhaft negativ Getesteten unter den nicht betroffenen bzw. gesunden Personen. Würde bspw. eine Diagnosemethode bei 9 von 10 betroffenen Personen anschlagen, so läge die Sensitivität bei 90%.
In Untersuchungen von Till et al. (2016) an 257 Rugby-Ligaspielern wurden bspw. mithilfe von in der U15 erhobenen Anthropometrie- und Fitnessdaten Prognosen der 7 bis 9 Jahre späteren Leistungserfolge (Amateur/Professional) erstellt. In diesen Erfolgsprognosen konnte eine Sensitivität von 83,3% und eine Spezifität von 63,8% erreicht werden.
Höner und Votteler (2016) erzielten in ihrer Talentstudie an 22.843 U12-Fußballern (siehe Abb. 17) eine Sensitivität von 75,3% und eine Spezifität von 62,0% (bei einem Cut-off-Wert von 60%). Auch Giles et al. (2020) erreichten in ihrer Tennis-Studie Werte im Bereich von 65 bis 72% für die Sensitivität und 75 bis 85% für die Spezifität.

Sensitivität und Spezifität stehen üblicherweise in einem konträren Verhältnis. Dies bedeutet, dass in der Praxis mit steigender Sensitivität eine sinkende Spezifität erwartet werden kann. Der Zusammenhang ist jedoch nicht linear.

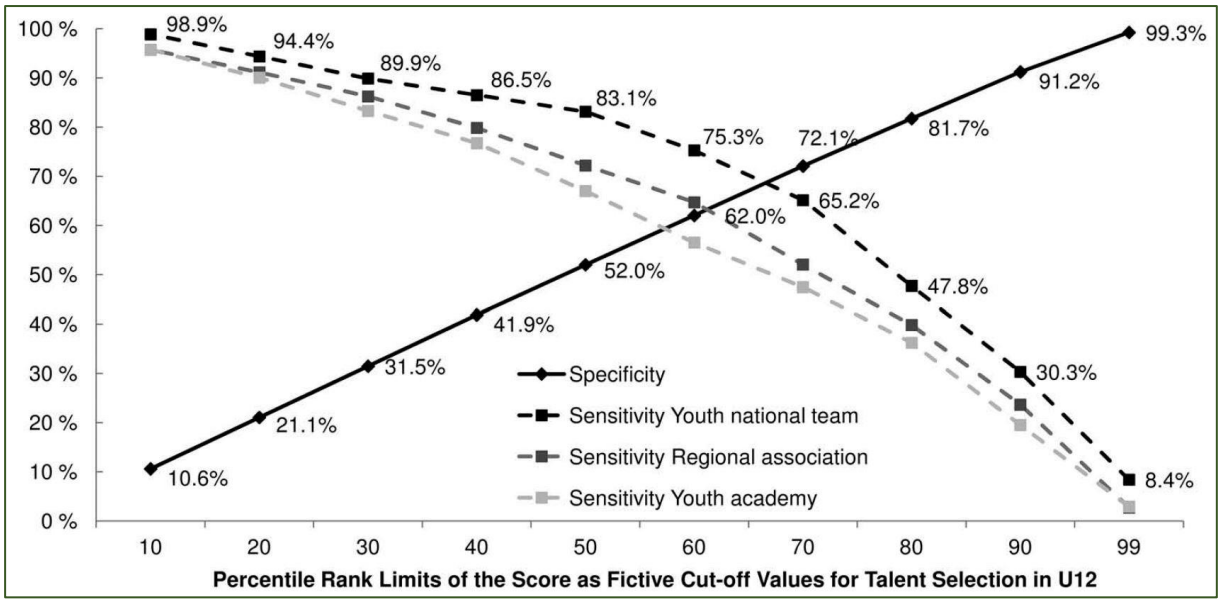

Abb. 17. Sensitivität und Spezifität in Abhängigkeit verschiedener Cut-Off-Werte (Höner & Votteler, 2016, S. 274)

In der Grafik von Höner und Votteler (2016) zu ihrer Studie der U12-Nachwuchsfußballer ist ersichtlich, dass mit steigendem Cut-Off-Wert bzw. sinkender Selektionsrate (siehe x-Achse bzw. Abzisse) die Sensitivität der Prognosen des in der U16 bis U19 erreichten Spielleistungslevels sinkt, während die Spezifität (annähernd linear) steigt (Abb. 17). Würde man der Grafik entsprechend 90% aller U12-Fußballer für das spätere Jugend-Nationalteam empfehlen (Cut-off bei 10% bzw. Selektionsrate bei 90%), so würde man 98,9% der späteren Nationalspieler korrekt vorhersagen, jedoch auch nur 10,6% der später nicht im Jugendnationalkader vertretenen Spieler korrekterweise ausschließen. Prognostizierte man hingegen von den U12-Nachwuchsfußballern nur 1% als für das spätere Jugendnationalteam geeignet, dann wären unter diesen immer noch 8,4% der tatsächlichen späteren Jugendnationalspieler zu finden und auch 99,3% der später niedrigklassiger spielenden Fußballer würden korrekt vorhergesagt.

Neben der Spezifität und Sensitivität werden häufig auch die beiden Kennwerte des positiven und negativen prädiktiven Wertes für die Prognosegüte herangezogen (Allen et al., 2015). Diese beiden Werte geben die Höhe der Wahrscheinlichkeit über das Zutreffen einer positiven bzw. negativen Vorhersage an. Der positive prädiktive Wert (precision) gibt hierbei den Anteil der korrekten positiven Voraussagen unter allen positiven Voraussagen (Formel 8), der negative prädiktive Wert entsprechend den Anteil der korrekten negativen Vorhersagen aller negativen Vorhersagen an (Davis & Goadrich, 2006; Formel 9).

$$Positive\ Predictive\ Value = \frac{A}{A + B} \tag{8}$$

$$Negativ\ Predictive\ Value = \frac{D}{C + D} \tag{9}$$

Allen et al. (2015) überprüften mithilfe von vier Prognosemethoden an 5.738 australischen Schwimmern, wie groß die Kader der letzten fünf Jahre gewesen sein müssten, damit sie 90% (Sensitivität) der australischen Schwimmer mit olympischer Qualifikation beinhaltet hätten. Bei ca. 67 Schwimmern, die die Qualifikationsziele erreichten, lagen die fünf Jahre zuvor berechneten Kadergrößen im Bereich von 413 bis 685 Personen, was wiederum einem positiven prädiktiven Wert von 13 bis 9% entspricht. Für Berechnungen des Vorjahreskaders lag der positive prädiktive Wert bei 38 bis 66%.
Für eine Gruppe von 50 Bogenschützen wurden in Klassifikations-Untersuchungen von Musa et al. (2019) eine deutlich höhere Zahl für den positiven prädiktiven Wert ausgewiesen (Tab. 6). Dabei wurden je nach Methode Werte im Bereich von 88,2 bis 94,1% erzielt. Aufgrund ihrer geringen Stichprobenanzahl und der fehlenden Langzeitprognose ist die Studie jedoch nur bedingt mit anderen Talentprognosen vergleichbar.

Tab. 6. *Prognosegüte zweier Klassifikationsmethoden (mod. nach Musa et al., 2019)*

	Accuracy	**Sensitivity**	**Specificity**	**Precision**
ANN	92%	94.1%	87.5%	94.1%
k-NN	80%	83.3%	71.4%	88.2%

Legend: ANN = Artificial Neural Network Analysis (Multilayer Perceptron); *k*-NN = *k*-Nearest Neighbour Analysis; Precision = positive predictive value

Als weiterer Parameter zur Bestimmung der Prognosegüte können die erklärte Streuung, das (korrigierte) Bestimmtheitsmaß oder die Fehlerquadratsumme des Analyseverfahrens herangezogen werden (Backhaus et al., 2015). So gibt bspw. die Fehlerquadratsumme (Quadratsummenfehler) die Güte eines neuronalen Netzes und analog das Bestimmtheitsmaß R^2 die Güte bei Regressionsanalysen an. Der Wert von R^2 sagt dabei aus, wie dicht die beobachteten Punkte an der Regressionsgeraden liegen. Sieghartsleitner et al. (2017) erreichten in ihrer Talentprognose zur späteren U17-Leistung von U14-Jugendfußballern mittels Trainermeinung, motorischer Testwerte und einer Kombination beider Prädiktoren in Regressionsanalysen ein R^2 von 0,38 (Trainereinschätzung) bis 0,62 (kombiniertes Modell). Die Modellgüte der Regressionsanalyse allein aufgrund der motorischen Testwerte lag bei $R^2 = 0{,}52$.
In Klassifikationsanalysen von Filipcic (1999) an 43 erfolgreichen Tennisspielern wurde anhand von motorischen Testwerten in einer Regression ein Bestimmtheitsmaß von $R^2 = 0{,}427$ erreicht.

3.7 Nützlichkeit (Utility)

Ein Vergleich von verschiedenen Methoden und Studien allein anhand der häufig beschriebenen Kennwerte Gesamttrefferquote, Sensitivität und Spezifität ist nicht immer leicht – sowohl Sensitivität als auch Spezifität sind in hohem Maße von der gewählten Selektionsrate bzw. dem Cut-off-Wert abhängig. Abbildung 17 verdeutlicht, dass sich je nach gewähltem Prozentrang ganz unterschiedliche Einzelwerte ergeben (Höner & Votteler, 2016). So entspricht ein Prozentrang von 40% einer Sensitivität von 86,5% und einer Spezifität von 41,9%, ein Prozentrang von 60% hingegen einer Sensitivität

von 75,3% und einer Spezifität von 62,0%. Ein optimales Verhältnis von Sensitivität und Spezifität kann dabei nicht pauschal bestimmt werden, da je nach Zielsetzung eine hohe Sensitivität oder hohe Spezifität besser sein könnte.

> It seems important to not only look at the percentage of correct classifications. Depending on your position in the sports system, type-I- or type-II-errors have different consequences. For instance, if the prediction model produces more type-I-errors, low achievers are supported and developed although they might not eventually reach elite levels of performance. However, in sport systems under financial constraints (i.e., most sport systems), these types of errors are not affordable. Prediction models that fail to identify an athlete with high potential result in lost talent, which can be particularly damaging if the numbers participating in the sport are low. […] The overall goal of athlete development systems is to facilitate the most talented persons to reach elite status; therefore, it is important to consider which type of error is more relevant for the improvement of the talent identification and development process. (Schorer et al., 2017, S. 1146)

Neben der inhaltlichen Ausrichtung des Talentidentifikationsprozesses – der Fokus liegt vermehrt auf einer Erkennung möglicher Hochleister (hohe Sensitivität) oder verstärkt auf einer Deselektion der Niedrigleister (hohe Spezifität) – spielt auch die Anzahl der erstellten Prognosen eine wichtige Rolle für die Bewertung von Prognoseergebnissen. Werden weniger Hochleister-Prognosen erstellt, als es in der Population tatsächliche Hochleister gibt, so ist eine Sensitivität von 100% nicht mehr erreichbar. Gibt es bspw. eine Grundrate von 10% an Hochleistern, es werden aber nur 9% Hochleisterempfehlungen ausgesprochen, so liegt das erreichbare Maximum der Sensitivität nur noch bei 90%.
Ähnlich verhält es sich bei der Spezifität. Werden mehr Hochleisterempfehlungen ausgegeben als es eigentliche Hochleister gibt, so ist jede dieser „Überempfehlungen" automatisch falsch und die Spezifität sinkt. Die gewählte Selektionsrate bzw. der Cut-off-Wert bestimmt also in einem erheblichen Teil die Sensitivitäts- und Spezifitätswerte.

Auch die Gesamttrefferquote ist häufig ohne weitere Informationen kaum interpretierbar (Marx & Lenhard, 2010). Zunächst lässt sich einerseits von der Gesamttrefferquote nur schwer auf die Einzelkennwerte von Sensitivität und Spezifität schließen, da die jeweiligen Gruppengrößen an Hochleistern und Niedrigleistern in der Praxis ungleich verteilt sind (siehe Kap. 3.3) und somit die Gesamttreffergenauigkeit durch die größere Stichprobenteilgruppe verzerrt wird. In einer retrospektiven Studie von Mostaert et al. (2020) mit 1.151 10 bis 13 Jahre alten Volleyballspielern zeigte die Diskriminanzanalyse in der Prognose der in der U16 bis U20 erreichten Volleyballleistung (n = 1.125 non-selected & n = 26 national selection) eine Gesamttreffergenauigkeit von 97,7%. Jedoch konnten nur 7,7% der späteren Nationalspieler korrekt klassifiziert werden (Sensitivität). Der hohe Wert in der Gesamttreffergenauigkeit wurde vorrangig durch eine hohe Spezifität von 99,9% erreicht. Da in der Veröffentlichung der Studie die Selektionsrate nicht mit angegeben wurde, lässt sich das Ergebnis nur teilweise bewerten. Sollte eine sehr geringe Selektionsrate von deutlich weniger als 2,25% (= Grundrate) vorliegen, wären der geringe Sensitivitätswert und der hohe Spezifitätswert besser einzuschätzen (vgl. Prozentrang „99" bei Abb. 17).

Andererseits ist außerdem die Höhe der Gesamttrefferquote im Vergleich zur Zufallstrefferquote zu werten (Formel 10), die allerdings nicht immer angegeben wird. Einem Beispiel von Marx und Lenhard (2010) zufolge, läge für ein Kriterium, das bei 10% der Stichprobe vorliegt (Grundrate = 10%), und einer Selektionsquote von 10% (10% der Probanden erhalten eine Vorhersage/Klassifikation für das Kriterium) die Zufallstrefferquote bereits bei 82% (Formel 11).

$$Random\ Hit\ Rate = \frac{S \times (A + C) + (1 - S) \times (B + D)}{A + B + C + D} \tag{10}$$

$$\Rightarrow Random\ Hit\ Rate = \frac{0{,}1 \times 10 + (1 - 0{,}1) \times 90}{100} = 0{,}82 \tag{11}$$

Eine Gesamttrefferquote einer Talentprognose von 80% wäre für das vorliegende Beispiel von Marx und Lenhard (2010) dementsprechend als „ungenügend“ einzustufen. In anderen Studien bei einer anderen Zufallstrefferquote könnte die gleiche Gesamttrefferquote jedoch wiederum als „hoch“ einzustufen sein.
In einer 10-Jahres-Langzeitstudie von Schorer et al. (2017) an 68 Handballspielern wurde daher auch eine Referenzwahrscheinlichkeit (hier: die a-priori Wahrscheinlichkeit) angegeben, um die Gesamttrefferquoten verschiedener Prognosemethoden besser vergleichen zu können. Am besten schnitt die Klassifikation anhand der motorischen Tests ab (85,2% Gesamttrefferquote), gefolgt von den Empfehlungen der Nationaltrainer (79,3%), der a-priori Wahrscheinlichkeit (75,9%) und den Empfehlungen der Handballspieler (75,8%) und Novizen (z.B. Zuschauern, 72,4%). An dem Beispiel lässt sich erkennen, dass sowohl die Prognosen der Handballspieler selbst als auch die der Novizen (Personen ohne eigene Handballerfahrung als Spieler) für eine Talentidentifikation ungeeignet scheinen.

Ähnlich wie Sensitivität und Spezifität hängen auch die Werte des positiven und negativen prädiktiven Wertes stark von der Stichprobenverteilung der Kriteriumsgruppen ab. Nach Güllich (2019) gehen wissenschaftliche Studien typischerweise von einer Gleichverteilung der Leistungsgruppen (Hochleister und Niedrigleister) aus, die Grundrate sei in der Praxis jedoch viel geringer. Dies habe zur Folge, dass der positive prädiktive Wert in der Praxis deutlich geringer ausfalle, als es die Studien vermuten ließen. Güllich verdeutlich dies anhand eines Beispiels: Nimmt man an, dass 1 von 1.000 Personen in der Praxis ein erfolgreicher Spitzensportler werden würde und eine Studie mit einer Gesamttreffergenauigkeit von 70% (Spezifität und Sensitivität liegen in diesem Beispiel ebenfalls bei 70%) vorliegt, dann würden von 999 Niedrigleistern ca. 300 (30%) Personen fälschlicherweise als Talent identifiziert/vorhersagt. Die Wahrscheinlichkeit, mit einer positiven Talentprognose auch tatsächlich einen erfolgreichen Sportler zu identifizieren, läge daher nur noch bei 0,23% (positiver prädiktiver Wert). Selbst bei einer Gesamttrefferquote und Spezifität von 90% würde der positive prädiktive Wert lediglich bei 0,9% liegen (Güllich, 2019).

Eine Betrachtung weiterer Kennwerte scheint daher notwendig, um die Ergebnisse adäquat einschätzen zu können.

Eines der einfacheren Verfahren stellt der *Youden Index* bzw. *Youdens J* dar (Youden, 1950). Hierbei werden die Ergebnisse aus Sensitivität und Spezifität addiert und die Summe um 1 subtrahiert (Formel 12). In der Praxis haben Sensitivitäts- und Spezifitätsveränderungen teilweise unterschiedliche Auswirkungen. So kann bspw. ein falsch positives Testergebnis höhere Kosten verursachen als ein falsch negatives Ergebnis. In solchen Fällen kann eine zusätzliche Gewichtung ω (mit $0 < \omega < 1$) eingefügt werden (Hedderich & Sachs, 2016; Formel 13).

$$Youden's\ J\ =\ \text{Sensitivity} + \text{Specificity} - 1 \tag{12}$$

$$Youden's\ J'\ =\ \omega \times \text{Sensitivity} + (1 - \omega) \times \text{Specificity} \tag{13}$$

Youdens J nimmt somit üblicherweise Werte zwischen 0 und 1 an. Ein Wert von 1 entspricht einem perfekten Ergebnis und einer Gesamttreffergenauigkeit von 100%. Theoretisch sind aber auch negative Werte denkbar. *Youdens J* besitzt den Nachteil, dass ein Wert von 1 aufgrund von Überempfehlungen (Selektionsrate ≠ Grundrate) einzelner Gruppen teilweise gar nicht erreicht werden kann. So würde bspw. bei einer Grundrate von 50% und einer Selektionsrate von 60% nur noch eine Spezifität und folglich auch ein *Youdens J* von maximal 80% erzielt werden können.

Neben *Youdens J* ist auch der F_1-Wert (eigentlich F_α mit $\alpha = 1$) für binäre Klassifikationen ein möglicher Kennwert der Prognosegüte (Giles et al., 2020). Er basiert auf dem harmonischen Mittelwert und setzt sich zu gleichen Teilen aus Sensitivität (Recall) und positivem prädiktiven Wert (Precision) zusammen (Formel 14). Ebenso wie *Youdens J* können auch hier Werte größer 0 und kleiner oder gleich 1 angenommen werden, wobei ein Wert von 1 für die maximale Vorhersagestärke steht. Anders als bei *Youdens J* werden die *true negatives* hingegen nicht in die Formel miteinbezogen. Der F_1-Wert zielt somit vermehrt auf eine Erkennung von Hochleistern bzw. betroffener Personen ab.
In bestimmten Fällen könnte dadurch jedoch eine hohe Prognosegüte berechnet werden, obwohl das Prognoseverfahren für die Praxis augenscheinlich ungeeignet ist. Dies ist bspw. dann der Fall, wenn keiner der Niedrigleister korrekt erkannt wird, jedoch die Grundrate an Hochleistern sehr groß ist. So ließe sich bei einer Grundrate von 95% und einer Prognosemethode, die allen Personen pauschal eine Hochleister-Empfehlung aussprechen würde (TP=95, FP=5, TN=0, FN=0), dennoch ein hoher F_1-Wert von $F_1 \approx 97{,}4\%$ erreichen.

$$F_1\ =\ 2 \times \frac{\text{Sensitivity} \times \text{Positive Predictive Value}}{\text{Sensitivity} + \text{Positive Predictive Value}} = \frac{2A}{2A + B + C} \tag{14}$$

$$MCC = \frac{A \times D - B \times C}{\sqrt{(A + B) \times (A + C) \times (D + B) \times (D + C)}} \tag{15}$$

Aufgrund der beschriebenen Einschränkungen des F_1-Wertes wird in verschiedenen Studien mit binärer Klassifikation stattdessen der *Matthews-Korrelationskoeffizient* (MCC) als Kennwert der Prognosegüte angegeben (Matthews, 1975). Der *Matthews-Korrelationskoeffizient* (Formel 15) ist ein von der Problematik der unausgewogenen (binären) Datensätze unbeeinflusstes alternatives Gütemaß, bei dem der Korrelationskoeffizient (Phi-Koeffizient ϕ) zwischen tatsächlichen und vorhergesagten Werten berechnet wird (Chicco & Jurman, 2020; Hedderich & Sachs, 2016). Der MCC kann Werte zwischen −1 (perfekte Fehlklassifikation) und +1 (perfekte Klassifikation/Prognose) annehmen, wobei ein Wert von 0 auf eine rein zufällige Prognose hinweist (Musa, Majeed et al., 2019).
Greift man bspw. das vorherige Beispiel auf, in dem ein F_1-Wert von 97,4% erreicht wird, und bestimmt mithilfe des Grenzwertes (gegen Null) näherungsweise den MCC, so geht dieser gegen MCC = 0, was auf ein zufälliges Ergebnis hinweist. Dennoch, so Farrington und Loeber (1989), sei der MCC nicht immer fehlerfrei zu interpretieren, weil ein Wert von MCC = +1 nur dann ermöglicht würde, wenn Grundrate und Selektionsrate gleich groß wären. In vielen Fällen würde die Selektionsrate aber von der Grundrate abweichen und ein maximaler MCC-Wert somit deutlich unter dem theoretisch möglichen Maximum von +1 liegen. Dies könne Vergleiche von MCC-Werten unterschiedlicher Studien erschweren.

Ein ähnliches Problem tritt ebenfalls bei *Cohen's Kappa* κ auf (J. Cohen, 1960; Formel 16). Auch hier wird das theoretisch Maximum von $\kappa = 1$ nur bei einer Gleichverteilung von Grundrate und Selektionsrate erreicht.

$$K = \frac{Hit\ Rate - Random\ Hit\ Rate}{1 - Random\ Hit\ Rate} \tag{16}$$

$$RIOC = \frac{Hit\ Rate - Random\ Hit\ Rate}{Maximum\ Hit\ Rate - Random\ Hit\ Rate} \tag{17}$$

$$Maximum\ Hit\ Rate = 1 - \frac{|B - C|}{A + B + C + D} \tag{18}$$

Ein Kennwert, der diese Problematik des nach oben hin gedeckelten stichprobenabhängigen Maximums umgeht, ist der Wert des *Relative Improvement Over Chance* (RIOC, Formel 17). Der RIOC-Wert orientiert sich sowohl an einem zufälligen Ergebnis als auch an dem tatsächlich möglichen (Cairney & Streiner, 2011; Marx & Lenhard, 2010). Somit kann er unabhängig von der Stichprobenverteilung und Selektionsrate Werte zwischen 0 (zufälliges Ergebnis) und 1 (maximal mögliches Ergebnis) annehmen. Für Vorhersagemethoden, die schlechter als eine zufällige Ziehung ausfallen, sind auch negative Werte denkbar. Die Berechnung der für die Bestimmung des RIOC-Wertes notwendigen *maximal möglichen Trefferquote* geschieht hierbei bspw. über die Gegenwahrscheinlichkeit (Formel 18). Dabei werden die aufgrund der von der Grundrate abweichenden Selektionsrate entstandenen Überempfehlungen einer Gruppe von der Gesamtzahl möglicher Treffer abgezogen. Ab einem Wert von RIOC > 0,33 wird die

Prognosegüte als „gut“ angegeben, ab einem Wert von RIOC > 0,66 als „sehr gut“ (Marx & Lenhard, 2010). In einer Klassifikationsstudie zum Drop-Out-Verhalten junger Turnerinnen von Pion et al. (2017) wurden umgerechnete RIOC-Werte zwischen 0,278 (MLP) und 0,537 (Gewichtung) erreicht. Modelle zur Talentidentifikation im Jugendfußball von Sieghardsleitner et al. (2019) kamen sogar auf RIOC-Werte von 0,866 und höher.

Ein weiteres Verfahren, das sich neben der Bestimmung geeigneter Cut-off-Werte ebenfalls zur Einschätzung der Prognosegüte eignet, ist das *receiver operating characteristic Verfahren* (ROC). Bei diesem Vorgehen wird die Selektionsrate der Prognosemethode schrittweise erhöht und die zugehörige Sensitivität und Spezifität berechnet (vgl. Abb. 17). Das Ergebnis wird als Diagramm in einer sogenannten ROC-Kurve festgehalten (Abb. 18). Hierbei werden die Werte der Sensitivität auf der Ordinate und die Rate der Falsch-Positiven (1−Spezifität) auf der Abszisse abgetragen (Schmidt-Atzert et al., 2018). Befindet sich die erstellte Kurve in der Nähe der gedachten Diagonalen, was einer ähnlichen Treffer- und „false-positive“-Quote entspricht, so weist dieser Umstand auf ein eher zufälliges Prognoseergebnis hin. Eine ROC-Kurve deutlich oberhalb der Diagonalen deutet hingegen auf eine hohe Prognosegüte hin (Hedderich & Sachs, 2016). Werte unterhalb können ebenfalls auf eine hohe Prognosegüte hinweisen, wobei hier die Prädiktoren dann evtl. neu interpretiert werden müssten.

Ein theoretisch „optimaler“ Trennwert (Cut-off-Wert bzw. Selektionswert) ist gegeben, wenn eine hohe Sensitivität bei gleichzeitig hoher Spezifität vorliegt (Hedderich & Sachs, 2016). Dieser Wert ergibt sich rechnerisch aus demjenigen Punkt, der von der Diagonalen den größten Abstand besitzt. Dieser Wert ist gleichbedeutend mit dem Maximum der durch Youdens *J* erhaltenen Kennwerte. Für die Praxis muss dieser „optimale“ Trennwert jedoch nicht unbedingt den für die jeweilige Situation geeignetsten Trennwert darstellen, da, wie bereits genannt, die Folgen der Fehler 1. und 2. Art häufig in einem ungleichen Verhältnis stehen (Kap. 3.6).

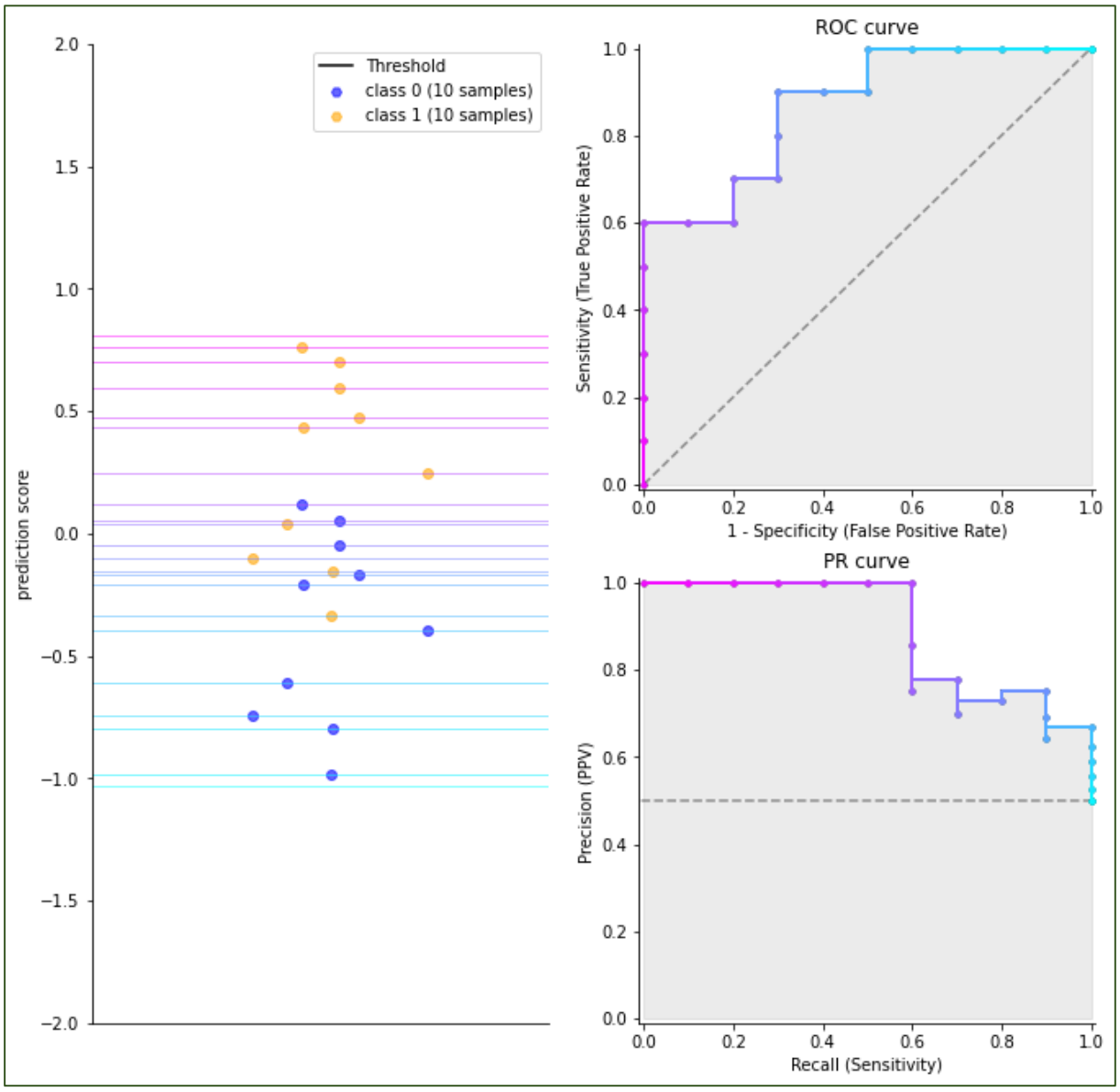

Abb. 18. Beispiel einer „receiver operating characteristic" Kurve (ROC Curve) und Precision-Recall-Kurve (PR Curve) in Abhängigkeit des jeweils gewählten „precision score" Grenzwerts (Azevedo, 2020)

Zur Erfassung der Prognosegüte mittels einer ROC-Kurve hat sich die Prozentzahl der *Area Under [the] Curve* (AUC, für ROC-Kurven auch teilweise AUROC genannt) etabliert. Hierbei spiegelt ein Wert von AUC = 0,5 ein zufälliges Ergebnis wider, während ein Wert von AUC = 1 auf eine perfekte Klassifikation hinweist. Ein Wert von AUC < 0,70 wird häufig als „suboptimal", Werte 0,70 ≤ AUC ≤ 0,80 als „gut" und Werte darüber als „exzellent" bezeichnet (D'Agostino et al., 2013). In einer Talentstudie von Till et al. (2016) an 257 U15-Rugbyligaspielern wurden mithilfe von anthropometrischen und sportmotorischen Daten bspw. AUC-Werte zwischen 0,603 (Professional vs. Academy) und 0,766 (Professional vs. Amateur) erreicht.
Faber et al. (2017) untersuchten mithilfe der vier Testaufgaben „sprint", „agility", „speed while dribbling" and "throwing a ball", ob sich von 1.191 U11-

Tischtennisspielern die zukünftige Leistung der U13 bis U18 vorhersagen lässt. Für männliche Tischtennisspieler zeigten sich nur die beiden Testaufgaben Sprint und Ballweitwurf als signifikante Prädiktoren. Die berechneten AUC-Werte lagen unabhängig von der Altersklasse bei ca. 60%. Für weibliche Tischtennisspielerinnen war hingegen die Testaufgabe „speed while dribbling" ein signifikanter Prädiktor. Die AUC-Werte gingen aber auch dort nicht über AUC = 65,2% hinaus.
Eine ähnliche Beobachtung trifft auch auf Studien von Sieghartsleitner et al. (2019) zu. In Untersuchungen an insgesamt 218 Schweizer Juniorenfußballern (U13–U17) zur Klassifikation der späteren U20-Spielerfolge (Kriterium: Erreichen der 1 bis 3 Liga oder einen Platz im Nationalteam) kamen sie nur auf AUC-Werte von knapp 70% (Tab. 7).
Anders verhält es sich in einer Langzeitstudie von Ribeiro Junior et al. (2021) an 4.692 brasilianischen Basketballern. Anhand der Prädiktor-Werte von „birthdates, height, body mass, playing position, geographic region, club, competition category and team performance" konnte sogar eine Fläche von 93% unter der ROC-Kurve erreicht werden.

Tab. 7. *Descriptive values of the receiver operating characteristic curves (Sieghartsleitner, Zuber, Zibung, Charbonnet & Conzelmann, 2019, S. 7)*

	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

Legend: AUC = Area Under the Curve, CI = Confidence Interval, YI = Youden Index

Im klinischen Bereich sind jedoch häufig nur die Grenz- bzw. Randbereiche der Kurve interessant, da oft eine vergleichsweise geringe Anzahl an Typ-I oder Typ-II-Fehlern gesucht wird. Die Angabe der gesamten Fläche unter dem Graphen kann diese Güte der Grenzbereiche aber nicht immer widerspiegeln. Je nach Zielsetzung ist ein Vergleich von AUC-Werten von ROC-Kurven daher nicht immer sinnvoll. An dieser Stelle können andere Kurven, z.B. *Precision-Recall-Kurven* (PR-Kurven bzw. Sensitivität–positiver-prädiktiver-Wert-Kurven), eingesetzt werden, um den Fokus eher auf bestimmte Trefferbereiche (hier: „true positives") zu legen (Davis & Goadrich, 2006). Entgegen der ROC-Kurven liegt bei PR-Kurven eine hohe Prognosegüte vor, wenn der Graph sich der rechten oberen Ecke annähert (Abb. 18). Ein zufälliges Ergebnis liegt dabei nicht auf der Diagonalen, sondern auf einer von der jeweiligen Grundrate abhängigen Geraden – diese verläuft parallel zur Abszisse auf Höhe des Grundratenwerts. Auch hier lässt sich die (prozentuale) Fläche unter dem Graphen (Area Under [the] Precision-Recall-Curve) als Vergleichswert heranziehen (Zhou et al., 2021).

Eine gewisse Ähnlichkeit mit dem AUC besitzt ebenfalls eine von Brier vorgeschlagene Methode (Brier, 1950). Er berechnet die Güte des Klassifikators gleichermaßen über alle Prognosen hinweg. Dabei wird ein eigener Güte-Kennwert, der *Brier-Wert*, für jede individuelle Prognose anhand der zugehörigen Prognosewahrscheinlichkeit (Wahrscheinlichkeitswert für einen späteren Hochleister) bestimmt und dieser Kennwert später über alle Prognosen gemittelt (Formel 19). Inhaltlich entspricht das Vorgehen der mittleren quadratischen Abweichung (*mean squared error*) der einzelnen Prognosewahrscheinlichkeiten (Beleites et al., 2013; Murphy & Winkler, 1977).

$$P = \frac{1}{n}\sum_{k=1}^{n}(Prognosewahrscheinlichkeit_k - E_k)^2 \tag{19}$$

In der Formel ist $E_k = 1$, wenn das Kriterium zutrifft – also die *k*-te Person ein Hochleister ist – und $E_k = 0$, wenn das Kriterium für diese Person nicht zutrifft. Der Mittelwert *P* kann dabei Werte zwischen 0 und 1 annehmen. Geht *P* gegen Null, so ist die Prognosegüte als hoch einzuschätzen. Der Brier-Wert stößt jedoch an seine Grenzen, wenn er auf sehr seltene (oder sehr häufige) Ereignisse sowie auf Prognosen von schlechter Qualität angewendet wird (Benedetti, 2010).

Eine weitere Möglichkeit, die Aussagekraft von Diagnosen zu bewerten, ist mittels des *Likelihood-Quotienten* möglich. Der Likelihood-Quotient gibt an, wie sich die Prognose auf die Chance (Odds) des Vorliegens eines Kriteriums auswirkt (Hedderich & Sachs, 2016). Hierbei werden zwei verschieden Quotienten herangezogen. Der Quotient LR^+ gibt Aufschluss darüber, wie sich die Prätest-Chance für das tatsächliche Vorliegen eines Kriteriums/Krankheit (Formel 20) und LR^- für das tatsächliche Nicht-Vorliegen eines Kriteriums (Formel 21) ändert.

$$LR^+ = \frac{\text{Sensitivity}}{1 - \text{Specificity}} \tag{20}$$

$$LR^- = \frac{1 - \text{Sensitivity}}{\text{Specificity}} \tag{21}$$

Durch Multiplikation der Prätest-Chance mit dem entsprechenden Quotienten wird die Posttest-Chance berechnet. Läge beispielsweise die Prätest-Wahrscheinlichkeit (Prävalenz), ein späterer Hochleister zu werden, bei 10%, so würde die Prätest-Chance, ein Hochleister zu werden, einem Verhältnis von 1 zu 9 entsprechen. Ein Likelihood-Quotient von $LR^+ = 3{,}0$ (z.B. für eine Sensitivität und Spezifität von 75%) erhöht die Chance, ein Hochleister zu werden, um das Dreifache. Die Posttest-Chance liegt somit bei 3 zu 9, was einer Posttest-Wahrscheinlichkeit von nun 25% entspricht (nach Hedderich & Sachs, 2016).

Exkurs: Vergleich von Klassifikatoren
Neben den bereits beschriebenen Gütekennwerten vorrangig binärer Klassifikationen, die eine Abwägung verschiedener Prognosemethoden am selben Datensatz ermöglichen, existieren weitere statistische Maßnahmen, um verschiedene Klassifikatoren zu bewerten. Hierzu gehören verschiedene Pseudo-Bestimmtheitsmaße, wie bspw. Nagelkerkes R^2. Dietterich (1998) beschreibt zudem gleich fünf verschieden Methoden, um Klassifikatoren zu vergleichen: Ein Zweistichproben-*t*-Test für unabhängige Stichproben, ein Zweistichproben-*t*-Test für verbundene Stichproben, ein Zweistichproben-*t*-Test für verbundene Stichproben mit 10-facher Kreuzvalidierung (10-fold CV), der McNemar-Test und ein Zweistichproben-*t*-Test für verbundene Stichproben mit fünf Wiederholungen einer 2-fachen-Kreuzvalidierung (5 × 2 cv paired *t* test). Der Einsatz dieser Methoden ist jedoch jeweils abhängig vom vorhanden Datensatz und der genutzten Klassifikationsmethode.
Ein Vergleich zweier Talentprognosen anhand von AUCs kann mithilfe des DeLong-Tests geschehen (DeLong et al., 1988). Unterschiedlicher Brier-Werte können anhand des *Brier Skill Scores* bewertet werden (Brier, 1950).

Insgesamt macht eine Betrachtung der verschiedenen Kennzahlen der Prognosegüte nur dann Sinn, wenn diese anhand eines unabhängigen Test-Datensatzes erhoben wurden. Nutzt man zu Bestimmung der Gütewerte den bereits für das Training der Klassifikationsmethode bereitgestellten Datensatz, so werden schon bekannte Fälle erneut klassifiziert. Dies hat zur Folge, dass die berechneten Treffergenauigkeiten üblicherweise höher ausfallen, als es in der Praxis mit noch unbekannten, neuen Fällen zu erwarten wäre.
Dass dieses Problem nicht trivial ist, zeigt eine Studie von Mostaert et al. (2021) an 243 heranwachsenden Radfahrern. Die Autoren versuchten mithilfe einer Diskriminanzanalyse, verschiedene Typen von Radfahrern zu unterscheiden (cyclists from road, cycling, track cycling, cyclo-cross, and mountain bike). Bei einer Klassifikation ohne unabhängigem Testdatensatz wurde eine Gesamttreffergenauigkeit von 80,7% erreicht. Die Genauigkeit eines unabhängigen Testdatensatzes lag hingegen nur noch bei 42,1%.

In der Praxis hat sich aus diesem Grund die *k*-fache Kreuzvalidierung (CV) etabliert. Hierbei wird der Datensatz in *k* (gleichgroße) Teile unterteilt und jeweils ein Teil anhand der anderen $k-1$ Teile klassifiziert/prognostiziert (Pion et al., 2017). Durch diese Methode dient jeder Datenpunkt der Stichprobe $k-1$-mal als Trainingspunkt (und/oder Validierungspunkt) und wird genau einmal klassifiziert. Bei kleinen Datensätzen kann es zusätzlich Sinn machen, die *k*-fache Aufteilung zu variieren und weitere *k*-fache Trennungen des Datensatzes vorzunehmen (Dietterich, 1998; Kolias et al., 2021). Diese Verfahrensweise kann verhindern, dass durch eine zufällige Verteilung – z.B. von sehr ähnlichen Datenpunkten in einer einzelnen Partition – das Gesamtergebnis verfälscht wird. Eine besondere Form der *k*-fachen CV ist die *Leave-one-out*-Methode. In diesem Verfahren wird ein Datensatz mit *N* Datenpunkten in genau *N* Teile geteilt und jeweils immer nur ein einziger Datenpunkt anhand aller anderen $N-1$ Fälle bestimmt (Hagenbuchner et al., 2015). Am Ende des Verfahrens werden die Ergebnisse der verschiedenen Partitionsergebnisse gemittelt und so die gesuchten Gütekennwerte berechnet.

4. Einordnung der Publikationen in die wissenschaftliche Talentforschung

Die Validierung einer Talentidentifikation kann aufgrund ihrer vielen Stellschrauben innerhalb des Validierungsprozesse durchaus als „herausfordernd“ beschrieben werden. Nicht allein die Prädiktoren, sondern auch die richtige Auswahl der Stichprobe, des passenden Kriteriums und der geeigneten Prognosemethode können sich auf die letztendliche Güte der Talentidentifikation auswirken. Zudem muss der jeweils erstellte Prognosekennwert wiederum in Anbetracht der Zielsetzung der zu überprüfenden Talentidentifikation bewertet werden. Hierbei haben üblicherweise Fehler 1. Art und Fehler 2. Art eine unterschiedlich hohe Bedeutung.

Die nachfolgende Publikationsreihe versucht, sich diesen methodischen Schwierigkeiten zu stellen und validiert die Talentprognose einer Talentidentifikation.
Die einzelnen Artikel entstanden durch eine sich über die Jahre entwickelte Validierungsprozedur, deren methodische Konzeption in die Validierungsschritte mündete, die in den vorangegangenen Kapiteln erläutert wurden. Die ersten Publikationen beschränken sich somit nur auf einzelne Aspekte der Validierung von Talentidentifikation, während in den abschließenden Studien mehr und mehr grundsätzliche Validierungsschritte angewendet wurden. Allen Artikeln gemein ist, dass diese sich jeweils nur auf einzelne Bereiche der Talentidentifikation (z.B. die Talentorientierung oder Talentidentifikation i.e.S.) und mitunter nur auf einzelne Sportarten beschränken.

Als Grundlage aller Studien dienen die Daten des *Fuldaer Bewegungs-Checks* (FBC), einer nachhaltigen Kampagne zur Bewegungs- und Gesundheits- sowie Sport- und Talentförderung der hessischen Region Fulda. Der Fuldaer Bewegungs-Check wurde bereits im Jahr 2010 als Projekt der Universität Bayreuth unter der Beteiligung des Landkreises und der Stadt Fulda ins Leben gerufen. Das übergeordnete Ziel der Kampagne ist es, die Bildungsregion sportlich zu fördern und die teilnehmenden Kinder nachhaltig zu mehr sportlicher Aktivität und einem gesünderen Lebensstil zu motivieren. Als Grundlage für dieses Vorhaben dienen sportmotorische Tests und psychologische Befragungen. Neben eines individuellen Stärken-Schwächen-Profils und einer passenden Sportartempfehlung, die jeder Teilnehmer im Anschluss an die Testungen erhält, werden in der Kampagne motorisch schwache und besonders talentierte Kinder identifiziert und in passende Fördermaßnahmen vermittelt. Identifikation und Sportartorientierung von Talenten stehen im Fokus der nachfolgenden Validierungsstudien.

Prädiktoren

Als Prädiktoren der Talentidentifikation werden vorrangig die Ergebnisse der von der Deutschen Vereinigung für Sportwissenschaft vorgestellten Testaufgaben des *Deutschen Motorik Tests 6–18* genutzt (Bös et al., 2009). Ergänzt werden diese acht Tests durch drei weitere Testaufgaben, die das allgemeine Fähigkeitsprofil der populäreren Sportarten vervollständigen sollen. Die Tests konzentrieren sich überwiegend auf

technisch unspezifische Fähigkeiten, weshalb ihre Ergebnisse die allgemeine sportliche Leistungsfähigkeit eines jeden Teilnehmers widerspiegeln.

Zur Testbatterie des Fuldaer Bewegungs-Checks gehören insgesamt elf Einzeltests, die auf die allgemeinen sportmotorischen Fähigkeiten wie Ausdauer, Kraft Schnelligkeit, Beweglichkeit und Koordination sowie auf einzelne grundlegende motorische Kompetenzen abzielen (Tab. 8). Zusätzlich werden Daten zu Alter (in Monaten), Geschlecht, Körpergröße und Körpergewicht erhoben.

Tab. 8. *Sportmotorische Testaufgaben des Fuldaer Bewegungs-Checks*

Prädiktoren	**Zielsetzung**	**Objektivität**	**Reliabilität**
Standweitsprung	Schnellkraft	0,99[1]	0,89[1]
20m-Sprint	Schnelligkeit	0,86[3]	0,96[3]
Bauchaufzüge	Kraftausdauer	0,92[4]	0,74[4]
Liegestütz	Kraftausdauer/Koordination	0,98[1]	0,63[1]
Seitliches Hin- und Herspringen	Koordination/Schnelligkeit	0,99[1]	0,89[1]
Rückwärts Balancieren	Koordination – Gleichgewicht	0,99[3]	0,73[3]
6min-Ausdauerlauf	Aerobe Ausdauer	0,87[1]	0,92[1]
Rumpfbeuge	Allgemeine Beweglichkeit	0,99[1]	0,94[1]
Hand-/Griffkraft*	Maximalkraft	-	0,97[2]
Ballweitwurf	Schnellkraft/Koordination	-	0,77
Wahlreaktionslauf* (Sternlauf/Buzzertest)	Koordination/Agilität	-	-

Legende: Die ersten acht Testaufgaben sind Teil des *Deutschen Motorik Tests 6–18*; [1] Bös et al., 2009, [2] Artero et al., 2011, [3] Utesch et al., 2015, [4] Klein et al., 2012; *diese beiden Tests *Fuldaer Bewegungs-Checks* wurden nicht in die Validierungsstudien der vorliegenden Arbeit miteinbezogen

Die beiden Hauptgütekriterien Reliabilität und Objektivität der acht Testaufgaben des Deutschen Motorik Tests 6–18 wurden bereits in einer Reihe von Studien verschiedener Autoren überprüft (Bös et al., 2009; Klein et al., 2012; Utesch et al., 2015). Die Validität der Testverfahren ist jedoch noch nicht hinreichend nachgewiesen. Während Bös et al. (2009) sich vor allem auf die inhaltliche Validität konzentriert und hierfür Expertenratings herangezogen haben (zwei Expertenratings: $M_1 = 1{,}88$ und $M_2 = 2{,}1$; mit Noten von 1 bis 5), verwendeten andere Autoren Korrelationen zur Überprüfung der kriteriumsbezogenen Validität oder konfirmatorische Faktorenanalysen zur Bestimmung der Konstruktvalidität (Bardid et al., 2019). Für die meisten Einzeltests kann innerhalb der beiden letztgenannten Validitätskategorien eine ausreichende bis sehr gute Testvalidität attestiert werden ($r = 0{,}69$).

Stichprobe

Im Rahmen der Kampagne werden jährlich in einem regionalen Talentscreening landkreisweit alle Schüler der jeweiligen zweiten Klassen (Alter: 94,3±5,4 Monate ≈ 8 Jahre) zu den Testungen eingeladen ($N \approx 2.000$). Die unter dem Aspekt der Repräsentativität sehr hohe Teilnehmerzahl genügt mit etwa 90 Prozent Jahrgangsabdeckung allgemein hohen Ansprüchen − auch im Vergleich zu motorischen Untersuchungen in Großstädten wie z.B. Düsseldorf, Berlin, Hamburg oder Frankfurt am Main. Die Stichprobe wird daher einer regionalen Gesamterhebung aller Schüler der zweiten Klassenstufe gerecht und ist demensprechend als sehr heterogen einzuschätzen. Die Stadt Fulda entspricht dabei mit 67.980 Einwohnern ($n_{♂} = 33.388$, $n_{♀} = 34.592$; $n_{Ausländer} \approx 11.150$; Einwohner/km^2 = 653; Anteil an Personen mit Migrationshintergrund ≈ 28%) und einer Fläche von 104 km^2 einer Mittelstadt (20–100 Tsd. Einwohner) und befindet sich mit seiner Einwohnerzahl auf Platz 128 der insgesamt 666 Mittel- und Großstädte Deutschlands (Der Magistrat der Stadt Fulda, 2019; Statistisches Bundesamt, 2021). Lässt man Kleinstädte außen vor (ca. 1.400), so entspricht die Einwohneranzahl Fuldas nahezu dem Durchschnitt der Mittel- und Großstädte Deutschlands ($N_{\phi} \approx 73.000$). Der Landkreis Fulda kommt zusätzlich auf ca. 220.000 Einwohner bei einer Fläche von nahezu 1.400 km^2. Die gesamte Region Fulda erreicht somit eine Gesamteinwohneranzahl von ca. 290.000 Personen.

Seit dem Beginn der Kampagne im Jahr 2010 wurden bisher rund 20.000 Zweitklässler der gesamten Bildungsregion (~6% der regionalen Bevölkerung) innerhalb der Kampagne getestet. Die Stichprobe der publizierten Validitätsstudien setzt sich entsprechend einer retrospektiven Betrachtung der Talentprognosen jedoch nur aus den Probanden der frühen Kampagnenjahre zusammen. Die tatsächlich zur Verfügung stehende Stichprobe orientiert sich zusätzlich an der jeweils untersuchten Sportart, deren Talentempfehlungen validiert wurden.

Kriterium

Als Kriterium dient das von den Teilnehmern in jeweils ausgewählten Altersbereichen (Initialalter [U9–U11], Basisalter-1 [U12–U13], Basisalter-2 [U14–U15], Anschlussalter-1 [U16–U17], Anschlussalter-2 [U18–U19], Finalalter [U20+]) erreichte individuelle Leistungsniveau. Um einen Vergleich auch verschiedener Sportarten zu gewährleisten, wurde dafür ein eigenes standardisiertes, 20-stufiges Skalenniveau entwickelt, das die verschiedenen Wettkampfklassen adäquat darstellt (Tab. 9). Inhaltlich unterteilt es sich in lokale, regionale, nationale sowie internationale Leistungserfolge. Das höchste Leistungslevel (Level 20) entspricht dabei einem Olympiasieger oder Weltmeister.
Gerade im jüngeren Altersbereich gibt es allerdings häufig noch kein differenziertes Ligen- oder Wettkampfsystem, das erfolgreiche von nicht-erfolgreichen Sportlern klar trennen kann. Diese Problematik wurde ebenfalls bei der Erstellung der Leistungsskalen berücksichtigt und eine entsprechende Eingruppierung der Juniorenerfolge vorgenommen.

Tab. 9. *Leistungslevel entsprechend der erreichten Wettkampferfolge*

Level	Erfolgsbasierte Leistungsskala	Success-based performance scale
0	Kein Wettkampfsport	No show
1	Vereinsregistrierter Nachwuchssportler	Club-registered youth athlete
2	Wettkampfteilnahme in der untersten lokalen altersbezogenen Leistungsklasse (z.B. Kreisliga); lokal erfolgreiche Wettkampfteilnahme im Altersbereich bis zur U11	Competition participation in the lowest local age-related performance class; local successful competition participation in the age range up to U11
3	Wettkampfteilnahme in einer mittleren altersbezogenen Kreisklasse oder Kreisoberliga; lokal erfolgreiche Wettkampfteilnahme im Altersbereich bis zur U13	Competition participation in a middle age-related district class; locally successful competition participation in the age range up to U13
4	Wettkampfteilnahme in der höchsten altersbezogenen Bezirks- oder Gauklasse (z.B. Bezirksliga); Finalteilnahme auf der Bezirksebene im Altersbereich bis zur U15	Competition participation in the highest age-related district class; Final participation at the district level in the age range up to U15
5	Wettkampfteilnahme in der höchsten altersbezogenen Landesklasse (z.B. Verbandsliga) oder Finalteilnahme auf der Bezirksebene im Altersbereich bis zur U17	Participation in competitions in the highest age-related state class or participation in finals on the county level in the age range up to U17
6	Medaille auf Landesebene U17 (z.B. Hessenliga oder Hessische Meisterschaft); Teilnahme am Wettkampfbetrieb der höchsten altersbezogenen Landesgruppenklasse (z.B. Regionalliga) oder an der altersbezogenen Landesgruppenmeisterschaft (Süddeutsche Meisterschaft)	Medal at the regional level U17; Participation in the competition of the highest age-related state group class (e.g. Hessian league) or in the age-related state group championship
7	Teilnahme am U17-Wettkampfbetrieb der höchsten altersbezogenen nationalen Leistungsklasse (z.B. B-Junioren-Bundesliga) oder dem internationalen altersbezogenen Meisterschaftsbetrieb (z.B. Nominierung für U15-Nationalmannschaft); Gewinn der nationalen Meisterschaft im Altersbereich bis einschließlich U15 (Platz 1–3); Lehrgangsteilnahme Jugendnationalmannschaft (U17)	Participation in the U17 competition of the highest age-related national performance class (e.g., youth national league) or the international age-related championship (e.g., nomination for U15 national team); winning the national championship in the age range up to and including U15 (place 1–3); participation in a training course for the youth national team (U17).
8	Teilnahme am U19-Wettkampfbetrieb der höchsten altersbezogenen nationalen Leistungsklasse (z.B. A-Junioren-Bundesliga): Wettkampfteilnahme am internationalen Meisterschaftsbetrieb (z.B. Nominierung für U17-, U19- oder U21-Nationalmannschaft); Gewinn der nationalen Meisterschaft im Altersbereich ab U17 (Platz 1–3); Senioren: 2. Bundesliga	Participation in the U19 competition of the highest age-related national performance class (e.g., junior national league): Participation in the international championship (e.g., nomination for U17, U19 or U21 national team); winning the national championship in the age range from U17 (place 1–3); seniors: 2nd national league

Level	Erfolgsbasierte Leistungsskala	Success-based performance scale
9	Junioren Europa Meisterschaft (EM) / Junioren Weltmeisterschaft (WM) – Platz 9–16; Senioren: 1. Bundesliga / Deutschen Meisterschaft (DM) – Platz 9–20	European Youth Championships / Youth World Championships – place 9–16; Seniors: 1st national league / German Championships (GC) – place 9–20;
10	Junioren-EM / Junioren-WM – Platz 4–8; Senioren: 1. Bundesliga / DM – Platz 5–8	European Youth Championships / Youth World Championships – place 4–8; Seniors: 1st national league / GC – place 5–8
11	Junioren-EM / Junioren-WM – Medaille; Senioren: 1. Bundesliga / DM – Platz 3–4	Medal at European Youth Championships or Youth World Championships; Seniors: 1st national league / GC – place 3–4
12	Senioren: 1. Bundesliga / DM – Platz 2	Seniors: 1st national league / GC – place 2
13	Senioren: 1. Bundesliga / DM – Platz 1	Seniors: 1st national league / GC – place 1
14	B-Nationalkader / A-Länderkampf; Start bei EM / WM / Olympische Spiele (OS)	B-national squad / A-country competition; start at European Championships (EC) / World Championships (WC) / Olympic Games (OG)
15	A-Nationalkader; EM – Platz 6–8; WM / OS – Platz 9–16	A-national squad; EC – place 6–8; WC / OG – place 9–16
16	EM – Platz 4–5; WM / OS – Platz 5-8	EC – place 4-5; WC / OG – place 5–8
17	EM – Platz 3; WM / OS – Platz 4	EC – place 3; WC / OG – place 4
18	EM – Platz 2; WM / OS – Platz 3	EC – place 2; WC / OG – place 3
19	Europameister; WM / OS – Platz 2	European Champion; WC / OG – Place 2
20	Weltmeister; Olympiasieger	World Champion; Olympic Champion

Die Leistungserfolge der Teilnehmer wurden mittels eines mehrperspektivischen Verfahrens erhoben, das sich neben aktuellen Informationen der in Frage kommenden Berichterstattungen auch auf Meldelisten (Wettkampfmeldungen, Stützpunktlisten, Kaderaufstellungen, etc.) stützt. Diese Meldelisten wurden je nach Sportart direkt über den Veranstalter der jeweiligen Wettkämpfe oder über offizielle Internetauftritte der Vereine und Verbände erhoben. Zusätzlich wurden einzelne Leistungen erfolgreicher Sportler in persönlichen Befragungen der jeweiligen Stützpunktleiter und Trainer überprüft und ergänzt. Gerade die höchsten Leistungsklassen können somit als vollständig angesehen werden, wohingegen sich in den unteren Ligen und Wettkampfstrukturen aufgrund der Menge an Ausrichtern und der häufig wechselnden Teilnehmer eine genaue Aufschlüsselung der jeweiligen Wettkampferfolge schwieriger gestaltet. Aus diesem Grund wurden die Leistungsklassen (wie in Kap. 3.6 beschrieben) üblicherweise in zwei grundlegende Erfolgslevel („Hochleister“ und „Niedrigleister“) eingruppiert, da es in der Talentforschung vor allem interessiert, welche Sportler die höchsten bzw. das höchste Leistungslevel erreichen konnten oder können werden. Neben der erreichten Leistung (in Leveln) der Probanden wurde auch das zu diesem Zeitpunkt erreichte Alter (Leistungsalter; in Monaten) festgehalten.

Selektionsrate und Prognosemethode:

In allen Studien wurden – mit Ausnahme der ersten Studie aus 2017 – unterschiedliche Selektionsraten mithilfe eines Sensitivität-Spezifitäts-Diagramms oder anhand der AUC berücksichtigt. Exemplarisch wurde zudem die Klassifikationsgüte an einzelnen, zumeist standardmäßig durch die gewählten statistischen Methoden und die verwendeten Statistikprogramme voreingestellten Selektionsraten aufgezeigt. Die Prognosemethoden setzten sich hierbei stets aus ausgewählten linearen (z.B. Diskriminanzanalyse) und nicht-linearen Methoden (z.B. Multilayer Perceptron) zusammen.

4.1 Talentscreening und Talentorientierung: Zur Validität von Bewegungs-Checks im Grundschulalter

(Hohmann*, Fehr, Siener, Hochstein, 2017)

Im Kontext der zu diesem Zeitpunkt aufgekommenen Diskussion einer wissenschaftlich gestützten und fundierten Talentidentifikation soll der Beitrag den allgemeinen Diskurs weiter vorantreiben und einen Übergang von den bisher dominierenden subjektiven und meist nur punktuellen Interpretationen von Einzelfällen (Güllich, 2013; Issurin, 2017; Joch, 1994) hin zu einer objektiven, längsschnittlichen und empirisch fundierten Befundinterpretation aufzeigen (vgl. Höner & Votteler, 2016). Die vorliegende gemischte Quer- und Längsschnittstudie dient daher der allgemeinen Einführung eines ersten Validierungskonzepts für Bewegungs-Checks. Die hierfür notwendigen Arbeitsschritte zielen sowohl auf die Sportartorientierung als auch auf die Talentidentifikation (i.e.S.) ab. Insgesamt werden vier Schritte vorgestellt (Abb. 19), die jeweils unterschiedliche Validierungsbereiche ins Auge fassen und anhand der frühen Daten (bis einschließlich 2016) des Fuldaer Bewegungs-Checks (FBC) verdeutlicht werden (Hohmann et al., 2015).

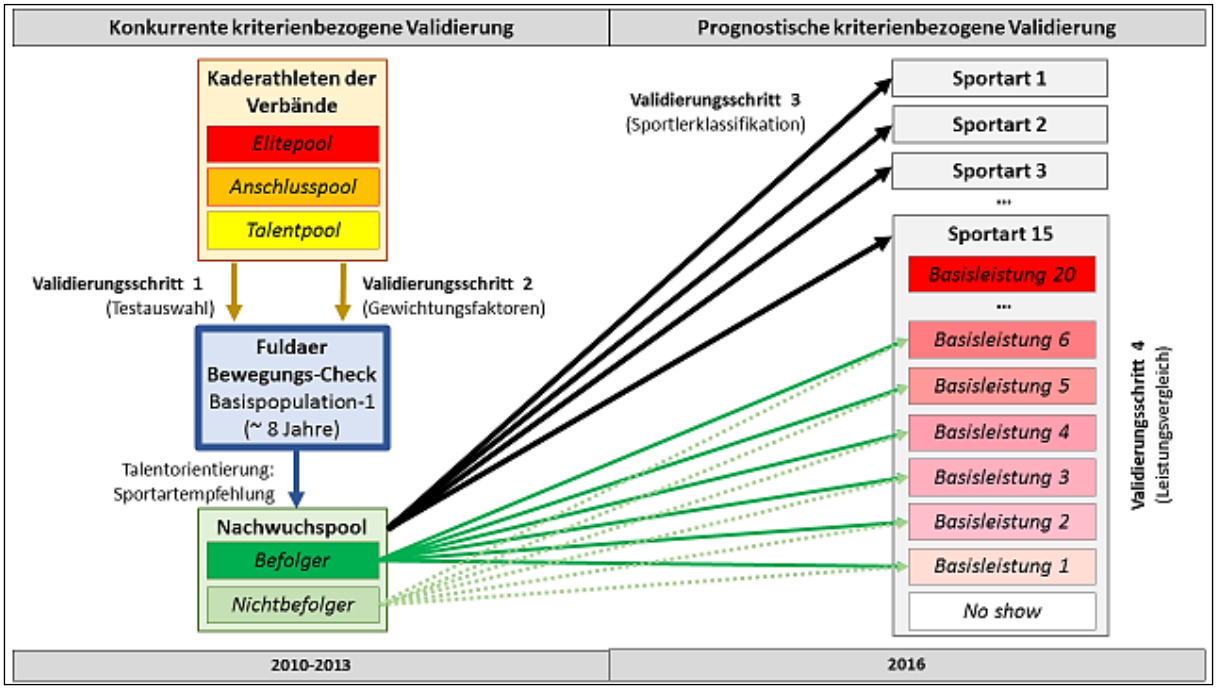

Abb. 19. Erstes Modell zur Validierung von Talentidentifikation

Validität der Testauswahl (Prädiktorvalidierung)

Entsprechend der in Kap. 3.1 vorgestellten Möglichkeiten der Validierung der Prädiktor-Auswahl werden hier exemplarisch die elf Testaufgaben des FBCs auf ihre Gültigkeit hin überprüft. Dafür wurden die Testaufgaben bei älteren, schon erfolgreichen Sportlern (Kaderathleten) unterschiedlicher Sportarten eingesetzt und anschließend eine Klassifikationsanalyse auf die Herkunftssportart angewendet. Sollte es mithilfe der gewählten Prädiktoren nicht möglich sein, Sportlern querschnittlich ihre aktuell betriebene Sportart zuzuweisen, so scheint auch eine prognostische Sportartempfehlung bzw. Talentorientierung für jüngere Probanden anhand dieser Prognosemethode nicht zielführend. Das Vorgehen entspricht somit inhaltlich einer „unechten Quasiprognose“ (Hohmann et al., 2015). Die Ergebnisse dieser und teilweise auch nachfolgender Klassifikationsanalysen (verwendet wurden Diskriminanzanalyse [DA], Multilayer Perceptron [MLP] und Kohonen Feature Map [KFM]) wurden bereits in Hohmann et al (2015) vorgestellt.

Validität der Sportartgewichtungen (Methodenvergleich)

Untersucht wurde in diesem Schritt, ob Gewichtungen einen Mehrwert für die querschnittlichen Klassifikationsprognosen der Kaderathleten darstellen. Zunächst wurde daher die Gesamttreffergenauigkeit der Klassifikationsmethoden anhand ungewichteter Merkmalsausprägungen berechnet, um sie anschließend mit der Genauigkeit einer Klassifikation anhand gewichteter Merkmalsausprägungen zu vergleichen. Da die Ergebnisse der gewichteten Prädiktoren besser ausfallen als die der ungewichteten Prädiktoren, lässt sich ein Vorteil von Gewichtungen in der Talentorientierung vermuten.

Validierung der Sportlerklassifikation (Sportartorientierung)

Während sich die ersten beiden Prüfschritte an den Testergebnissen bereits erfolgreicher Sportler orientierten, findet im dritten Validierungspunkt eine längsschnittliche Betrachtung und erste prognostische Validierung der Talentorientierung statt. Als Kriterium wurde hierbei die 3 bis 6 Jahre nach der Testung jeweils ausgeübte Sportart verwendet, welche es nun anhand des Merkmalsprofils aus der zweiten Grundschulklasse zu klassifizieren gilt. Die Prognosegüte wird dabei mithilfe der Gesamttreffergenauigkeit ermittelt.

Validierung der sportartspezifischen Talentprognose (Talentidentifikation)

Neben der in Schritt drei beschriebenen korrekten Zuordnung eines Probanden in seine spätere Sportart, ist ebenfalls von Interesse, auf welchem Leistungslevel diese Sportart später ausgeübt wird bzw. werden kann. Als Kriterium dient somit nicht eine Sportartcodierung, sondern die in Tabelle 9 vorgestellten Leistungslevel. Aus diesem Grund wurde eine Korrelation zwischen dem Prognosewert (gewichtet und ungewichtet) und den erreichten Leistungsleveln berechnet.

Die Ergebnisse der vier Validierungsschritte deuten insgesamt auf eine positive Gesamtvalidität des Fuldaer Bewegungs-Checks hin. Nichtsdestotrotz sind die Befunde aufgrund des geringen Prognosezeitraums und der hinzukommenden geringen Teilnehmeranzahl noch limitiert.

4.2 Prognostic validity of talent orientation in soccer

(Hohmann, Siener*, He, 2018)

Im Kontext der Talentidentifikation existieren mit wenigen Ausnahmen (Carling et al., 2012; Höner & Votteler, 2016; Le Gall et al., 2010) überwiegend Studien mit querschnittlichem Design oder nur vergleichsweise kurzen Zeitintervallen von weniger als drei Jahren. Obwohl die auch durch diesen Umstand bedingte Befundlage der Erfolge von langfristigen Talentprognosen noch nicht gänzlich geklärt werden konnte, gibt es Zweifel an der prognostischen Validität der mithilfe von sportmotorischen Tests in jungen Jahren erhobenen Prognosen zukünftiger Spitzensportler (Carling & Collins, 2014; Lidor, 2007). Die vorliegende Studie hatte daher das Ziel, die prognostische Validität von mittelfristigen (2–6 Jahre) Talentprognosen der Sportart Fußball zu überprüfen.

In der Studie wurde zunächst das Merkmalsprofil eines typischen Fußballers erfasst und anschließend die Vorhersagekraft der verwendeten Prädiktoren (vgl. Kap 3.1) mithilfe von ORs geprüft. Daraufhin wurde anhand einer Varianzanalyse untersucht, ob sich die durch die gewählten Kriteriums-Partitionen entstandenen Leistungsgruppen bereits in ihren U9-Testleistungen signifikant voneinander unterscheiden. Im Fokus des Artikels stehen jedoch die binären Talentprognosen von 316 Fußballspielern.

Als Methode wurde neben einer Diskriminanzanalyse (leave-on-out-CV) mit dem MLP (Partitionsaufteilung von Training–Validation–Test: 80–10–10) auch ein nicht-lineares Klassifikationsverfahren eingesetzt.
Die berechneten ORs, mit einer überdurchschnittlichen Merkmalsausprägung später das nationale Leistungslevel zu erreichen, lagen für alle Testaufgaben über eins. Das Ergebnis deutet darauf hin, dass alle Prädiktoren geeignet scheinen, um eine Talentprognose zu erstellen. Es zeigte sich zudem, dass Sportler, die später ein höheres Leistungslevel erreichen, bereits in der U9 in den meisten Testaufgaben den später schwächeren Fußballspielern überlegen sind – was wiederum für das gewählte Kriterium und eine vielversprechende Klassifikationsmöglichkeit spricht. Die Talentprognose erreicht je nach gewähltem Selektionswert eine Gesamttreffergenauigkeit von knapp über 70%. Die erreichte Validität dieser Talentidentifikation mittels sportmotorischer Tests liegt somit im oberen Durchschnitt anderer vergleichbaren Studien (Güllich & Krüger, 2019).

Die vorliegende Studie betrachtet gleich vier Teilperspektiven der Validierung von Talentidentifikation: Erstens wird ein Merkmalsprofil von Fußballern erstellt und Prädiktoren werden im Hinblick auf ihre Prognoseeignung bewertet (Leyhr et al., 2018), zweitens werden die Einteilungsstufen des Kriteriums hinsichtlich ihrer Trennkraft der verschiedenen Leistungsgruppen abgeschätzt, drittens werden zwei verschiedene Prognosemethoden eingesetzt und hinsichtlich ihrer Prognosegüte nebeneinandergestellt (Mostaert et al., 2020) und viertens die prognostischen Kennwerte von Sensitivität und Spezifität hinsichtlich verschiedener Selektionsraten bestimmt. Gerade der letzte Punkt ist bisher mit wenigen Ausnahmen (Höner & Votteler, 2016) in den Studien zur Validierung von Talentidentifikationen noch nicht weit verbreitet und stellt damit eine wichtige Ergänzung in diesem Bereich dar. Insgesamt fehlt jedoch eine erweiterte Einschätzung der Nützlichkeit (Youden`s *J*, RIOC, MCC, AUC, etc.) der erhaltenen Ergebnisse.

4.3 Talent orientation: the impact of motor abilities on future success in table tennis

Siener* & Hohmann, 2019

Obwohl gerade die deutschen Tischtennis-Sportler gegenwärtig eine gute Figur machen und mit *Timo Boll* sogar zwischenzeitlich die Weltrangliste der Herren angeführt haben (International Table Tennis Federation [ITTF], 2020), liegen vor allem für den Kinder- und Jugendbereich nur wenige (deutsche) Studien zur Talentidentifikation in dieser Sportart vor (Faber, et al. 2014). Der Umstand lässt vermuten, dass sich die Identifikation junger Nachwuchstalente im Tischtennis bisher schwieriger gestaltet als bspw. in anderen Sportarten wie Fußball (Höner et al., 2019; Sieghartsleitner, 2020) oder Basketball (Baker et al., 2020). Nicht zuletzt ist diese Tatsache der Grund, warum mit dem vorliegenden Artikel die Validität einer Talentidentifikation im Tischtennis exemplarisch geprüft wurde.

Als Grundlage diente eine Stichprobe des Fuldaer Bewegungs-Checks von 225 Junioren-Tischtennisspielen der Altersklassen U13 bis U17. Als Kriterium wurde die höchste

erreichte Jugendleistung gewählt, die wiederum in vier Leistungslevel bzw. Leistungsgruppen (orientiert an Tab. 9) eingeteilt werden konnte.
Inhaltlich strukturiert sich der Artikel in vier Validierungsbereiche: Prüfung auf Sportartprofile (Sportartorientierung), Prüfung von Unterschieden zwischen den Kriteriums-Leveln, Bedeutung der Prädiktoren für die Talentprognose und Klassifikation der Leistungsgruppen.

Zunächst wurde geprüft, ob Tischtennisspieler ein einheitliches Profil aufweisen und ob dieses im Sinne einer möglichen Talentorientierung von Profilen anderer Sportarten abweicht. In einem zweiten Schritt wurden die Unterschiede der anhand des gewählten Kriteriums entstandenen Leistungsgruppen geprüft, um festzustellen, ob sich erfolgreiche Juniorensportler von weniger erfolgreichen Juniorensportlern bereits in der U9 unterscheiden. Um signifikante Gruppenunterschiede in den einzelnen Testaufgaben nachzuweisen, wurden hierfür ANOVAs berechnet und ein post-hoc-Vergleich der vier Leistungsgruppen angestellt. Sollte keiner der Tests eine Unterscheidung der Leistungsgruppen ermöglichen, so wäre auch eine mögliche Klassifikation von zukünftig erfolgreichen Sportlern fraglich.
Um einen Einblick in die Bedeutung der einzelnen Tests bzw. Prädiktoren auf die späteren Spielerfolge zu erhalten, wurden ORs berechnet und geprüft, mit welcher Chance ein Teilnehmer mit einer überdurchschnittlichen Einzeltestleistung ($z \geq 1$) das höchste Leistungslevel erreicht. Neben den ORs wurde ebenfalls die Verteilung dieser *best movers* unter den Sportlern der höchsten Leistungsgruppe bestimmt.
In einem letzten Schritt wurden schließlich die Validitätskennwerte (Sensitivität, Spezifität und Gesamttrefferquote) dreier Prognosemethoden berechnet.

Die Ergebnisse verdeutlichen, dass Tischtennisspieler zwar ein relativ einheitliches Merkmalsprofil aufweisen, bezüglich ihrer Testwerte im Vergleich zu anderen Sportartgruppen in den meisten Testaufgaben jedoch schwächer abschneiden. Die geringen Ausprägungen der meisten konditionellen Testwerte hatten sich auch bereits in ähnlichen Studien gezeigt (K. Robertson et al., 2018). In den koordinativ geprägten Testaufgaben konnten Tischtennisspieler hingegen gute Ergebnisse erzielen (siehe auch Nikolić et al., 2014). Dementsprechend sind in diesen Testaufgaben die berechneten ORs auch am höchsten. Die ANOVA macht deutlich, dass die gewählte Level-Einteilung des Kriteriums für das Vorhaben tendenziell geeignet scheint. Jedoch war nur in fünf der neun Testaufgaben die höchste Leistungsgruppe auch mit dem höchsten Gruppenmittelwert vertreten. Die Gesamttreffergenauigkeit der drei Prognosemethoden liegt jeweils im oberen Drittel der möglichen Skala. Das künstliche neuronale Netz schneidet hierbei mit 78,4% korrekten Treffern am besten ab. Die Diskriminanzanalyse kommt auf nur 74,8%, und die Gesamttreffergenauigkeit der gewählten Gewichtung liegt unterhalb von 74%. Alle drei Kennwerte liegen jedoch über der in anderen Studien erreichten Marke von 70% (Güllich, 2019).

Inhaltlich werden in dieser Studie die Validierungsbereiche der vorangegangenen Studie aufgegriffen und auf die Sportart Tischtennis übertragen. Zusätzlich wird versucht, die

Ergebnisse von Sensitivität und Spezifität in Abhängigkeit einer rein zufälligen Klassifikation zu bewerten. Hierfür kommt mit Youden`s *J* erstmals ein Kennwert zur Nützlichkeit der Talentprognose zum Einsatz.

4.4 Talent Identification in Youth Soccer: Prognosis of U17 Soccer Performance on the Basis of General Athleticism and Talent Promotion Interventions in Second-Grade Children

Hohmann* & Siener, 2021

Dieser Artikel basiert auf einer Follow-Up-Studie des bereits 2018 veröffentlichten Artikels *Prognostic validity of talent orientation in soccer*. Die entsprechende Studie verfügt – zwei Jahre später – nun neben einer größeren Stichprobe auch über einen ausgedehnteren Prognosezeitraum, da Vorhersagen bis zur U17-Juniorenleistung (zuvor U15) bestimmt werden konnten.
Der Artikel möchte somit die von einigen Autoren (Carling & Collins, 2014; Côté et al., 2009) infrage gestellte Vorhersagbarkeit späterer Erfolge über einen längeren Zeitraum überprüfen. Im Fokus steht hierbei verstärkt die Aussagekraft der einzelnen Prädiktoren und weniger die bereits 2018 erhobene Prognoseleistung unterschiedlicher Prognosemethoden (Gewichtung, DA und MLP; siehe Kapitel 4.2). Neben den allgemeinen motorischen Tests des Fuldaer Bewegungs-Checks bzw. des Deutschen Motorik-Tests 6–18 (DMT) wurden daher erstmals auch Prädiktoren der Talentförderung in das Studiendesign miteinbezogen.

Die U17-Juniorenleistung von 502 Fußballern (im Vergleich zu 316 U15-Fußballern zuvor) wurde in insgesamt fünf Level-Einteilungen (Kriteriumsvariablen) eingeteilt. Als Prädiktoren wurden ergänzend zu den bereits bekannten motorischen U9-Testaufgaben des DMT sowohl die vier Merkmale „(i) Club Diversity, (ii) local talent promotion course, (iii) sport class, and (iv) local soccer training center“ der Talentförderung, als auch die in der U13 erreichten Juniorenerfolge eingesetzt. Entsprechend des vorangegangenen Artikels wurde nun zunächst ein sportartspezifisches Merkmalsprofil erstellt. Es wurde zudem die gewählte Kriteriumseinteilung mithilfe einer ANOVA an den einzelnen Prädiktoren geprüft und ein Eindruck gewonnen, welche Prädiktoren für eine Unterscheidung der Kriteriumsgruppen infrage kommen könnten. In einem zweiten Schritt wurde die Aussagekraft der einzelnen Prädiktoren anhand von ORs und einer Pfadanalyse betrachtet (siehe auch Höner et al., 2017), um in einem dritten Schritt abschließend die Kennwerte einer Gewichtung (Prognosemethode) zu bestimmen und zu bewerten.

Insgesamt spiegeln die erzielten Ergebnisse in großen Teilen die Befunde anderer Studien wider (Figueiredo et al., 2009; Reilly et al., 2000; Sarmento et al., 2018). Insbesondere die Prädiktoren Sprintgeschwindigkeit und Ausdauer scheinen demnach für eine Talentidentifikation geeignet (Gonaus & Müller, 2012; Höner & Votteler, 2016; Le Gall et al., 2010). Im Bereich der Prädiktoren der Talentförderung zeigt das Merkmal „DFB Trainingscenter“ die höchsten OR-Werte. Entgegen der Schlussfolgerungen anderer

Autoren (LaPrade et al., 2016) konnten sich jedoch die Vorteile einer frühen Ausübung von mehreren Sportarten nicht gänzlich bestätigen.

Im Hinblick auf das übergeordnete Konzept der Validierung von Talentidentifikation greift dieser Artikel in der vorgestellten Publikationsreihe erstmals Prädiktoren der Talentförderung auf und betrachtet zusätzlich die erbrachte Spielleistung im Juniorenalter. Durch dieses Vorgehen wird zwar die langzeitliche Betrachtungsweise durch die Hinzunahme erst später auftretender Prädiktoren zeitlich eingeschränkt, jedoch mit dieser Prädiktorerweiterung dem dynamische Talentbegriff Rechnung getragen. Gleichzeitig werden die Fördermechanismen im Deutschen Jugendfußball und ihre Effektivität näher beleuchtet.
Als weitere Neuerung kam in dieser Studie erstmals eine Pfadanalyse zum Einsatz, um den Zusammenhang einzelner Prädiktoren mit der abschließend erreichten U17-Spielleistung zu vergleichen. Im Bereich der Nützlichkeit der Talentprognose wurde hier der bereits in der vorherigen Studie eingeführte Youden Index wieder aufgegriffen und zusätzlich ein Vergleich zur Grundrate angefügt: „In case of such a cutoff limit, the success rate in the group of the 120 drafted U9 talents would be 19.16%. (…) Comparatively, the base rate in the tested 502 U9 soccer players was 7.97%" (S. 11). Somit lässt sich zum ersten Mal auch ein direkter Benefit der Talentprognose bei einer relativ praxisnahen Empfehlungsrate ausmachen.

4.5 Early talent identification in tennis: A retrospective study

Siener*, Ferrauti, & Hohmann, 2022

Eine zunächst als ein Artikel geplante Studie zum Erfolg im Junioren-Tennis wurde aufgrund ihres Umfangs in zwei Artikel geteilt und inhaltlich sowie methodisch deutlich gekürzt – das Ergebnis ist der vorliegende erste Artikel zur frühen Talentidentifikation von Junioren-Tennisspielern. Im Fokus stehen das tennisspezifische Merkmalsprofil junger Sportler sowie der Zusammenhang von Prädiktor und Kriterium. Der zweite Artikel umfasst dann die genauere Betrachtung der Talentprognosen.

Insgesamt 174 heutiger Junioren-Tennisspieler (U12–U18) wurden bereits in der U9 im Rahmen des Fuldaer Bewegungs-Checks getestet. Diese Stichprobe besteht aus 16 stärkeren Spielern mit einer Platzierung auf der nationalen Junioren-Rangliste des *Deutschen Tennis Bundes* und 158 schwächeren Tennisspielern ohne eine solche Ranglistenplatzierung. Das Erreichen der nationalen Rangliste wurde somit als Kriterium angesetzt. Wie bereits zuvor dienten die motorischen Tests des Fuldaer Bewegungs-Checks als Prädiktoren. Prädiktoren der Talentförderung wurden nicht verwendet. Mithilfe von *t*-Tests und einer MANCOVA wurde überprüft, ob sich die beiden Junioren-Leistungsgruppen hinsichtlich ihrer in der U9 erbrachten Testleistung signifikant unterscheiden. Die Ergebnisse wurden zudem in einem Merkmalsprofil und einer Boxplot-Übersicht grafisch verdeutlicht. Weiterhin wurden die Zusammenhänge von Prädiktor und Kriterium anhand einer Korrelationsanalyse untersucht.

Die Ergebnisse zeigen, dass sowohl bessere als auch schwächere Tennisspieler ein ähnliches Merkmalsprofil aufweisen. Die Ausprägungen der Merkmale sind bei den später erfolgreicheren Sportlern jedoch deutlich höher als die Ausprägungen der später nicht erfolgreichen Tennisspieler.
In der MANCOVA zeichnete sich zudem ab, dass sich beide Leistungsgruppen in den meisten Tests signifikant unterscheiden. Dieser Umstand stimmt mit den Befunden anderer Studien überein (z.B. Kramer et al., 2017; Roetert et al., 1992; Ulbricht et al., 2016). Die berechneten Korrelationswerte gingen jedoch nicht über einen moderaten Wert hinaus. Dies muss dennoch nicht gleichzeitig für eine schwache Prognosegüte im Allgemeinen stehen, da für die Talentprognose-Erstellung nicht nur einzelne Tests, sondern die Gesamtheit aller Testvariablen herangezogen wird (siehe Kap. 3.6).

Im Hinblick auf das übergeordnete Konzept der Talentvalidierung erarbeitet der vorliegende Artikel die Grundlagen der in einem nachfolgenden Artikel überprüften Talentprognose und betrachtet im Kern, ob und in welcher Art es einen Zusammenhang zwischen Prädiktor und Kriterium gibt. Im Fokus stehen also keine Prognosen oder Klassifikationen, sondern die Erkenntnis, welche Prädiktoren in welcher Güte für eine Prognose geeignet wären (vgl. Kap. 3.1). Die verwendeten Methoden entsprechen hierbei einer eher klassischen Herangehensweise, wie sie im Bereich des Tennissports häufig angewendet wird (vgl. Faber et al., 2015; Fett et al., 2017; Filipcic et al., 2010; Roetert et al., 1992; Ulbricht et al., 2016), gewährleisten dadurch jedoch auch eine hohe Vergleichbarkeit.

4.6 Prognostic Validity of Statistical Prediction Methods Used for Talent Identification in Youth Tennis Players Based on Motor Abilities

Siener*, Faber, & Hohmann, 2021

Dieser Artikel beinhaltet den zweiten Teil der Tennis-Studie und befasst sich mit der prognostischen Validität der Talentprognosen unterschiedlicher Prognosemethoden.
In den meisten bisherigen Studien aus dem Bereich der Talentidentifikation lag der Fokus verstärkt auf der Bestimmung möglicher Talent-Prädiktoren und ihrer Güte hinsichtlich einer möglichen Talentprognose (Faber, Bustin et al., 2016; Filipcic & Filipcic, 2005; Kovacs, 2007). Eine Einschätzung der Nützlichkeit von Talentprognosen war bis auf wenige Ausnahmen (z.B. Höner & Votteler, 2016; Schorer et al., 2017) – und dabei zumeist auch nur in Ansätzen – jedoch kaum in den bisherigen Talentstudien zu finden. Aus diesem Grund werden hier erstmals die bereits in anderen Disziplinen, z.B. der Psychologie (Copas & Loeber, 1990) oder Kriminalistik (Cairney & Streiner, 2011), erfolgreich eingesetzten Kennwerte zur Bestimmung der Nützlichkeit von Diagnosen bzw. Klassifikationsmethoden herangezogen. Ziel der Studie ist es, die Güte und Nützlichkeit verschiedener Talentprognosen zu vergleichen und daran die Validität der Talentidentifikation zu bemessen.

Als Grundlage der Studie diente die bereits zuvor beschrieben Stichprobe aus 174 Junioren-Tennisspielern, die Prädiktoren und das Kriterium wurden ebenfalls entsprechend

des vorangehenden Artikels gewählt. Mithilfe einer Merkmalsgewichtung, einer Diskriminanzanalyse, einer logistischen Regression und eines künstlichen neuronalen Netzes wurden für jeden Teilnehmer vier Talentprognosen erstellt und diese Prognosen anschließend verglichen. Neben den klassischen Kennwerten der Sensitivität, Spezifität, des positiven und negativen prädiktiven Wertes und der Gesamttreffergenauigkeit wurden auch Werte zu Bestimmung der Nützlichkeit der Talentprognose angewandt, darunter der Wert des *Relative Improvement Over Chance* (RIOC oder zu Deutsch RATZ), der F_1-Wert, Youdens Index und die Fläche unter der „receiver operating characteristic"-Kurve (ROC-Kurve). Um einen besseren Überblick über die Verteilung der Prognosen zu erhalten, wurden zusätzlich eine Klassifikationskarte der Einzelklassifikationen aller vier Methoden erstellt und die von den einzelnen Methoden berechneten Prognosewerte der späteren Hochleister verglichen.

Insgesamt konnten alle vier Prognosemethoden gute bis sehr gut RIOC-Werte erzielen. Das neuronale Netz schnitt hierbei mit einem RIOC-Wert von 68% am besten ab. Vergleichbare Studien zu Talentidentifikation anhand sportmotorischer Fähigkeiten kamen mitunter nur auf ca. 55% (Pion et al., 2017; Sieghartsleitner, Zuber, Zibung & Conzelmann, 2019). Auch die anderen Kennwerte der Nützlichkeit unterstrichen die hohe Validität der Tennis-Talentstudie. Es zeigte sich jedoch, dass sowohl Kennwerte als auch Methoden in Abhängigkeit der Zielsetzung der Studien unterschiedliche Aussagekraft haben (Farrington & Loeber, 1989). So deutet eine geringe Sensitivität und ein womöglich damit verbundener geringer Youden Index gerade bei sehr niedrigen oder hohen Selektionsraten nicht gleichzeitig auf eine schwache Prognosegüte hin, wie eine genauere Betrachtung des Ergebnisses der logistischen Regression verdeutlicht. Für eine Selektionsrate nahe der Grundrate scheint dieser Kennwert jedoch wiederum sehr gut geeignet, um die Validität der Talentprognose abschätzen zu können.

Im Hinblick auf das übergeordnete Konzept der Talentvalidierung greift der vorliegende Artikel den letzten Validierungsschritt einer Talentidentifikation auf und vergleicht das Prognoseergebnis hinsichtlich der spezifischen Zielsetzung in Relation zu einem zufälligen Ergebnis. In Anbetracht des aktuellen Diskurses zur Validität der Talentforschung sollen damit nicht nur der Nutzen früher Talentidentifikationen unterstrichen, sondern auch gleichzeitig der methodische Grundriss und die Interpretation von Validierungsstudien in der Talentforschung vorangetrieben werden. Exemplarisch wurden dabei vier Kennwerte der Nützlichkeit von Talentprognosen anhand von vier Klassifikatoren bestimmt und vor dem derzeitigen Forschungshintergrund beleuchtet. Zudem wurde die Problematik einseitiger Validitätskennwerte aufgezeigt.

5. Diskussion

Das Ziel der Artikel war es, die Validität einer frühen Talentidentifikation und -orientierung darzulegen. Zum einen sollte die vorliegende Arbeit Aufschluss über die Güte der von einigen Autoren (LaPrade et al., 2016; Seibold, 2011; Till & Baker, 2020) kritisch gesehenen, anhand sportmotorischer Tests erhobenen, frühen Talenterkennung geben und zum anderen ein sowohl inhaltliches als auch methodisches Konzept für eine erfolgreiche Talentprognose finden (Baker et al., 2018). Gerade die Auswahl passender Prädiktoren (Faber et al., 2014; Pickering et al., 2019) und einheitlicher Erfolgskriterien (Bergkamp et al., 2019; Johnston et al., 2018), wie auch die systematische Auswertung von Talentprognosen, folgte bislang keiner einheitlichen Strukturierung (Sarmento et al., 2018). Für diese Zielsetzung wurden daher exemplarisch anhand einzelner Sportarten die Prädiktoren, Sportartempfehlungen und Talentprognosen des Fuldaer Bewegungs-Checks, einer nachhaltigen Kampagne zur Bewegungs- und Gesundheits-, Sport- und Talentförderung in der hessischen Region Fulda, herangezogen und validiert. Im Rahmen einer jährlichen Vollerhebung der sportlichen Leistungsfähigkeit aller Schüler der zweiten Grundschulklassen wurden in dieser Kampagne seit ihrem Beginn im Jahr 2010 die Merkmalsprofile von bereits rund 20.000 Kindern mithilfe sportmotorischer Testaufgaben überprüft und ausgewertet. Die in dieser Kampagne erhobenen Daten von zwei anthropometrischen Testwerten (Körpergröße und Körpergewicht) und insgesamt neun sportmotorischen Testaufgaben der Bereiche *physical fitness* und *motor competence* dienten in den vorliegenden Artikeln als Prädiktoren der aufgestellten Talentprognosen. Zu den neun sportmotorischen Testaufgaben gehören: 20m-Sprint, Standweitsprung, Liegestütz, Bauchaufzüge, rückwärts Balancieren, seitliches Hin- und Herspringen, Rumpfbeuge, Ausdauerlauf und Ballweitwurf. Darüber hinaus kamen in einer Studie (siehe Kap. 4.4) außerdem Merkmale der Talentförderung und Talententwicklung hinzu. Als Kriterium fungierte über alle Studien und Sportarten hinweg die einheitlich erhobene Juniorenleistung, die anhand einer 20-stufigen Skala eingeschätzt und bewertet wurde. Neben linearen Klassifikatoren wurden in den Studien zumeist auch nichtlineare Methoden eingesetzt, um eine Talentprognose zu erstellen.

Die Artikel selbst bauen methodisch aufeinander auf und durchliefen einen ständig fortschreitenden Prozess der strukturellen Verbesserung des zugrundeliegenden Studiendesigns (sowohl auf inhaltlicher als auch methodischer Ebene). Damit entwickelte sich auch das übergeordnete Konzept der Validierung von Talentidentifikation, das erst mit dem letzten Artikel seinen Abschluss fand. Der Fokus einzelner Artikel liegt allerdings zumeist nur auf bestimmten Bereichen des Validierungsschemas, weshalb sie in ihrer Gesamtheit gemeinsam betrachtet werden müssen, um ein vollständiges Bild der Validität der Talentprognosen einzelner Sportarten des Fuldaer Bewegungs-Checks zu erhalten.

Validierungskonzept

Der erste Artikel der vorliegenden Validierungsstudien versuchte ein erstes Validierungskonzept zu erstellen und anhand von vier Schritten die Zielsetzungen der einzelnen

Teilbereiche abzugrenzen. Die vier Teilbereiche des Findens geeigneter Prädiktoren, der Auswahl vielversprechender Klassifikatoren, des Abschätzens der Prognosegüte der Sportartorientierung (vgl. Pion et al., 2015) und der Bewertung einzelner Talentempfehlungen (prognostische kriterienbezogene Validität) wurden auch in darauffolgenden Artikel immer wieder aufgegriffen und optimiert.

Zusätzlich zu den, üblicherweise über mehrere Jahre laufenden Längsschnittstudien erfolgreich eingesetzten, pro- und retrospektiven Ansätzen der Betrachtungen der Prädiktoren des Grundschulalters kamen in diesem Beitrag auch querschnittliche Analysen und sogenannte „unechte Quasiprognosen" zum Einsatz (Hohmann et al., 2015). Beide Varianten wurden später durch „echte" Langzeitprognosen zukünftiger Leistungen von Sportlern ersetzt. Nichtsdestotrotz könnten die vorgeschlagenen Quasiprognosen an Kaderathleten gerade für junge Talentkampagnen mit einem kurzen Prognosezeitraum (Herrmann et al., 2019) eine Möglichkeit darstellen, auf relativ einfache Weise erste Vermutungen zur Auswahl und Güte möglicher Prädiktoren anzustellen (Hohmann & Seidel, 2005; Pfeiffer & Hohmann, 2012).

Die beiden prognostischen Validierungsschritte umfassen im ersten Artikel einen Prognosezeitraum von 3 bis 6 Jahren – später wurden diese Zeiträume sukzessive erweitert und auch das Validierungskonzept ausgebaut. Die Sportartklassifikation und Talentorientierung nahm jedoch in den darauffolgenden Artikeln eine kleinere Rolle ein, da sie vermehrt auf eine erfolgreiche Talentprognose zukünftiger Spitzensportler einzelner Sportarten abzielten. Ein konkreter Vorschlag expliziter Validierungsschritte von Talentidentifikation war bis dato kaum in der vorhandenen Talentliteratur zu finden (in Ansätzen bei Höner & Votteler, 2016) und stellt somit einen wichtigen konzeptionellen Fortschritt im Talentdiskurs dar. Auch die entwickelte 20-stufige Leistungsskala ermöglichte es zukünftigen Studien, eine Einschätzung des Leistungslevels der Spitzensporterfolge der Probanden zu erhalten und einen Vergleich von Wettkampfleistungen über verschiedene Sportarten hinweg zu gewährleisten. Bisher waren solche Leistungsskalen zwar auch bspw. von Pfeiffer und Hohmann (2012) bekannt, die Skalenbereiche erstreckten sich jedoch häufig nur auf bestimmte Ausschnitte der möglichen Wettkampferfolge oder beschränkten sich auf einzelne Sportarten (Hohmann, 2009; Johnston et al., 2018).

Prädiktoren (Predictors)

Die sich anschließenden Artikel betrachteten überwiegend nur noch einzelne Sportarten und konzentrierten sich bis auf den Vergleich von Merkmalsprofilen unterschiedlicher Sportarten hauptsächlich auf die Erkenntnisse zur Talentidentifikation erfolgreicher Juniorensportler. Unter den beobachteten Disziplinen finden sich zwei Beiträge zur Sportart Fußball, zwei weitere zum Tennis sowie ein Artikel zum Tischtennis.

Die beiden Fußballartikel orientieren sich methodisch an der Studie von Höner & Votteler (2016), die über 20.000 U12-Junioren-Fußballer bis zur Altersgruppe U16 bis U19 beobachtet hatten. In ihren Analysen griffen sie auf ein in vier Level gegliedertes

Kriterium zurück (youth national team, regional association, youth academy, not selected) und berechneten die Gruppenunterschiede der verschiedenen Leistungsgruppen in den einzelnen Testaufgaben mithilfe einer ANOVA. Testaufgaben mit hoher Trennschärfe konnten für vielversprechende Talent-Prädiktoren sprechen. In einem zweiten Schritt wurden mittels logistischer Regression ORs berechnet, um wiederum die Güte der Prädiktoren hinsichtlich der späteren Fußballerfolge abschätzen zu können. Abschließend bestimmten sie für verschiedene Selektionsraten Sensitivität und Spezifität einer eigens aufgestellten Talent-Gewichtung. Die Ergebnisse zeigten, dass sich die meisten der gewählten Prädiktoren für eine Talentprognose eignen und die mithilfe der Gewichtung erstellte Talentprognosen deutliche Vorteile gegenüber einer zufälligen Prognose besitzen. Leider fehlt in dieser Studie eine nähere Betrachtung niedriger Selektionsraten, diese wäre jedoch bei der vorhandenen niedrigen Grundrate von 0,9% von Bedeutung gewesen.

Analog zur Studie von Höner und Votteler (2016) wurde auch in den vorliegenden Validierungsstudien im Fußball die Güte der Prädiktoren mittels ANOVA und ORs bestimmt. Die Leistungsgruppen der ANOVA erschlossen sich dabei aus der zuvor angesprochenen 20-stufigen Skala, wobei hier entsprechend der Altersklasse nur die untersten Ebenen zum Tragen kamen. Aufgrund von unterschiedlichen Gruppengrößen wurden mehrere einzelne Level zu insgesamt vier Leistungsgruppe zusammengefasst. Wie auch bei Höner und Votteler (2016) waren die erfolgreichsten Teilnehmer den anderen Sportlern in den meisten Testaufgaben (mitunter auch signifikant) überlegen. Die gewählten Prädiktoren scheinen daher bis auf wenige Ausnahmen (Körpergewicht, Körpergröße und Beweglichkeit) für eine Talentprognose im Fußball geeignet. Außerdem stimmen diese Ergebnisse überwiegend mit den Befunden anderer Studien an älteren Probanden überein: Vergleichbare Studienergebnisse gab es bspw. zur Sprintschnelligkeit (Gravina et al., 2008; Höner & Votteler, 2016; Vaeyens et al., 2006) und Ausdauer (Emmonds et al., 2016; Malina et al., 2007).

Die berechneten ORs aller sportmotorischen Tests lagen für den gewählte Prozentrang von 84% über einem Wert von eins. Dieses Ergebnis verdeutlicht, dass neben dem herausragenden Abschneiden in speziellen Testaufgaben eine allgemeine hohe Leistungsfähigkeit in allen sportlichen Bereichen für einen späteren Hochleister im Fußball erwartet werden muss (Gulbin et al., 2013). Während Studien von Höner et al. (2017) und Opstoel et al. (2015) diese allgemeine Athletik von (späteren) Hochleistern bestätigen konnten, wurde jedoch in Studien von Sieghartsleitner et al. (2019) in einem holistischen Modell ein negativer Zusammenhang von sportmotorischer Leistung und dem späterem professionellen Spielerstatus festgestellt ($OR = 0{,}11$; $p < 0{,}01$). Den Autoren zufolge sei nicht die allgemeine Leistung in den Testaufgaben, sondern vielmehr die *in-game performance* ($OR = 12{,}53$; $p < 0{,}01$) und der *family support* ($OR = 5{,}20$; $p < 0{,}01$) für den späteren Erfolg entscheidend.

Nicht in der Studie von Höner und Votteler (2016) enthalten ist das für die Fußballartikel sowie für den Tischtennis- und Tennisartikel erstellte sportartspezifische

Merkmalsprofil (vgl. Bös & Schlenker, 2011; Hohmann & Seidel, 2005). Mit seiner Hilfe lassen sich die Stärken und Schwächen der Sportler in den einzelnen Tests erkennen und abschätzen, auf welche Fähigkeit ein später guter Sportler bauen sollte. Solche Profildiagramme kennt man ebenfalls von Teunissen et al. (2021), die in ihrer Studie die Merkmalsprofile 19 verschiedener Sportarten verglichen haben, um im Rahmen eines Talenttransfers geeignete Alternativsportarten aufzuzeigen. Da es nicht auszuschließen war, dass regional vorkommende Begabungen oder Spezialisierungen das allgemeine Bild eines Sportlers verzerren könnten, wurden die Daten des Profils der vorliegende Validierungsstudie zunächst an der regionalen Bevölkerung normiert. Somit wurden nur die Unterschiede und nicht das regional geprägte Leistungsvermögen der Sportler betrachtet. Dies war insofern notwendig, da Vergleiche mit den Normwerten des Deutschen Motorik-Tests 6–18 gezeigt haben, dass die Schüler der Region vor allem in den Testaufgaben seitliches Hin- und Herspringen wie auch Liegestütz deutlich über dem Bundesdurchschnitt zu liegen scheinen (Bös et al., 2009). Diese Annahme könnte dem Umstand geschuldet sein, dass die Normwerte von Bös et al. anhand von zumeist einer (teilweise auch zwei) linearen Regressionskurven über den kompletten Zeitraum des Alters von 6 bis 18 Jahren berechnet wurden (Bös et al., 2009). Einzelne Schwankungen in einer bestimmten Altersgruppe können durch diese Berechnungsweise nur schwer aufgefangen werden. Aufgrund der deutlichen Diskrepanz beider Mittelwerte von den vorgeschlagenen Normen sollten ggf. die Normwerttabellen für bestimmte Altersgruppen in zukünftigen Studien nochmals überprüft und gegebenenfalls angepasst werden.

Das Merkmalsprofil wurde in späteren Studien auch für die Sportarten Tischtennis und Tennis erstellt und so ein Vergleich dieser drei Sportarten ermöglicht. In einzelnen Studien wurde darüber hinaus auch das Profil von der Gesamtgruppe an Sportlern aller Sportarten betrachtet. Die gefundenen Leistungsprofile decken sich überwiegenden mit den Ergebnissen anderer Studien (Gil et al., 2007; K. Robertson et al., 2018; Teunissen et al., 2021). Auffällig ist daneben, dass die beiden fußballspezifischen Sportartprofile des frühen und späten Juniorenalters eine ähnliche Merkmalsausprägung aufwiesen, woraus geschlossen werden kann, dass dieselben Fähigkeiten der U9 sowohl für frühe als auch für späte Juniorenerfolge verantwortlich sein können. Dies ist insofern bedeutend, da sich die körperlichen Anforderungen bestimmter Sportarten bspw. durch die Pubertät ändern können und zuvor von Agilität geprägte Wettkämpfe sich hin zu verstärkt kraftbetonten Auseinandersetzungen wandeln, wie dies bspw. im Tennis bereits nachgewiesen werden konnte (Kovalchik & Reid, 2017; Ulbricht et al., 2016). Um dieser Vermutung genauer nachgehen zu können, sind weitere Studien zur Frage nach U9-Profilvoraussetzungen für unterschiedliche Altersbereiche und der Stabilität von Leistungen (vgl. Höner et al., 2015) notwendig.

Während im ersten Fußballartikel auch die Prognosegenauigkeit der Klassifikatoren betrachtet wurde, lag der Fokus des zweiten Artikels stärker auf der Güte der verwendeten Prädiktoren. Aus diesem Grund wurden ergänzend zu den allgemeinen Testaufgaben des Deutschen Motorik-Tests 6–18 hier auch Talentförder- und

Talententwicklungsparameter (Güllich & Krüger, 2013) in die Untersuchungen miteinbezogen. Es zeigte sich in einer Pfad-Analyse (vgl. Höner et al., 2015), dass neben der Sprintfähigkeit, der Ausdauer und dem Gleichgewicht auch die Sportförderklasse einen direkten Einfluss auf die U17-Leistung besitzt. Der größte Zusammenhang besteht jedoch zwischen der U17-Leistung und der früheren U13-Leistung, welche wiederum stark vom DFB-Stützpunkt bedingt wird ($r = 0,77$). Die große Bedeutung von schulischer Sportförderung und Talentförderzentren des Deutschen Fußball-Bundes für die Leistungsentwicklung und somit auch für den späteren Erfolg konnte ebenfalls in anderen Studien belegt werden (Granacher & Borde, 2017; Höner et al., 2015).
Der Einsatz von zusätzlichen Parametern der Talentförderung und -entwicklung scheint daher für die Talentidentifikation einen signifikanten Mehrwert zu bieten. Studien von Sieghartsleitner et al. (2019) zeigten bspw., dass durch die Hinzunahme von Prädiktoren der Talentförderung (z.B. *family support*; OR = 5.2) die Treffergenauigkeit von Talentprognosen um durchschnittlich bis zu 13% ansteigen kann. Die Fläche unter der ROC-Kurve erhöhte sich sogar durchschnittlich um bis zu 14% ($p < 0,01$).

Einige Prädiktoren der Talentförderung und -entwicklung lassen sich jedoch erst mit wachsendem Alter erheben (z.B. die Teilnahme an bestimmten Kadern), wodurch sich zwangsweise ebenfalls das Prognoseintervall verringert und die Prognosegüte steigt (Güllich, 2013). Auch zeigten Studien von Leyhr et al. (2020), dass sich die Leistungsentwicklung im Alter U12 bis U15 von später erfolgreichen Sportlern nicht zwangsweise von der Entwicklung später schwächerer Sportler unterscheidet. Welche Talentförder- und Talententwicklungsprädiktoren jedoch bereits in der U9 einen entscheidenden Mehrwert zur Talentprognose besitzen, ist bisher noch nicht umfassend geklärt.

Der erste Tennisartikel prüfte den Zusammenhang von Prädiktor und Kriterium mittels einer Korrelationsanalyse. Diese Vorgehensweise zur Validierung von Testaufgaben ist bereits durch eine Vielzahl an Studien bekannt (Kramer et al., 2017; Meckel et al., 2015; Roetert et al., 1992; Ulbricht et al., 2016). Die Korrelationswerte der Tests mit der bis zu neun Jahre später erreichten Spielleistung liegen in der vorliegenden Tennisstudie mit Ausnahme des Ballweitwurfs ($r = 3,6$; $p < 0,01$) unter einem Wert von $r = 0,3$ und verdeutlichen somit nur geringe Zusammenhänge. Nichtsdestotrotz können auch diese vermeintlich schwächeren Prädiktoren in ihrer Gesamtheit einen signifikanten Mehrwert zur Talentprognose bieten, was sich wiederum im zweiten Tennisartikel bewahrheiten konnte.

Talentprognose (Prediction)

Im ersten Fußballartikel wie auch im Tischtennisartikel kamen zusätzlich zu den bereits genannten Methoden die drei Klassifikatoren lineare Gewichtung, Diskriminanzanalyse (DA) und mit dem Multilayer Perceptron (MLP) auch ein künstliches neuronales Netz zum Einsatz. Anhand der durch diese Methoden berechneten Validitätskennwerte aus Sensitivität, Spezifität und Gesamttreffergenauigkeit wurden erste Erkenntnisse zur prognostischen Gesamtvalidität der Talentprognosen gewonnen. Die Gesamtgenauigkeit der Vorhersagen erreichte dabei unabhängig von der Analysemethode jeweils über

70%. Dieser Grenzwert bezeichnet nach Güllich (2019) den oberen Rand der empirischen beobachteten Genauigkeit üblicher Talentidentifikationsstudien, weshalb die Ergebnisse als vergleichsweise gut einzuschätzen sind. Studien von Robertson et al. (2015) erreichten in Klassifikationen von ca. 5.000 australischen Footballspielern nur eine Gesamttrefferquote von ca. 65%. Auch Dugdale et al. (2019) kamen in ihrer Querschnitts-Studie an 373 schottischen Fußballerspielern in Klassifikationsanalysen zum aktuellen Leistungslevel (amateur, development, performance) nur auf Gesamtgenauigkeiten von knapp 70%. In Talentprognosen von Sieghartsleitner et al. (2019) von insgesamt 218 Schweizer Junioren-Fußballern unterschiedlicher Altersklassen wurden ebenfalls Treffergenauigkeiten von durchschnittlich 70% (66%–78%) erreicht.
Es gibt jedoch auch Ausnahmen. So erreichten bspw. Pion et al. (2017) bei Untersuchungen an 243 Turnerinnen in Klassifikationsanalysen mithilfe eines neuronales Netzes (MLP) Gesamttreffergenauigkeiten von 79,8%. Die Prognosen der befragten Trainer trafen hingegen nur zu 51,9% zu.

Vergleicht man in den vorliegenden Validierungsstudien die Sensitivität der verschiedenen Prognosen der Sportartgewichtung bzw. des sportartspezifischen Empfehlungswertes bei einer vorgegebenen Spezifität von 82,9% (dieser Prozentwert wurde in allen vier Studien angegeben) miteinander, so wird ersichtlich, dass die einzelnen Ergebnisse stark voneinander abweichen. Während die Ergebnisse im Tischtennis nur einen gering höheren Wert als ein rein zufälliges Prognoseergebnis zeigen, können in der verwandten Rückschlagsportart Tennis deutlich besser Befunde erzielt werden. Die Prognosegenauigkeit der vorgeschlagenen Gewichtung ist im Tennis mit einer Sensitivität von über 70% die vielversprechendste Klassifikation unter den drei Sportartprognosen. Die Treffergenauigkeit in der Gruppe der späteren Hochleister fällt für beide Fußballanalysen trotz des für die zweite Studie um zwei Jahre verlängerten Prognosebereichs mit ca. 42,5% nahezu identisch aus (Tab. 10).

Tab. 10. *Sensitivitätswerte der sportartspezifischen Gewichtungen bei einer vorgegebenen Spezifität von 82,9%*

	Soccer #1	Soccer #2	Table Tennis	Tennis
Sensitivity	42.4 %	42.6 %	23.1 %	75.0 %

Zum einen könnten die Unterschiede in der Prognosegüte der verschiedenen Sportarten direkt auf die gewählte Gewichtung zurückzuführen sein. So könnten bspw. im Fußball eigene Gewichtungen für einzelne Spielerpositionen (Stürmer, Verteidiger, Mittelfeld) notwendig sein, um die Prognose zu verbessern (Slimani & Nikolaidis, 2019). Eine alleinige Berechnungsweise (Gewichtung) über alle Fußballertypen hinweg, wie sie in den vorliegenden Studien angewendet wurde, kann die Vielfalt an Spielertypen eventuell nur schwer abdecken. Auch im Vergleich zur Tennisstudie erreichen die Ergebnisse der Fußballstudien nur geringe Sensitivitätswerte. Im Tennissport gibt es zwar auch unterschiedliche Spielertypen, die Unterschiede machen sich jedoch im Hinblick auf andere Sportarten nicht so deutlich bemerkbar. Die schwache Prognosegüte im Tischtennis

könnte wiederum auf die im Vergleich zu anderen Sportarten eventuell schwächeren Merkmalsausprägungen des Sportartprofils zurückzuführen sein.
Robertson et al. (2018) untersuchten bspw. mithilfe von Fragebögen, wie 177 Trainer der Sportarten Tischtennis, Tennis und Badminton die sportmotorischen Anforderungen ihrer Sportart einschätzen. Von insgesamt 21 verschiedenen Testaufgaben wurde nur die Kompetenz „ball handling: hits“ (jedoch nicht signifikant) für den Tischtennis bedeutsamer eingeschätzt als für die anderen beiden Disziplinen. In nahezu allen anderen Bereichen (Beweglichkeit, Kraft, Schnelligkeit, etc.) lagen die Anforderung der Sportarten Tennis oder Badminton laut Trainereinschätzungen zumeist signifikant über denen des Tischtennis (siehe auch Teunissen et al., 2021). Aufgrund dieser scheinbar sehr spezifischen Anforderungen im Tischtennis (Faber, Bustin et al., 2016; Munivrana et al., 2015) könnte deshalb die Prognosegüte so gering ausfallen. Diese Vermutung wird zusätzlich dadurch begünstigt, dass in den bisherigen Testaufgaben des U9-Altersbereichs vermehrt konditionelle Fähigkeiten und nur in geringem Maße technische Fertigkeiten geprüft wurden. Zudem ist ungeklärt, ob sich sportartspezifische Kompetenzen nicht auch erst im Laufe der Trainingsjahre entwickeln müssen (Baker et al., 2018) und daher in früher Jahren noch gar nicht vorliegen können.
Zum anderen könnten die Unterschiede in der Prognosegüte der drei vorgestellten Sportartgewichtungen auch durch die Stichprobe und den Prognosezeitraum bedingt sein (Baker et al., 2018). In den Altersklassen bis zum frühen Juniorenalter finden Wettkämpfe überwiegend auf regionaler oder auf Landesebene statt. Gewichtungen, die anhand der Profile von international erfolgreichen Leistungssportlern entwickelt wurden, könnten sich jedoch von den regionalen und altersbedingten Anforderungsprofilen noch unterscheiden (Filipcic et al., 2010; Ulbricht et al., 2016). Ob und wann sowie in welchem Ausmaß diese Unterschiede zum Tragen kommen, werden aber erst zukünftige Langzeituntersuchungen über den vollen Prognosezeitraum bis hin zur internationalen Spitzensportleistung im Erwachsenenalter zeigen können.

Nützlichkeit (Utility)

Neben den Berechnungen von Merkmalsprofilen, ORs und ANOVAs kam in den beiden Fußballstudien auch die von Höner und Votteler (2016) bereits vorgestellte Grafik zum Verlauf von Sensitivität und Spezifität bzgl. verschiedener Selektionsraten zum Einsatz (Abb. 17). Inhaltlich entspricht die Grafik einer ROC-Kurve. Für die Sportpraxis stellt diese Kurve eine bedeutende Ergänzung der ansonsten bisher üblichen Ergebnisdarstellungen dar. Je nach individueller Zielsetzung der Talentidentifikation kann somit eine unterschiedliche Anzahl an Hochleister-Empfehlungen (Selektionsquote) ausgegeben und ihre Güte hinsichtlich der Empfehlung abgeschätzt werden. Für Förderprogramme (Talentaufbaugruppen, etc.) mit größeren Personenanzahlen sind bspw. höhere Selektionsraten notwendig als für Talentförderungen mit kleinere Gruppengröße.
Zudem lässt sich neben der je nach Zielsetzung notwendigen Aufklärung der Vorhersagestärke verschiedener Selektionsraten auch der Verlauf von Spezifität und Sensitivität erkennen und dadurch abwägen, wie groß die Fördergruppen optimalerweise sein müssten, um möglichst viele spätere Hochleister zu unterstützen oder möglichst viele spätere

Niedrigleister aus den Fördermaßnahmen auszuschließen. Diese Ergebnisdarstellung ist neben der vereinfachten Auswertbarkeit sportpraktischer Fragestellung auch für eine wissenschaftliche Betrachtungsweise von Vorteil. Zum einen wird dadurch ermöglicht, dass Ergebnisse verschiedener Studien trotz unterschiedlicher Selektionswerte (Cut-off-Werten) verglichen werden können, zum anderen lässt sich bei bekanntem Selektionswert das Prognoseergebnis deutlich besser abschätzen.
Vergleicht man bspw. das Klassifikationsergebnis von Mostaert et al. (2020) – das auf den ersten Blick mit einer Sensitivität von 11,5% sehr gering aussehen mag – mit den Ergebnissen von Höner und Votteler (2016), so lässt sich feststellen, dass bei ähnlicher Spezifität (~99,4%) von Höner und Votteler nur eine Sensitivität von 8,4% erreicht wird. Die geringen Sensitivitätswerte der Studie von Mostaert et al. (2020) sind demnach dem niedrigen Selektionswert (Höner & Votteler: 1%; Mostaert, et al.: NN) geschuldet und sprechen nicht zwangsweise für ein schwaches Klassifikationsergebnis.

Um die Validität einer Talentprognose dementsprechend besser abschätzen zu können, bieten sich jene Kennwerte an, die hier unter dem Überbegriff *Nützlichkeit (Utility)* aufgenommen wurden (Kap. 3.7). Darunter befindet sich bspw. der RIOC-Wert (Cairney & Streiner, 2011; Marx & Lenhard, 2010), der Youden Index (Youden, 1950), der F_1-Wert oder auch die Fläche unter der ROC-Kurve.
Der RIOC-Wert betrachtet bspw. das Prognoseergebnis in Relation zu einer rein zufälligen und maximal möglichen Treffergenauigkeit. Dieses Vorgehen ist insofern notwendig, da einerseits die Zufallstrefferquote von dem gewählten Selektionswert abhängt und andererseits nicht für alle Prognosen immer eine maximale Treffergenauigkeit von 100% angenommen werden kann. Ersteres ist durch die zumeist ungleiche Verteilung innerhalb der Leistungsgruppen bedingt. Dies soll hier an einem Beispiel verdeutlich werden: Läge die Grundrate der Hochleister in der Bevölkerung bei 10% und die Selektionsrate bei 33,3%, würde per Zufall in einem Klassifikationsverfahren jede dritte Person als späterer Hochleister und von 100 Personen (10 Hochleister + 90 Niedrigleister) 60 Niedrigleister und 3 Hochleister korrekt vorhergesagt. Die zufällige Gesamttreffergenauigkeit läge mit 63 korrekten Vorhersagen von insgesamt 100 Prognosen bereits bei 63%. Bei einer Selektionsrate von 10% würde sich die Zufallstrefferquote sogar auf 82% erhöhen (Marx & Lenhard, 2010).
Für die maximal mögliche Trefferquote verhält es sich ähnlich. Würden bspw. mehr Hochleister-Empfehlungen ausgesprochen als es tatsächliche Hochleister gibt – dies wäre der Fall, wenn die Selektionsrate höher sein sollte als die Grundrate –, dann fiele jede zu optimistisch gestellte Vorhersage (Überempfehlung) automatisch falsch aus und die Spezifität würde sinken.
Bleibt man beim genannten Beispiel mit insgesamt 33 ausgesprochenen Hochleisterempfehlungen (Selektionsrate bei 33,3%), obwohl laut Grundrate nur zehn Hochleister existieren (Grundrate bei 10%), würde dies bedeuten, dass mindestens 23 Hochleisterempfehlungen falsch wären. Die maximal mögliche Treffergenauigkeit läge in diesem Beispiel folglich bei 77%.

Auch der Youden Index bzw. Joudens *J* lässt einen guten Vergleich mit einem rein zufälligen Ergebnis zu, da bei einem Wert von $J = 0$ ein zufälliges Trefferverhältnis angenommen werden kann. Jedoch lässt sich die Güte der Prognose nicht allein an der Höhe des Youden Index festmachen, da ein Wert von $J = 1$ nur dann erreicht wird, wenn Selektionsrate und Grundrate gleich groß sind. Auch die Fläche unter der ROC-Kurve lässt kein abschließendes Gesamtresümee zur Güte der Talentprognosen zu, da in vielen Talentstudien nur niedrige Selektionsraten nahe der Grundrate von Interesse sind, der AUC hingegen über alle Selektionsraten berechnet wird. Jeder Parameter weist somit unterschiedliche Schwächen im Hinblick auf die mögliche Interpretation auf.

Aus diesem Grund wurden alle vier Parameter in der zweiten Tennisstudie exemplarisch für vier Klassifikatoren bestimmt und verglichen. Es zeigte sich, dass das künstliche neuronale Netz und die tennisspezifische Gewichtung bzgl. der gewählten Kennwerte ähnliche Resultate erzielen. Beide Klassifikationsmethoden erreichten RIOC-Werte von über 0,66, was nach Marx und Lenhard (2010) einem guten Ergebnis entspricht.
In Studien von Pion et al. (2017) wurden anhand vergleichbarer Methoden RIOC-Werte zwischen umgerechnet 0,278 (MLP) und 0,537 (Gewichtung) erreicht. Auch Marx und Lenhard (2010) erlangten in ihren Studien nur einen RIOC-Wert von 0,55. Weitaus vielversprechender waren hingegen die Befunde der Studien von Sieghartsleitner et al. (2019). Sie kamen mithilfe eines holistischen Talentmodells in einer logistischen Regression auf einen RIOC-Wert von 0,866.
Relativ ähnlich fielen auch die Werte für F_1-Wert und Youden Index der beiden Methoden MLP und Gewichtung aus. Youdens *J* erreicht bei beiden Klassifikatoren für den gewählten Selektionswert ca. 58%. Dieser Kennwert ist für die Sportart Tennis im Vergleich mit den anderen Artikeln der vorliegenden Validierungsstudien der höchste. In den beiden anderen Sportarten Fußball und Tischtennis kam der Youden Index nicht über 34% hinaus.
In den genannten Untersuchungen von Sieghartsleitner et al. (2019) wurde neben dem RIOC-Wert auch der Youden Index betrachtet. Analysiert wurde dabei, wie sich die Auswahl unterschiedlicher Prädiktoren auf die Talentprognose einer logistischen Regression auswirkt. Hierbei wurden je nach Modell (Prädiktorauswahl) für den Youden Index Werte zwischen 43% und 77% erreicht. Den geringsten Wert erzielte dabei ein Modell bestehend nur aus sportmotorischen Tests. Ein Wert von 58% kann in der hier vorliegenden Studie daher als vielversprechend angesehen werden.
Neben dem Youden Index betrachteten Sieghartsleitner et al. (2019) auch die AUC-Kennwerte der unterschiedlichen Modelle und verglichen diese mittels des DeLong-Tests. Die Fläche unter der Kurve kam dabei für das *motor performance model* auf einen Wert von 71%.
Faber et al (2016) erzielten in ihren Studien an über 400 Junioren-Tischtennisspieler in Untersuchungen von sportspezifischen Prädiktoren AUC-Werte von 0,589 bis 0,652. Diese Werte liegen ebenfalls unter den hier ermittelten Werten von 80,0 bis 85,2%, die laut D'Agostino et al. (2013) als „exzellent“ einzuschätzen sind. Ein DeLong-Test wurde jedoch in den vorliegenden Validierungsstudien nicht eingesetzt. Seine Durchführung

wäre für zukünftige Studien trotzdem interessant, obwohl der AUC aufgrund der zuvor beschrieben Schwächen nicht pauschal für alle Zielsetzungen geeignet scheint.

Limitationen

Gerade die Befunde von Zuber et al. (2015) und Sieghartsleitner et al. (2019) verdeutlichen die hohe Bedeutung der Prädiktorauswahl (siehe auch Davids & Baker, 2007). In den vorliegenden Studien wurden nur Prädiktoren der allgemeinen sportmotorischen Fähigkeiten berücksichtigt. Die Testauswahl entspricht dadurch den in Europa gängigen Tests des ausgewählten Altersbereichs (Niessner et al., 2020; Pion, 2015; Potočnik et al., 2020). Mit der Hinzunahme weiterer Prädiktoren aus den Bereichen Persönlichkeit, Motivation, Umwelt oder Förderung könnten sich die Talentprognosen noch einmal deutlich verbessern (Mosoi, 2013; Zuber et al., 2016).
Zudem kam in den letzten Jahren verstärkt das Thema genetischer Tests auf (Pickering et al., 2019). Inwieweit diese jedoch für eine frühe TID geeignet scheinen, ist noch unklar. Auch Untersuchungen von Faber et al. (2014) zu sportspezifischen Tests im Tischtennis zeigen, dass bestimmte spezifische Testaufgaben bereits in frühem Alter für eine verbesserte Talentprognose herangezogen werden können.

Jedoch ist nicht jeder vielversprechende Test auch für die Aufnahme in die Testbatterie geeignet, da Tests entweder ethisch/moralisch nicht vollumfänglich vertretbar sind (darunter bspw. bestimmte medizinische Eingriffe), sehr teuer erscheinen (z.B. die Magnetresonanztomographie) oder aus anderen Gründen nicht aufgenommen werden können (bspw., weil sie sehr zeitaufwändig sind).
Hinzu kommt, dass einige Fähigkeiten und Fertigkeiten sich erst im Juniorenalter ausprägen und dann erst ihre Trennschärfe bzgl. einer Klassifikation aufweisen könnten. Dieser Umstand kann vor allem auf technisch anspruchsvollere Übungen zutreffen, da diese mitunter stark vom Trainingsalter und der jeweiligen Vorerfahrung bedingt werden (Opstoel et al., 2015). In Studien von Faber et al. (2020) zeigte sich bspw., dass ein signifikanter Zusammenhang zwischen Trainingsstunden und der sportspezifischen Leistungsfähigkeit (speed while dribbling, aiming a target, eye-hand coordination) besteht. Für allgemeine sportmotorische Fähigkeiten, z.B. Sprintschnelligkeit oder Agilität, konnten diese Zusammenhänge nicht nachgewiesen werden.
Nichtsdestotrotz sollte auch im U9-Altersbereich schon über eine holistische Talentidentifikation nachdacht werden, da die Befunde von Sieghartsleitner et al (2019) zeigen, dass ein holistisches Prädiktor-Modell mit einem AUC = 0,93 einen um 22% höheren AUC-Wert als ein Modell auf Basis rein sportmotorischer Tests besitzt. Ob sich diese für die Juniorenbereiche gezeigten Befunde (Koopmann et al., 2020; Mosoi, 2013; Sieghartsleitner, Zuber, Zibung & Conzelmann, 2019) daneben auch für die U9 bewahrheiten können, muss erst noch geprüft werden.

Die in den vorliegenden Studien durchgeführten Prognosen erstrecken sich bisher nur über den Zeitraum bis zum frühen und späten Juniorenalter. Die prognostische Validität von Prognosen über den gesamten langfristen Trainingsaufbau bis hin zur Finalleistung im Erwachsenalter wurde in unseren Studien noch nicht betrachtet. Alle gefunden

Ergebnisse können daher nur als erstes Zwischenergebnis der langfristigen Talentidentifikation gesehen werden. Einige Studien – wie auch die Ergebnisse der vorliegenden Studien – deuten aber auf einen signifikanten Zusammenhang zwischen früher und später Juniorenleistung/-leistungsfähigkeit und der Finalleistung/-leistungsfähigkeit hin.
So zeigten Kramer et al. (2016) in ihren Studien an Junioren-Tennisspielerinnen, dass die Prädiktoren *lower-body power*, *speed*, und *agility* von der U14 bis zur U16 ihre Vorhersagekraft beibehalten konnten.
Ähnliche Befunde wurden auch von Höner et al. (2015) nachgewiesen. In ihren Fußball-Studien korrelierten die Testleistungen des Sprints, der Agilität und des Dribbling-Tests der U12 mit den Testleistungen der U15 zu $r \approx 0{,}5$ (siehe auch Güllich & Emrich, 2014; Hohmann, 2009). In den meisten Studien, so Güllich und Krüger (2013), ließen sich oft signifikante Zusammenhänge zwischen den Leistungen zu unterschiedlichen Zeitpunkten finden – Leistungsunterschiede in Zeiträumen von bis zu zwei Jahren würden typischerweise zu 20 bis 80 % durch frühere Leistungen erklärt. Jedoch sanken bei längeren Zeiträumen ab drei Jahre die Zusammenhänge von früherer und späterer Leistung auf nur 20 bis 50 %.

Eine der wenigen Studien, die den Verlauf von sportmotorischen Leistungsfähigkeit über den kompletten langfristigen Leistungsaufbau von der U7 bis zur U16/U19 betrachtet, unterteilte verschiedene sportmotorische Leistungserfolge in Tertiale und überprüfte, ob die getesteten Probanden ihr anfangs erreichtes Leistungs-Tertial beibehalten haben (True et al., 2021). Unabhängig von der sportmotorischen Disposition konnten 50–60% aller Sportler, die in der U7 das höchste Tertials erreichten, auch nach 9 Jahren in der U16 wieder dieses höchste Tertial behaupten. Die Korrelationsanalyse der Leistungen der männlichen U7 und U19 zeigte dabei für die Testaufgaben (*flexed arm hang, flexed arm hang/weight, jump and reach, standing long jump, standing long jump/height, agility shuttle run, 30-yard dash, endurance shuttle run* und *Wells sit and reach*) Werte im Bereich $0{,}210 \leq r \leq 0{,}567$. Die höchsten Korrelationswerte erreichten die Testitems *30 yard dash* ($r = 0{,}567$), *sit and reach* ($r = 0{,}501$) sowie *jump and reach* ($r = 0{,}447$) und fielen somit deutlich höher aus als die Befunde von Brouwers et al (2012; siehe unten). Es lässt sich folglich vermuten, dass gerade diese beiden Testvariablen durch ihre Stabilität als Prädiktor für die Talentidentifikation geeignet sind.
Ob jedoch die in den vorliegenden Artikeln prognostizierten Sportler des Juniorenbereichs auch im Erwachsenenbereich erfolgreich sind, hängt wiederum von der Güte des Kriteriums ab.

Für die Sportart Tennis hingegen scheinen gewisse Zusammenhänge zwischen Junioren- und späterer Erwachsenenleistung zu existieren. Es gibt zwar Befunde von Brouwers et al. (2012), dass lediglich 16,8% der männlichen Junioren-Top20-Spieler auch einen Platz unter den besten 20 Sportlern der Erwachsenenrangliste erreichen und nur 43,2% der U14-Turniersieger sich später unter den besten 200 Spielern der Herrenrangliste finden lassen – die bestimmten Korrelationswerte fielen zudem mit $r_{male} = 0{,}208$ und $r_{female} = 0{,}296$ relativ gering aus –, betrachtet man aber die Gegebenheiten aus Sicht

des Erwachsenenalters und fragt nach der Notwendigkeit eines früheren Juniorenerfolges, fallen die Zusammenhänge im Tennis zumeist höher aus.
Befunde von Reid und Morris (2013) belegen bspw., dass 91% der männlichen Tennisspieler mit einem Ranglistenplatz unter den erst 100 Spielern der Weltrangliste bereits im Juniorenalter einen Platz auf der Juniorenrangliste erzielt hatten. Studien von Li et al. (2018) zeigten, dass 75% der Herren-Top10-Weltrangliste ihren ersten Titel bereits vor dem 16. Lebensjahr erzielten. So deutet alles darauf hin, dass viele der besten Tennisspieler der *Aktivenrangliste* bereits früh Erfolge erzielten, jedoch ein früher Erfolg nicht zwangsweise spätere Erfolge bedingt.

In anderen Sportarten sind diese Zusammenhäng weit weniger deutlich. So zeigen Umfragen von Güllich und Emrich (2014), dass nur 4% der Weltklasse-Sportler aus 47 Olympischen Sportarten bereits in der U15 eine internationale Top10-Plazierung und nur 31% eine nationale Top10-Platzierung erreichten. Selbst in der U19 fuhren immer noch 20% der späteren Weltklasse bisher nur auf regionaler Ebene oder niedriger Erfolge ein. Ob sich demnach die in der vorliegenden Studienreihe erstellten Prognosen der Juniorenerfolge auch auf die Finalleistung übertragen lassen oder die gleichen Juniorensportler später Erfolge im Erwachsenenalter erzielen, muss in weiteren Studien noch überprüft werden.

Die prognostische Validität von Talentprognosen sollte grundsätzlich ausschließlich über Langzeitstudien erfolgen, da nur dort der tatsächliche Zusammenhang zwischen frühen Prädiktoren und spätem Kriterium bestimmt werden kann (Baker et al., 2020; Höner et al., 2017; Johnston et al., 2018).
Untersuchungen über einen solch langen Zeitraum bergen jedoch die Gefahr einer *Validitätsverschiebung* der verwendeten Testverfahren (Hohmann et al., 2015). Unter diesem Begriff versteht man eine Änderung der allgemeinen Voraussetzungen, bspw. durch den Wechsel der sportartspezifischen Anforderungen oder durch Regeländerungen, und einer dadurch veränderten Einschätzung der Prädiktoren. Hierunter fallen bspw. offensichtliche Änderungen in einer Sportart, wie die Abschaffung der Abseitsregel (1996) im Feldhockey, die Einführung des Fosbury Flop (1986) im Hochsprung und des V-Stils (Anfang der 1990er Jahre) im Skisprung oder die Einführung und das spätere Verbot der Radschlag-Technik (Cartwheel) im Kugelstoßen sowie des sogenannten „spanischen“ Wurfstils im Speerwurf.
Es können neben neuen Techniken und Wettkampfregeln jedoch auch weniger offensichtliche Anpassungen des Sportartprofils vorkommen. Ein Beispiel hierfür sind die Olympiasieger *Usain Bolt* und *Swetlana Wassiljewna Chorkina*, die anfangs für ihre Sportart als von der Körpergröße her zu groß eingeschätzt wurden (Abbott et al., 2005; Baker et al., 2018). Noch Anfang des 21. Jahrhunderts wurde die Schrittfrequenz als wichtigster Indikator für eine gute Sprintgeschwindigkeit eingeschätzt und eine hohe Körpergröße als Nachteil für diesen Prädiktor gesehen (Baker et al., 2018). Nach den Erfolgen von Bolt zeigte sich jedoch, dass eine Kombination aus Schrittfrequenz und Schrittweite, welche wiederum von der Körpergröße bedingt wird, die 100m-Sprint-Zeiten erklärt (Salo et al., 2011) und der Fokus auf beiden Prädiktoren liegen muss.

Nach solchen Weiterentwicklungen und Änderungen können andere Sportler als zuvor im Wettkampf bevorteilt sein oder in den Fokus der Talentidentifikation rutschen und sich das entsprechende, gesuchte Leistungsprofil „verschieben“. Dies hätte zur Folge, dass zuvor valide Prädiktoren und die damit verbundene prognostische Validität der Talentprognose neu bewertet werden müssten.

Diesen Überlegungen schließen sich mögliche Änderungen des sportartspezifischen Anforderungsprofils auch zwischen verschiedenen Altersbereichen (U17, U19, U21, etc.) (*Validitätsfluktiation*; Güllich & Krüger, 2013) und der Umstand an, dass der erreichte Erfolg von sich ständig ändernden Trainings- und Förderbedingungen abhängt, die nur schwer durch einzelne Prädiktoren abgedeckt werden können (Schnabel et al., 2016).
Ein weiteres Problem in Langzeitstudien kann die vorkommende *Stichprobenmortalität* (*Drop-Outs*) sein (Hohmann et al., 2015).

So sehen sich gerade die für eine Abschätzung der prognostischen Validität wichtigen Langzeitstudien einer Reihe von Schwierigkeiten gegenüber, die die Güte der Talentprognose beeinflussen können. Aus diesem Grund ist es notwendig, die Validität der Talentidentifikation immer wieder neu zu prüfen und Talentstudien den Gegebenheiten der Sportpraxis anzupassen.

6. Zusammenfassung

Insgesamt sind die Kennwerte zur prognostischen Validität in den drei Sportarten sehr unterschiedlich ausgeprägt. Die schwächsten Ergebnisse zeigten sich im Tischtennis. Zwar lässt sich ein Mehrwert gegenüber einer rein zufälligen Prognose erkennen, jedoch fallen gerade in dem für die Sportpraxis wichtigen Grenzbereich der Selektionswertskala nahe der Grundrate die Unterschiede vergleichsweise gering aus. Dies könnte an den zuvor genannten Schwierigkeiten der Testauswahl oder an der Auswahl der Probanden liegen.
Die Ergebnisse im Tennis sind hingegen äußerst vielversprechend, da die meisten Kennwerte im Vergleich zu bisherigen Studien mit ähnlichen Prädiktoren sehr hoch ausfallen. Die prognostische Validität kann daher gerade für diese Sportart als hoch eingeschätzt werden.
Betrachtet man die Validitätskennzahlen, kommen die Befunde im Fußball nicht an die hohe prognostische Validität der Sportart Tennis heran. Dennoch lässt sich trotz der unterschiedlichen Merkmalsprofile der Fußballer mithilfe einer Einzel-Gewichtung eine signifikante Verbesserung hinsichtlich einer zufälligen Vorhersage erzielen.
In der Gesamtheit zerstreuen die Ergebnisse der einzelnen Validitätsstudien somit die teilweise vorherrschenden Zweifel (Baker et al., 2018; Güllich, 2019; Güllich & Emrich, 2014) an einer frühen Talentidentifikation. Gerade in bestimmten Sportarten ist eine mittel- und langfristige Prognose zukünftiger Leistungen schon in jungen Jahren möglich und die bereits in dieser frühen Phase des Diskurses der zukünftigen Talentidentifikation gefunden Ergebnisse lassen für zukünftige Studien noch viel erwarten.

7. Empfehlung für die Praxis

Im Rahmen vorliegenden Studien zu Validierung von Talentprognosen hat sich ein Validierungskonzept ergeben, das auch für zukünftige Untersuchungen der Talentidentifikation bedeutsam scheint. Die Empfehlungen für die Praxis richten sich daher vorrangig an Wissenschaftler und Forscher, die die prognostische Validität einer Erfolgsprognose im Rahmen einer (Leistungs-)Diagnostik bewerten wollen. Das empfohlene Validierungskonzept unterteilt sich in sieben Teilschritte (Abb. 20), die aufeinander aufbauen.

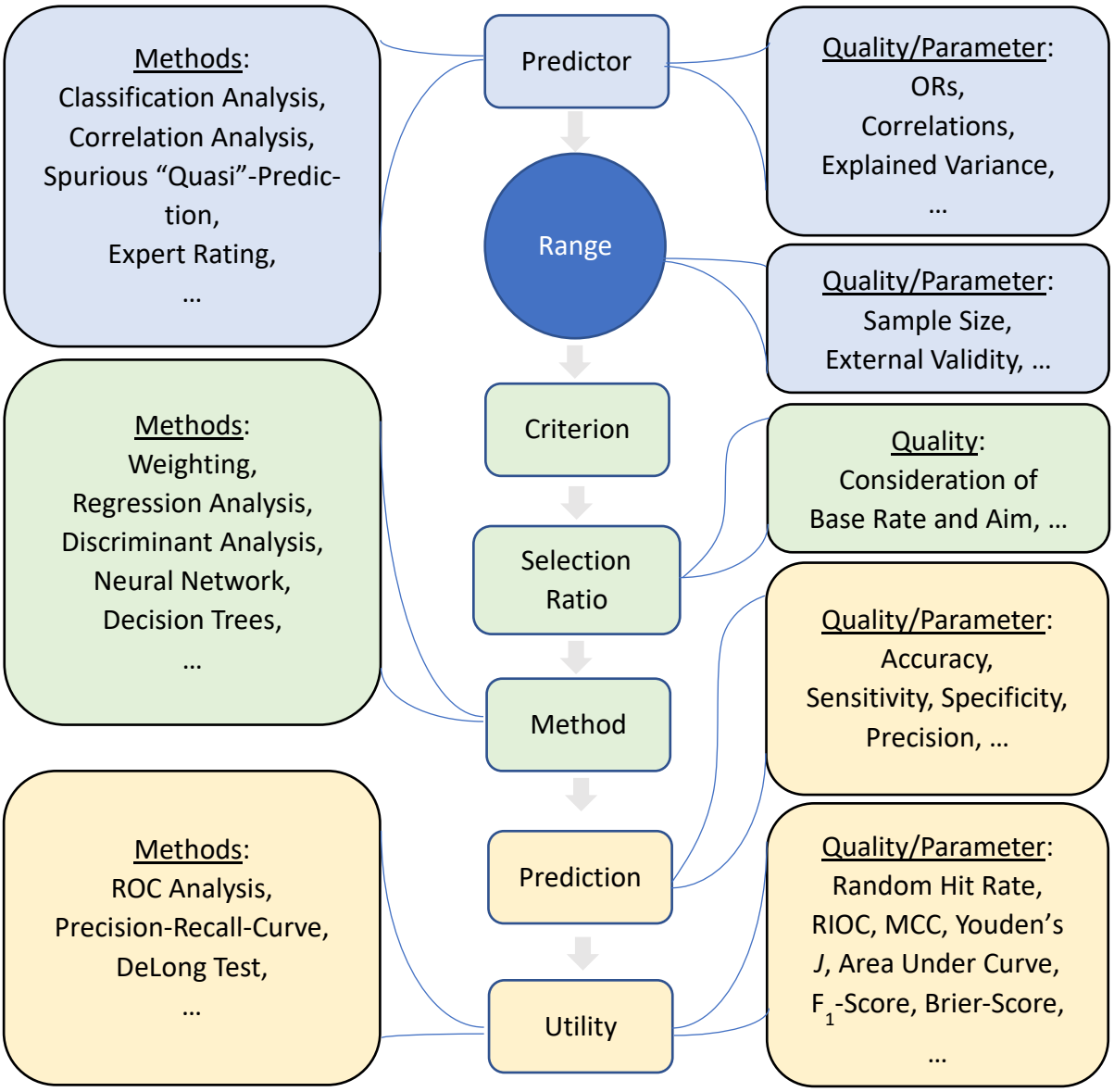

Abb. 20. Validierungskonzept mit möglichen Schritten und Kennwerten zur Bestimmung der prognostischen Validität von (Talent-)Prognosen

Geeignete **Prädiktoren** (Schritt 1) lassen sich bspw. aus den Anforderungen der Weltklasse ableiten oder über Trainerbefragungen erheben. Inwiefern diese ausgewählten Prädiktoren für eine Talentprognose geeignet sind, lässt sich unter anderem über eine

Korrelationsanalyse, die Zusammenhänge von Prädiktor und Kriterium berechnet, oder mithilfe von ORs einer binären logistischen Regression prüfen.
Um abschätzen zu können, inwieweit das später gefundene Ergebnis der prognostischen Validität verallgemeinert werden kann, sollten in den Beschreibungen der **Stichprobe** (Schritt 2) neben dem Stichprobenumfang auch Angaben zur externen Validität gemacht und mögliche Einflüsse (Vorauswahl durch Trainer, Teilnehmer haben die gleichen Fördermaßnahmen durchlaufen, etc.) der Stichprobe thematisiert werden.
Neben den Prädiktoren hat auch das gewählte **Kriterium** (Schritt 3) einen hohen Einfluss auf die schlussendliche Prognoseleistung. Das Kriterium sollte daher nahe an den Gegebenheiten der tatsächlichen Praxis und entsprechend der Zielsetzung gewählt werden. Dabei gilt es, feste Grenzen wenn möglich zu vermeiden. Ob die Einteilung der durch das Kriterium gegebenen (Leistungs-)Gruppen erfolgversprechend erscheint, lässt sich mithilfe von *t*-Tests und ANOVAs untersuchen.
Es sollten grundsätzlich keine festen **Selektionsraten** (Schritt 4) gewählt werden, da dadurch ein späterer Vergleich mit anderen Studien erschwert wird und sich gleichzeitig die Validität der Talentprognose nicht mehr vollumfänglich einschätzen lässt. Feste Cut-off-Werte zu verwenden, kann dann sinnvoll sein, wenn diese den in der Praxis tatsächlich möglichen Fördermöglichkeiten und Förderzielsetzungen entsprechen.
Unter den **Prognose-/Klassifikationsmethoden** (Schritt 5) hat sich in den vorliegenden Studien zumeist eine Kombination aus einem linearen und einem nicht-linearen Verfahren als die beste Variante herausgestellt. Es sollten daher immer mindestens zwei Methoden exemplarisch durchgeführt werden. Unabhängig von der gewählten Methode bietet sich als Verfahren eine mehrfache Kreuzvalidierung (z.B. 10-fold CV) an, da diese dem Vorgehen in der Praxis mit am ehesten entspricht.
Die **Prognosegüte** (Schritt 6) der schlussendlichen Talentprognose lässt sich mithilfe von Sensitivität, Spezifität, Gesamttreffergenauigkeit, positivem prädiktivem und negativem prädiktiven Wert angeben. Für eine einfache Betrachtung der Ergebnisse bieten sich zudem Streudiagramme (Abb. 10) und Vier-Felder-Tafeln (Abb. 14 und Tab. 5) an.
Die große Bandbreite an statistischen Kennwerten der **Nützlichkeit** (Schritt 7) macht es schwer, einen einheitlichen Parameter als zukünftigen Validitätskennwert der (Talent-)Forschung auszuweisen. Je nach Zielsetzungen können einzelne Kennwerte Vor- und Nachteile haben, die es abzuwägen gilt. In naher Zukunft wird man daher wohl nicht darum herumkommen, unterschiedliche Variablen zur Bestimmung der prognostischen Validität zu berechnen. In den vorliegenden Studien wurden deshalb exemplarisch nachfolgende vier Kennwerte genutzt: Prognosegüte einer zufälligen Talentprognose, RIOC-Wert, AUC-Wert und der Youden Index. Für eine anschauliche Einschätzung der Ergebnisse können auch Sensitivität-Spezifität-Diagramme (Abb. 17), ROC-Kurven oder Precision-Recall-Kurven (Abb. 18) eine sinnvolle Ergänzung der Ergebnispräsentation darstellen. Diese Kurven sollten zusätzlich zur Verdeutlichung des Verlaufs der Trefferquoten von Selektionsraten nahe der Grundrate genutzt werden.

8. Literaturverzeichnis

Abbott, A., Button, C., Pepping, G.-J., & Collins, D. (2005). Unnatural selection: talent identification and development in sport. *Nonlinear dynamics, psychology, and life sciences*, *9*(1), 61–88.

Ali, A., Williams, C., Hulse, M., Strudwick, A., Reddin, J., Howarth, L., Eldred, J., Hirst, M., & McGregor, S. (2007). Reliability and validity of two tests of soccer skill. *Journal of Sports Sciences*, *25*(13), 1461–1470. https://doi.org/10.1080/02640410601150470

Allen, S. V., Vandenbogaerde, T. J., Pyne, D. B., & Hopkins, W. G. (2015). Predicting a nation's olympic-qualifying swimmers. *International journal of sports physiology and performance*, *10*(4), 431–435. https://doi.org/10.1123/ijspp.2014-0314

Artero, E. G., España-Romero, V., Castro-Piñero, J., Ortega, F. B., Suni, J., Castillo-Garzon, M. J., & Ruiz, J. R. (2011). Reliability of field-based fitness tests in youth. *International Journal of Sports Medicine*, *32*(3), 159–169. https://doi.org/10.1055/s-0030-1268488

Azevedo, C. (12. April 2020). On ROC and Precision-Recall curves - Towards Data Science. *Towards Data Science*. Abruf unter https://towardsdatascience.com/on-roc-and-precision-recall-curves-c23e9b63820c

Backhaus, K., Erichson, B., & Weiber, R. (2015). *Fortgeschrittene Multivariate Analysemethoden*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-46087-0

Baker, J., Schorer, J., & Wattie, N. (2018). Compromising Talent: Issues in Identifying and Selecting Talent in Sport. *Quest*, *70*(1), 48–63. https://doi.org/10.1080/00336297.2017.1333438

Baker, J., Wilson, S., Johnston, K., Dehghansai, N., Koenigsberg, A., Vegt, S. de, & Wattie, N. (2020). Talent Research in Sport 1990-2018: A Scoping Review. *Frontiers in psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.607710

Bardid, F., Utesch, T., & Lenoir, M. (2019). Investigating the construct of motor competence in middle childhood using the BOT-2 Short Form: An item response theory perspective. *Scandinavian journal of medicine & science in sports*, *29*(12), 1980–1987. https://doi.org/10.1111/sms.13527

Barron, D., Ball, G., Robins, M., & Sunderland, C. (2018). Artificial neural networks and player recruitment in professional soccer. *PloS one*, *13*(10), e0205818. https://doi.org/10.1371/journal.pone.0205818

Beleites, C., Salzer, R., & Sergo, V. (2013). Validation of soft classification models using partial class memberships: An extended concept of sensitivity & co. applied to grading of astrocytoma tissues. *Chemometrics and Intelligent Laboratory Systems*, *122*, 12–22. https://doi.org/10.1016/j.chemolab.2012.12.003

Benedetti, R. (2010). Scoring Rules for Forecast Verification. *Monthly Weather Review*, *138*(1), 203–211. https://doi.org/10.1175/2009MWR2945.1

Bergkamp, T. L. G., Niessen, A. S. M., Hartigh, R. J. R. den, Frencken, W. G. P., & Meijer, R. R. (2019). Methodological Issues in Soccer Talent Identification Research. *Sports medicine (Auckland, N.Z.)*, *49*(9), 1317–1335. https://doi.org/10.1007/s40279-019-01113-w

Blanksby, B. (1980). Measures of talent in competitive swimming. *Sports Coach*, *4*(4), 13–19.

Bös, K., & Schlenker, L. (2011). Deutscher Motorik-Test 6–18 (DMT 6–18). In M. Krüger & N. Neuber (Hrsg.), *Bildung und Sport: Bd. 1. Bildung im Sport: Beiträge zu einer zeitgemäßen Bildungsdebatte* (1. Aufl., S. 337-355). VS Verl. für Sozialwiss. https://doi.org/10.1007/978-3-531-94026-7_21

Bös, K., & Schlenker, L. (2016). *Der Deutsche Motorik-Test 6-18 (2. Edition)*. Feldhaus.

Bös, K., Schlenker, L., Albrecht, C., Büsüch, D., Lämmle, L., Müller, H., Oberger, J., Seidel, I., Tittlbach, S., & Woll, A. (2009). *Deutscher Motorik-Test 6-18 (DMT 6-18)*. Feldhaus.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Brouwers, J., Bosscher, V. de, & Sotiriadou, P. (2012). An examination of the importance of performances in youth and junior competition as an indicator of later success in tennis. *Sport management review*, *15*(4), 461–475.

Büch, M.-P. (2005). Entwicklung von Potentialen für Sporteliten. In E. Emrich, A. Güllich & M.-P. Büch (Hrsg.), *Beiträge zum Nachwuchsleistungssport* (S. 347-368). Hofmann.

Bullock, N., Gulbin, J. P., Martin, D. T., Ross, A., Holland, T., & Marino, F. (2009). Talent identification and deliberate programming in skeleton: ice novice to Winter Olympian in 14 months. *Journal of Sports Sciences*, *27*(4), 397–404. https://doi.org/10.1080/02640410802549751

Cairney, J., & Streiner, D. L. (2011). Using relative improvement over chance (RIOC) to examine agreement between tests: three case examples using studies of developmental coordination disorder (DCD) in children. *Research in developmental disabilities*, *32*(1), 87–92. https://doi.org/10.1016/j.ridd.2010.09.001

Carl, K. (1989). *Talentsuche, Talentauswahl und Talentförderung*. *Studienbrief der Trainerakademie Köln des Deutschen Sportbundes: Bd. 24*. Hofmann.

Carling, C., & Collins, D. (2014). Comment on “Football-specific fitness testing: Adding value or confirming the evidence?”. *Journal of Sports Science*, *32*(13), 1206–1208.

Carling, C., Le Gall, F., & Malina, R. M. (2012). Body size, skeletal maturity, and functional characteristics of elite academy soccer players on entry between 1992 and 2003. *Journal of Sports Sciences*, *30*(15), 1683–1693. https://doi.org/10.1080/02640414.2011.637950

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, *21*(1), 6. https://doi.org/10.1186/s12864-019-6413-7

Cobley, S [Stephen], Baker, J., & Schorer, J. (2021). Talent Identification and Development in Sport. In J. Baker, J. Schorer & S. Cobley (Hrsg.), *Talent identification and development in sport: International perspectives* (S. 1-16). Routledge. https://doi.org/10.4324/9781003049111-1

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Copas, J. B., & Loeber, R. (1990). Relative improvement over chance (RIOC) for 2×2 tables. *British Journal of Mathematical and Statistical Psychology*, *43*(2), 293–307. https://doi.org/10.1111/j.2044-8317.1990.tb00942.x

Côté, J., Lidor, R., & Hackfort, D. (2009). ISSP position stand: To sample or to specialize? Seven postulates about youth sport activities that lead to continued participation and elite performance. *International Journal of Sport and Exercise Psychology*, *7*(1), 7–17. https://doi.org/10.1080/1612197X.2009.9671889

D'Agostino, R. B., Pencina, M. J., Massaro, J. M., & Coady, S. (2013). Cardiovascular Disease Risk Assessment: Insights from Framingham. *Global heart*, *8*(1), 11–23. https://doi.org/10.1016/j.gheart.2013.01.001

Davids, K., & Baker, J. (2007). Genes, environment and sport performance: why the nature-nurture dualism is no longer relevant. *Sports medicine (Auckland, N.Z.)*, *37*, 961–980.

Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. In W. Cohen (Hrsg.), *ACM Other conferences, Proceedings of the 23rd international conference on Machine learning* (S. 233-240). ACM. https://doi.org/10.1145/1143844.1143874

De Bosscher, V., De Knop, P., van Bottenburg, M., & Shibli, S. (2006). A Conceptual Framework for Analysing Sports Policy Factors Leading to International Sporting Success. *European Sport Management Quarterly*, *6*(2), 185–215. https://doi.org/10.1080/16184740600955087

DeLong, E. R., DeLong, D. M., & 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.

Deprez, D., Coutts, A. J., Lenoir, M., Fransen, J., Pion, J., Philippaerts, R. M., & Vaeyens, R. (2014). Reliability and validity of the Yo-Yo intermittent recovery test level 1 in young soccer players. *Journal of Sports Sciences*, *32*(10), 903–910. https://doi.org/10.1080/02640414.2013.876088

Dietterich (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural computation*, *10*(7), 1895–1923. https://doi.org/10.1162/089976698300017197

Dobbin, N., Highton, J., Moss, S. L., & Twist, C. (2019). The Discriminant Validity of a Standardized Testing Battery and Its Ability to Differentiate Anthropometric and Physical Characteristics Between Youth, Academy, and Senior Professional Rugby League Players. *International journal of sports physiology and performance*, *14*(8), 1110–1116. https://doi.org/10.1123/ijspp.2018-0519

Drid, P., Casals, C., Mekic, A., Radjo, I., Stojanovic, M., & Ostojic, S. M. (2015). Fitness and Anthropometric Profiles of International vs. National Judo Medalists in Half-Heavyweight Category. *Journal of strength and conditioning research*, *29*(8), 2115–2121. https://doi.org/10.1519/JSC.0000000000000861

Dugdale, J. H., Arthur, C. A., Sanders, D., & Hunter, A. M. (2019). Reliability and validity of field-based fitness tests in youth soccer players. *European journal of sport science*, *19*(6), 745–756. https://doi.org/10.1080/17461391.2018.1556739

Emmonds, S., Till, K., Jones, B., Mellis, M., & Pears, M. (2016). Anthropometric, speed and endurance characteristics of English academy soccer players: Do they influence obtaining a professional contract at 18 years of age? *International Journal of Sports Science & Coaching*, *11*(2), 212–218. https://doi.org/10.1177/1747954116637154

Emrich, E., Güllich, A., & Büch, M.-P. (Hrsg.). (2005). *Beiträge zum Nachwuchsleistungssport*. Hofmann.

Faber, I. R., Bustin, P. M. J., Oosterveld, F. G. J., Elferink-Gemser, M. T., & Nijhuis-Van der Sanden, M. W. G. (2016). Assessing personal talent determinants in young racquet sport players: a systematic review. *Journal of Sports Sciences*, *34*(5), 395–410. https://doi.org/10.1080/02640414.2015.1061201

Faber, I. R., Elferink-Gemser, M. T., Faber, N. R., Oosterveld, F. G. J., & Nijhuis-Van der Sanden, M. W. G. (2016). Can Perceptuo-Motor Skills Assessment Outcomes in Young Table Tennis Players (7-11 years) Predict Future Competition Participation and Performance? An Observational Prospective Study. *PloS one*, *11*(2), e0149037. https://doi.org/10.1371/journal.pone.0149037

Faber, I. R., Elferink-Gemser, M. T., Oosterveld, F. G. J., Twisk, J. W. R., & Nijhuis-Van der Sanden, M. W. G. (2017). Can an early perceptuo-motor skills assessment predict future performance in youth table tennis players? An observational study (1998-2013). *Journal of Sports Sciences*, *35*(6), 593–601. https://doi.org/10.1080/02640414.2016.1180421

Faber, I. R., Nijhuis-Van der Sanden, M. W. G., Elferink-Gemser, M. T., & Oosterveld, F. G. J. (2015). The Dutch motor skills assessment as tool for talent development in table tennis: a reproducibility and validity study. *Journal of Sports Sciences*, *33*(11), 1149–1158. https://doi.org/10.1080/02640414.2014.986503

Faber, I. R., Oosterveld, F. G. J., & Nijhuis-Van der Sanden, M. W. G. (2014). Does an eye-hand coordination test have added value as part of talent identification in table tennis? A validity and reproducibility study. *PloS one*, *9*(1), e85657. https://doi.org/10.1371/journal.pone.0085657

Faber, I. R., Pion, J., Munivrana, G., Faber, N. R., & Nijhuis-Van der Sanden, M. W. G. (2018). Does a perceptuomotor skills assessment have added value to detect talent for table tennis in primary school children? *Journal of Sports Sciences*, *6*, 593–601. https://doi.org/10.1080/02640414.2017.1316865

Faber, I. R., Zamoscinska, M., Teunissen, J. W., & Pion, J. (2020). The relationship between sport-specific training and a perceptuo-motor skills assessment as part of talent identification in young table tennis players (8-10 years). *International Journal of Racket Sports Science*, *2*(1), 42–54. https://doi.org/10.30827/Digibug.63720

Fahrmeir, L., Kneib, T., & Lang, S. (2009). *Regression: Modelle, Methoden und Anwendungen* (2. Aufl.). *Statistik und ihre Anwendungen*. Springer Berlin Heidelberg. Abruf unter http://nbn-resolving.org/urn:nbn:de:bsz:31-epflicht-1554927

Farrington, D. P., & Loeber, R. (1989). Relative improvement over chance (RIOC) and phi as measures of predictive efficiency and strength of association in 22 tables. *Journal of Quantitative Criminology*, *5*(3), 201–213. https://doi.org/10.1007/BF01062737

Fernandez-Fernández, J., Ulbricht, A., & Ferrauti, A. (2014). Fitness testing of tennis players: how valuable is it? *British journal of sports medicine*, *48*(1), 22–31.

Ferrauti, A. (Hrsg.). (2020). *Trainingswissenschaft für die Sportpraxis: Lehrbuch für Studium, Ausbildung und Unterricht im Sport* (1st ed. 2020). Springer Berlin Heidelberg; Springer Spektrum.

Ferrauti, A., Schneider, C., & Wiewelhove, T. (2020). Leistungssteuerung. In A. Ferrauti (Hrsg.), *Trainingswissenschaft für die Sportpraxis: Lehrbuch für Studium, Ausbildung und Unterricht im Sport* (1. Aufl., S. 67-186). Springer Berlin Heidelberg; Springer Spektrum.

Fett, J., Ulbricht, A., Wiewelhove, T., & Ferrauti, A. (2017). Athletic performance, training characteristics, and orthopedic indications in junior tennis Davis Cup players. *International Journal of Sports Science & Coaching*, *12*(1), 119–129. https://doi.org/10.1177/1747954116684393

Figueiredo, A. J., Gonçalves, C. E., Coelho e Silva, M. J., & Malina, R. M. (2009). Youth soccer players, 11-14 years: maturity, size, function, skill and goal orientation. *Annals of human biology*, *36*(1), 60–73. https://doi.org/10.1080/03014460802570584

Filipcic, A. (1999). Reliability and validity of motor tests in Tennis. *Sportonomics*, *5*(1), 86–92.

Filipcic, A., & Filipcic, T. (2005). The influence of tennis motor abilities and anthropometric measures on the competition successfulness of 11 and 12 year-old female tennis players. *Acta Universitatis Palackianae Olomucensis. Gymnica*, *35*(2), 35–41.

Filipcic, A., Pisk, L., & Filipcic, T. (2010). Relationship between the results of selected motor tests and competitive successfulness in tennis for different age categories. *Kinesiology*, *42*(2), 175–183.

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. https://doi.org/10.1177/1747954116655051

Franchini, E., Del Vecchio, F. B., Matsushigue, K. A., & Artioli, G. G. (2011). Physiological profiles of elite judo athletes. *Sports medicine (Auckland, N.Z.)*, *41*(2), 147–166. https://doi.org/10.2165/11538580-000000000-00000

Gabler, H., & Ruoff, B. A. (1979). Zum Problem der Talentbestimmung im Sport. Rahmentheoretische Vorueberlegungen. *Sportwissenschaft (Schorndorf)*, *9*(2), S. 164-180.

Gagné, F. (2004). Transforming gifts into talents: the DMGT as a developmental theory1. *High Ability Studies*, *15*(2), 119–147. https://doi.org/10.1080/1359813042000314682

Gil, S., Ruiz, F., Irazusta, A., Gil, J., & Irazusta, J. (2007). Selection of young soccer players in terms of anthropometric and physiological factors. *The Journal of sports medicine and physical fitness*, *47*(1), 25–32.

Giles, B., Kovalchik, S., & Reid, M. (2020). A machine learning approach for automatic detection and classification of changes of direction from player tracking data in professional tennis. *Journal of Sports Sciences*, *38*(1), 106–113. https://doi.org/10.1080/02640414.2019.1684132

Girard, O., & Millet, G. P. (2009). Physical determinants of tennis performance in competitive teenage players. *Journal of strength and conditioning research*, *23*(6), 1867–1872.

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. https://doi.org/10.1080/02640414.2012.713980

Granacher, U., & Borde, R. (2017). Effects of Sport-Specific Training during the Early Stages of Long-Term Athlete Development on Physical Fitness, Body Composition, Cognitive, and Academic Performances. *Frontiers in Physiology*, *8*, 810. https://doi.org/10.3389/fphys.2017.00810

Gravina, L., Gil, S., Ruiz, F., Zubero, J., Gil, J., & Irazusta, J. (2008). Anthropometric and physiological differences between first team and reserve soccer players aged 10-14 years at the beginning and end of the season. *Journal of strength and conditioning research*, *22*(4), 1308–1314. https://doi.org/10.1519/JSC.0b013e31816a5c8e

Grosser, M., Starischka, S., & Zimmermann, E. (2008). *Das neue Konditionstraining: Sportwissenschaftliche Grundlagen, Leistungssteuerung und Trainingsmethoden, Übungen und Trainingsprogramme* (10. Aufl.). *BLV-Sportwissen*. BLV.

Gulbin, J. P., & Ackland, T. (2009). Talent identification and profiling. In (S. 11-26).

Gulbin, J. P., Croser, M. J., Morley, E. J., & Weissensteiner, J. R. (2013). An integrated framework for the optimisation of sport and athlete development: a practitioner approach. *Journal of Sports Sciences*, *31*(12), 1319–1331. https://doi.org/10.1080/02640414.2013.781661

Güllich, A. (2013). Talente im Sport. In A. Güllich & M. Krüger (Hrsg.), *Bachelor. Sport: Das Lehrbuch für das Sportstudium* (1. Aufl., S. 623-654). Springer Berlin Heidelberg.

Güllich, A. (2019). Talente im Sport: Talententwicklung, Talenterkennung und Talentförderung. In A. Güllich & M. Krüger (Hrsg.), *Grundlagen von Sport und Sportwissenschaft* (S. 1-18). SPRINGER. https://doi.org/10.1007/978-3-662-53384-0_11-1

Güllich, A., & Cobley, S [Stephen]. (2019). On the efficacy of talent identification and talent development programmes. In J. Baker, S. Cobley, J. Schorer & N. Wattie (Hrsg.), *Routledge international handbooks. Routledge handbook of talent identification and development in sport* (S. 80-98). Routledge.

Güllich, A., & Emrich, E. (2014). Considering long-term sustainability in the development of world class success. *European journal of sport science*, *14 Suppl 1*, S383-97. https://doi.org/10.1080/17461391.2012.706320

Güllich, A., & Krüger, M. (Hrsg.). (2013). *Bachelor*. *Sport: Das Lehrbuch für das Sportstudium* (1. Aufl.). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-37546-0

Güllich, A., & Krüger, M. (Hrsg.). (2019). *Grundlagen von Sport und Sportwissenschaft*. SPRINGER. https://doi.org/10.1007/978-3-662-53384-0

Hagenbuchner, M., Cliff, D. P., Trost, S. G., van Tuc, N., & Peoples, G. E. (2015). Prediction of activity type in preschool children using machine learning techniques. *Journal of Science and Medicine in Sport*, *18*(4), 426–431. https://doi.org/10.1016/j.jsams.2014.06.003

Hahn, E. (1982). *Kindertraining: Probleme, Trainingstheorie und Praxis*. *BLV-Sportwissen: Bd. 405*. BLV Verl.-Ges.

Heazlewood, I., Walsh, J., Climstein, M., Kettunen, J., Adams, K., & DeBeliso, M. (2016). A Comparison of Classification Accuracy for Gender Using Neural Networks Multilayer Perceptron (MLP), Radial Basis Function (RBF) Procedures Compared to Discriminant Function Analysis and Logistic Regression Based on Nine Sports Psychological Constructs to Measure Motivations to Participate in Masters Sports Competing at the 2009 World Masters Games. In P. Chung, A. Soltoggio, C. W. Dawson, Q. Meng & M. Pain (Hrsg.), *Advances in Intelligent Systems and Computing. Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS)* (Bd. 392, S. 93-101). Springer International Publishing. https://doi.org/10.1007/978-3-319-24560-7_12

Hedderich, J., & Sachs, L. (2016). *Angewandte Statistik*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-45691-0

Heller, K. A., Perleth, C., & Lim, T. K. (2005). The Munich Model of giftedness designed to identify and promote gifted students. In R. J. Sternberg & J. E. Davidson (Hrsg.), *Conceptions of giftedness* (2. Aufl., S. 147-170). Cambridge University Press.

Herrmann, C., Heim, C., & Seelig, H. (2019). Construct and correlates of basic motor competencies in primary school-aged children. *Journal of sport and health science*, *8*(1), 63–70. https://doi.org/10.1016/j.jshs.2017.04.002

Hohmann, A. (2001). Leistungsdiagnostische Kriterien sportlichen Talents - dargestellt am Beispiel des leichtathletischen Sprints. *Leistungssport*, *31*(4), 14–22.

Hohmann, A. (2009). *Entwicklung sportlicher Talente an sportbetonten Schulen: Schwimmen, Leichtathletik, Handball*. Imhof.

Hohmann, A., Fehr, U., & Voigt, L. (2015). Heute im Talentpool – In Hamburg auf dem Podium. *Leistungssport*, *45*(5), 5–11.

Hohmann, A., Lames, M., & Letzelter, M. (2014). *Einführung in die Trainingswissenschaft* (6., unveränderte Auflage). Limpert Verlag.

Hohmann, A., Lames, M., Letzelter, M., & Pfeiffer, M. (2020). *Einführung in die Trainingswissenschaft*. Limpert Verlag GmbH.

Hohmann, A., & Seidel, I. (2005). Talententwicklung im Leistungssport: Die Magdeburger Talent- und Schnelligkeitsstudie MATASS. *BISp-Jahrbuch: Forschungsförderung …*, *2004*, S. 185-196.

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*(8), e0182211. https://doi.org/10.1371/journal.pone.0182211

Höner, O., Raabe, J., Murr, D., & Leyhr, D. (2019). Prognostic relevance of motor tests in elite girls' soccer: a five-year prospective cohort study within the German talent promotion program. *Science and Medicine in Football*, *3*(4), 287–296. https://doi.org/10.1080/24733938.2019.1609069

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. https://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. https://doi.org/10.1080/02640414.2014.928416

Howe, M. J. A., Davidson, J. W., & Sloboda, J. A. (1898). Innate talents: Reality or myth? Behavioral and Brain Sciences, 21(3): p. 399 - 442. *Behavioral and Brain Sciences*, *21*(3), 399–442.

Huijgen, B. C. H [B. C. H.], Elferink-Gemser, M. T., Ali, A., & Visscher, C [C.] (2013). Soccer skill development in talented players. *International Journal of Sports Medicine*, *34*(8), 720–726. https://doi.org/10.1055/s-0032-1323781

Hunter, J. E., & Schmidt, F. L. (2007). *Methods of meta-analysis: Correcting error and bias in research findings* (2. Aufl.). *35*. Sage.

International Table Tennis Federation. (2020). *ITTF - World Ranking*. Abruf unter https://ranking.ittf.com/#/home

Issurin, V. B. (2017). Evidence-Based Prerequisites and Precursors of Athletic Talent: A Review. *Sports medicine (Auckland, N.Z.)*, *47*(10), 1993–2010. https://doi.org/10.1007/s40279-017-0740-0

Joch, W. (1994). *Das sportliche Talent: Talenterkennung - Talentförderung - Talentperspektiven* (2. Aufl.). *Edition Sport & Wissenschaft: Bd. 15*. Meyer und Meyer.

Johansson, A., & Fahlén, J. (2017). Simply the best, better than all the rest? Validity issues in selections in elite sport. *International Journal of Sports Science & Coaching*, *12*(4), 470–480. https://doi.org/10.1177/1747954117718020

Johnston, K., Wattie, N., Schorer, J., & Baker, J. (2018). Talent Identification in Sport: A Systematic Review. *Sports medicine (Auckland, N.Z.)*, *48*(1), 97–109. https://doi.org/10.1007/s40279-017-0803-2

Jokuschies, N., Gut, V., & Conzelmann, A. (2017). Systematizing coaches' 'eye for talent': Player assessments based on expert coaches' subjective talent criteria in top-level youth soccer. *International Journal of Sports Science & Coaching*, *12*(5), 565–576. https://doi.org/10.1177/1747954117727646

Klein, M., Fröhlich, M., & Emrich, E. (2012). Zur Testgenauigkeit ausgewählter Items des Deutschen Motorik-Tests DMT 6-18. *Leipziger sportwissenschaftliche Beiträge*, *53*(1), 23–45.

Kolias, P., Stavropoulos, N., Papadopoulou, A., & Kostakidis, T. (2021). Evaluating basketball player's rotation line-ups performance via statistical markov chain modelling. *International Journal of Sports Science & Coaching*, 174795412110090. https://doi.org/10.1177/17479541211009083

Kondric, M., & Filipcic, A. (2009). *Scientific approach in table tennis and tennis in Slovenia*. Sport Books Publisher.

Koopmann, T., Faber, I. R., Baker, J., & Schorer, J. (2020). Assessing Technical Skills in Talented Youth Athletes: A Systematic Review. *Sports medicine (Auckland, N.Z.), 50*(9), 1593–1611. https://doi.org/10.1007/s40279-020-01299-4

Köster, P., Hohmann, A., Niessner, C., & Siener, M. (2021). Health-Related Motor Testing of Children in Primary School: A Systematic Review of Criterion-Referenced Standards. *Children (Basel, Switzerland), 8*(11). https://doi.org/10.3390/children8111046

Kovacs, M. (2006). Applied physiology of tennis performance. *British journal of sports medicine*, *40*(5), 381–386. https://doi.org/10.1136/bjsm.2005.023309

Kovacs, M. (2007). Tennis physiology: training the competitive athlete. *Sports Medicine*(37), 189–198. https://doi.org/10.2165/00007256-200737030-00001

Kovalchik, S. A., & Reid, M. (2017). Comparing Matchplay Characteristics and Physical Demands of Junior and Professional Tennis Athletes in the Era of Big Data. *Journal of sports science & medicine*, *16*(4), 489–497.

Kovar, R. (1981). *Human Variation in Motor Abilitiies and ist Genetic Analysis*. Faculty of Physical Education and Sport.

Kramer, T., Huijgen, B. C. H., Elferink-Gemser, M. T., & Visscher, C [Chris] (2016). A Longitudinal Study of Physical Fitness in Elite Junior Tennis Players. *Pediatric Exercise Science*, *28*(4), 553–564. https://doi.org/10.1123/pes.2016-0022

Kramer, T., Huijgen, B. C., Elferink-Gemser, M. T., & Visscher, C [Chris] (2017). Prediction of tennis performance in junior elite tennis players. *Journal of sports science and medicine*, *16*(1), 14–21.

Kubinger, K. D. (1996). *Einführung in die psychologische Diagnostik* (2., korrigierte Aufl.). Beltz Psychologie-Verl.-Union.

Kunz, M. (2007). 265 million playing football. *FIFA Magazine*, *7*, 11–15.

Kupper, K. (1993). Theorie und Methodologie der Talenterkennung im Sport. In G. Senf (Hrsg.), *Talenterkennung und -förderung im Sport* (S. 2-24). Academia.

LaPrade, R. F., Agel, J., Baker, J., Brenner, J. S., Cordasco, F. A., Côté, J., Engebretsen, L., Feeley, B. T., Gould, D., Hainline, B., Hewett, T., Jayanthi, N., Kocher, M. S., Myer, G. D., Nissen, C. W., Philippon, M. J., & Provencher, M. T. (2016). AOSSM Early Sport Specialization Consensus Statement. *Orthopaedic journal of sports medicine*, *4*(4), 2325967116644241. https://doi.org/10.1177/2325967116644241

Larkin, P., & Reeves, M. J. (2018). Junior-elite football: time to re-position talent identification? *Soccer & Society*, 1–10. https://doi.org/10.1080/14660970.2018.1432389

Le Gall, F., Carling, C., Williams, M., & Reilly, T. (2010). Anthropometric and fitness characteristics of international, professional and amateur male graduate soccer players from an elite youth academy. *Journal of science and medicine in sport*, *13*(1), 90–95. https://doi.org/10.1016/j.jsams.2008.07.004

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*(5), e0196324. https://doi.org/10.1371/journal.pone.0196324

Leyhr, D., Raabe, J., Schultz, F., Kelava, A., & Höner, O. (2020). 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

Li, P., Bosscher, V. de, & Weissensteiner, J. R. (2018). The journey to elite success: a thirty-year longitudinal study of the career trajectories of top professional tennis players. *International Journal of Performance Analysis in Sport*, *18*(6), 961–972. https://doi.org/10.1080/24748668.2018.1534197

Li, P., Weissensteiner, J. R., Pion, J., & Bosscher, V. de (2020). Predicting elite success: Evidence comparing the career pathways of top 10 to 300 professional tennis players. *International Journal of Sports Science & Coaching*, *15*(5-6), 793–802. https://doi.org/10.1177/1747954120935828

Lidor (2007). Measurement of talent in volleyball15-month follow-up of elite adolescent players. *J. Sports Med. Phys. Fit.*, *47*, 159.

Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). *Grundlagen Psychologie*. Beltz. Abruf unter http://www.content-select.com/index.php?id=bib_view&ean=9783621278454

Ljach, W. I. (1998). Kinderhochleistungssport in Russland. In R. Daugs (Hrsg.), *Schriftenreihe des Bundesinstituts für Sportwissenschaft: Bd. 95. Kinder und Jugendliche im Leistungssport: Beiträge des Internationalen Interdisziplinären Symposiums "Kinderleistungen" vom 7. bis 10. November 1996 in Saarbrücken* (1. Aufl., S. 203-216). Hofmann.

Der Magistrat der Stadt Fulda. (19. Juni 2019). *Stadt Fulda: Fachservice für Partner der Kinder- und Jugendhilfe: Sozialmonitoring Schwerpunktstadtteile 2010-2018*. Abruf unter https://www.fulda.de/fd/51_Amt_fuer_Jugend__Familie_und_Senioren/51.4_Service_fuer_Kooperationspartner/Staerken_Sozialer_Netze/Sozialmonitoring_Stadt_Fulda_Schwerpunktstadtteile_Stand_19.6.2019.pdf

Malina, R. M., Ribeiro, B., Aroso, J., & Cumming, S. P. (2007). Characteristics of youth soccer players aged 13-15 years classified by skill level. *British Journal of Sports Medicine*, *41*(5), 290-5; discussion 295. https://doi.org/10.1136/bjsm.2006.031294

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*, *5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Mann, D. L., Dehghansai, N., & Baker, J. (2017). Searching for the elusive gift: advances in talent identification in sport. *Current opinion in psychology*, *16*, 128–133. https://doi.org/10.1016/j.copsyc.2017.04.016

Marconi, M. (2013). *Die Identifikation von Talenten im Sport aus entwicklungstheoretischer Perspektive: Eine vierjährige Längsschnittstudie an Schweizer Nachwuchsskirennfahrerinnen und -fahrern*. Selbstverlag.

Martin, D., Nicolaus, J., Ostrowski, C., & Rost, K. (1999). *Handbuch Kinder- und Jugendtraining*. *Beiträge zur Lehre und Forschung im Sport, 125*. Hofmann Verlag.

Marx, P., & Lenhard, W. (2010). Diagnostische Merkmale von Screeningverfahren. In M. Hasselhorn & W. Schneider (Hrsg.), *Frühprognose schulischer Kompetenzen.* Hogrefe.

Matsudo, V. K. R. (1996). Prediction of Future Athletic Excellence. In O. Bar-Or (Hrsg.), *The Child and Adolescent Athlete* (S. 92-109). Blackwell Science Distributors Human Kinetics Books.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, *405*(2), 442–451. https://doi.org/10.1016/0005-2795(75)90109-9

Meckel, Y., Hophy, A., Dunsky, A., & Eliakim, A. (2015). Relationships between physical characteristics and ranking of young tennis players. *Central European Journal of Sport Sciences and Medicine*, *10*(2), 5–12.

Mohamed, H., Vaeyens, R., Matthys, S., Multael, M., Lefevre, J., Lenoir, M., & Philppaerts, R. (2009). Anthropometric and performance measures for the development of a talent detection and identification model in youth handball. *Journal of Sports Sciences*, *27*(3), 257–266. https://doi.org/10.1080/02640410802482417

Mosoi, A. A. (2013). Skills and Motivation of Junior Tennis Players. *Procedia - Social and Behavioral Sciences*, *78*, 215–219. https://doi.org/10.1016/j.sbspro.2013.04.282

Mostaert, M., Laureys, F., Vansteenkiste, P., Pion, J., Deconinck, F. J. A., & Lenoir, M. (2021). Discriminating performance profiles of cycling disciplines. *International Journal of Sports Science & Coaching*, *16*(1), 110–122. https://doi.org/10.1177/1747954120948146

Mostaert, M., Pion, J., Lenoir, M., & Vansteenkiste, P. (2020). A Retrospective Analysis of the National Youth Teams in Volleyball: Were They Always Faster, Taller, and Stronger? *Journal of strength and conditioning research.* Vorab-Onlinepublikation. https://doi.org/10.1519/JSC.0000000000003847

Munivrana, G., Furjan-Mandić, G., & Kondrič, M. (2015). Determining the structure and evaluating the role of technical-tactical elements in basic table tennis playing systems. *International Journal of Sports Science & Coaching*, *10*, 111–132.

Murphy, A. H., & Winkler, R. L. (1977). Reliability of Subjective Probability Forecasts of Precipitation and Temperature. *Applied Statistics*, *26*(1), 41. https://doi.org/10.2307/2346866

Murr, D., Raabe, J., & Höner, O. (2018). The prognostic value of physiological and physical characteristics in youth soccer: A systematic review. *European journal of sport science*, *18*(1), 62–74. https://doi.org/10.1080/17461391.2017.1386719

Musa, R. M., Majeed, A., Taha, Z., Abdullah, M. R., Maliki, A. B. H. M., & Kosni, N. A. (2019). The application of Artificial Neural Network and k-Nearest Neighbour classification models in the scouting of high-performance archers from a selected fitness and motor skill performance parameters. *Science & Sports*, *34*(4), e241-e249. https://doi.org/10.1016/j.scispo.2019.02.006

Musa, R. M., Taha, Z., Majeed, A., & Razali, A. M. (2019). *Machine Learning in Sports: Identifying Potential Archers*. *SpringerBriefs in Applied Sciences and Technology*. Springer Singapore; Imprint: Springer. https://doi.org/10.1007/978-981-13-2592-2

Mustafović, E., Čaušević, D., Covic, N., Ibrahimović, M., Alić, H., Abazović, E., & Mašić, S. (2020). Talent Identification in Youth Football: A Systematic Review. *Journal of Anthropology of Sport and Physical Education*, *4*, 37–43. https://doi.org/10.26773/jaspe.201007

Niessner, C., Utesch, T., Oriwol, D., Hanssen-Doose, A., Schmidt, S. C. E., Woll, A., Bös, K., & Worth, A. (2020). Representative Percentile Curves of Physical Fitness From Early Childhood to Early Adulthood: The MoMo Study. *Frontiers in public health*, *8*, 458. https://doi.org/10.3389/fpubh.2020.00458

Nikolić, I., Furjan-Mandić, G., & Kondric, M. (2014). The relationship of morphology and motor abilities to specific table tennis tasks in youngsters. *Collegium antropologicum*, *38*(1), 241–245.

Ochs, J. (2019). *Erarbeitung eines Modells zur Gewichtung von motorischen Fähigkeiten zur Sportartenempfehlung von Kindern im Rahmen des Fuldaer Bewegungschecks* [Masterarbeit]. Universität Bayreuth, Bayreuth.

Opstoel, K., Pion, J., Elferink-Gemser, M. T., Hartman, E., Willemse, B., Philippaerts, R. M., Visscher, C [Chris], & Lenoir, M. (2015). Anthropometric characteristics, physical fitness and motor coordination of 9 to 11 year old children participating in a wide range of sports. *PloS one*, *10*(5), e0126282. https://doi.org/10.1371/journal.pone.0126282

Pechtl, V., Ostrowski, C., & Klose, S. (1993). Positionen zur Erstellung bundeseinheitlicher Kaderkriterien. In Bundesvorstand Leistungssport (Hrsg.), *Beiträge zur Förderung im Nachwuchsleistungssport* (S. 7-30). DSB.

Pfeiffer, M., & 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

Pickering, C., Kiely, J., Grgic, J., Lucia, A., & Del Coso, J. (2019). Can Genetic Testing Identify Talent for Sport? *Genes*, *10*(12). https://doi.org/10.3390/genes10120972

Pion, J. (2015). *The Flemish Sports Compass: from sports orientation to elite performance prediction.* University Press.

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. https://doi.org/10.1080/02640414.2016.1192669

Pion, J., Segers, V., Fransen, J., Debuyck, G., Deprez, D., Haerens, L., Vaeyens, R., Philippaerts, R. M., & Lenoir, M. (2015). Generic anthropometric and performance characteristics among elite adolescent boys in nine different sports. *European journal of sport science*, *15*(5), 357–366. https://doi.org/10.1080/17461391.2014.944875

Pion, J., Teunissen, J. W., Welle, S. t., Spruijtenburg, G., Faber, I. R., & Lenoir, M. (2021). How Similarities and Differences between Sports Lead to Talent Transfer. In J. Baker, J. Schorer & S. Cobley (Hrsg.), *Talent identification and development in sport: International perspectives* (S. 184-196). Routledge. https://doi.org/10.4324/9781003049111-13

Potočnik, Ž. L., Jurak, G., & Starc, G. (2020). Secular Trends of Physical Fitness in Twenty-Five Birth Cohorts of Slovenian Children: A Population-Based Study. *Frontiers in public health*, *8*, 561273. https://doi.org/10.3389/fpubh.2020.561273

Reid, M., & Morris, C. (2013). Ranking benchmarks of top 100 players in men's professional tennis. *European journal of sport science*, *13*(4), 350–355. https://doi.org/10.1080/17461391.2011.608812

Reilly, T., Williams, A. M., Nevill, A., & Franks, A. (2000). A multidisciplinary approach to talent identification in soccer. *Journal of Sports Sciences*, *18*(9), 695–702. https://doi.org/10.1080/02640410050120078

Rey, G. D. (2017). *Methoden der Entwicklungspsychologie: Datenerhebung und Datenauswertung* (2., überarbeitete Auflage). Books on Demand.

Ribeiro Júnior, D. B., Werneck, F. Z., Oliveira, H. Z., Panza, P. S., Ibáñez, S. J., & Vianna, J. M. (2021). From Talent Identification to Novo Basquete Brasil (NBB): Multifactorial Analysis of the Career Progression in Youth Brazilian Elite Basketball. *Frontiers in psychology*, *12*, 617563. https://doi.org/10.3389/fpsyg.2021.617563

Robertson, K., Pion, J., Mostaert, M., Norjali Wazir, Mohd Rozilee Wazir, Kramer, T., Faber, I. R., Vansteenkiste, P., & Lenoir, M. (2018). A coaches' perspective on the contribution of anthropometry, physical performance, and motor coordination in racquet sports. *Journal of Sports Sciences*, *36*(23), 2706–2715. https://doi.org/10.1080/02640414.2018.1441941

Robertson, S., Woods, C., & Gastin, P. (2015). Predicting higher selection in elite junior Australian Rules football: The influence of physical performance and anthropometric attributes. *Journal of science and medicine in sport*, *18*(5), 601–606. https://doi.org/10.1016/j.jsams.2014.07.019

Robertson, S. J., Burnett, A. F., & Cochrane, J. (2014). Tests examining skill outcomes in sport: a systematic review of measurement properties and feasibility. *Sports medicine (Auckland, N.Z.)*, *44*(4), 501–518. https://doi.org/10.1007/s40279-013-0131-0

Roetert, E. P., Garrett, G. E., Brown, S. W., & Camaione, D. N. (1992). Performance Profiles of Nationally Ranked Junior Tennis Players. *Journal of Applied Sport Science Research*, *6*(4), 225–231. https://doi.org/10.1519/1533-4287

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: an expanded typology. *Journal of Applied Psychology*, *85*(1), 112–118. https://doi.org/10.1037/0021-9010.85.1.112

Salo, A. I. T., Bezodis, I. N., Batterham, A. M., & Kerwin, D. G. (2011). Elite sprinting: are athletes individually step-frequency or step-length reliant? *Medicine and science in sports and exercise*, *43*(6), 1055–1062. https://doi.org/10.1249/MSS.0b013e318201f6f8

Sarmento, H., Anguera, M. T., Pereira, A., & Araújo, D. (2018). Talent Identification and Development in Male Football: A Systematic Review. *Sports medicine (Auckland, N.Z.)*, *48*(4), 907–931. https://doi.org/10.1007/s40279-017-0851-7

Schmidt-Atzert, L., Amelang, M., Fydrich, T., & Moosbrugger, H. (2018). *Psychologische Diagnostik* (5. Aufl.). *Lehrbuch*. SPRINGER. Abruf unter http://www.springer.com/

Schnabel, G., Harre, H.-D., & Krug, J. (2016). *Trainingslehre - Trainingswissenschaft: Leistung - Training - Wettkampf* (3. Auflage). Meyer & Meyer.

Schorer, J., Rienhoff, R., Fischer, L., & Baker, J. (2017). Long-Term Prognostic Validity of Talent Selections: Comparing National and Regional Coaches, Laypersons and Novices. *Frontiers in psychology*, *8*, 1146. https://doi.org/10.3389/fpsyg.2017.01146

Seibold, M. (2011). *European Tennis: A Comparative Analysis of Talent Identification and Development (TID)*. University of Edinburgh. Abruf unter https://pdfs.semanticscholar.org/0ee0/319e9e25d79b3d28496dbcfee0193cb06edb.pdf

Senf, G. (Hrsg.). (1993). *Talenterkennung und -förderung im Sport*. Academia.

Serpiello, F. R., Cox, A., Oppici, L., Hopkins, W. G., & Varley, M. C. (2017). The Loughborough Soccer Passing Test has impractical criterion validity in elite youth football. *Science and Medicine in Football*, *1*(1), 60–64. https://doi.org/10.1080/02640414.2016.1254810

Sieghartsleitner, R. (2020). *Talentforschung im Fussball zwischen holistisch-interaktionistischer Theorie und reduktionistischer Empirie* [Dissertation]. DataCite.

Sieghartsleitner, R., Zuber, C., Zibung, M., Charbonnet, B., & Conzelmann, A. (2019). Talent selection in youth football: Technical skills rather than general motor performance predict future player status of football talents. *Current Issues in Sport Science (CISS)*, *4*(11). https://doi.org/10.15203/CISS_2019.011

Sieghartsleitner, R., Zuber, C., Zibung, M., & Conzelmann, A. (2017). *Talentsektion mittels Testbatterie oder Bauchgefühl? Sowohl als auch!*

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 Elite Youth Football. *Journal of sports science & medicine*, *18*, 32–43.

Silva, A. J., Costa, A. M., Oliveira, P. M., Reis, V. M., Saavedra, J., Perl, J., Rouboa, A., & Marinho, D. A. (2007). The use of neural network technology to model swimming performance. *Journal of sports science & medicine*, *6*(1), 117–125.

Singer, R. (1981). Allgemeine methodische Probleme der Talentbestimmung im Sport. In *Mainzer Studien zur Sportwissenschaft: 5/6. Leichtathletiktraining im Spannungsfeld von Wissenschaft und Praxis: Arbeitsbericht des Internationalen DLV-Fortbildungskongresses "Leichtathletiktraining vor Moskau" vom 23. - 25. 11. 1979 am Fachbereich Sport der Universität Mainz* (S. 14-27). Schors.

Siris, P. (1974). Das Wachstumstempo der motorischen Eigenschaften – Ein Faktor der potentiellen Möglichkeiten von Sportlern. *Leistungssport*, *4*(5), 339–341.

Slimani, M., & Nikolaidis, P. T. (2019). Anthropometric and physiological characteristics of male soccer players according to their competitive level, playing position and age group: a systematic review. *The Journal of sports medicine and physical fitness*, *59*(1), 141–163. https://doi.org/10.23736/S0022-4707.17.07950-6

Smirmaul, B. P. C., Bertucci, D. R., & Teixeira, I. P. (2013). Is the VO2max that we measure really maximal? *Frontiers in Physiology*, *4*, 203. https://doi.org/10.3389/fphys.2013.00203

Spies, F., Schauer, L., Bindel, T., & Pfeiffer, M. (2022). Talent detection—importance of the will and the ability when starting a sport activity. *German Journal of Exercise and Sport Research.* Vorab-Onlinepublikation. https://doi.org/10.1007/s12662-022-00796-0

Statistisches Bundesamt. (9. September 2021). *Städte (Alle Gemeinden mit Stadtrecht) nach Fläche, Bevölkerung und Bevölkerungsdichte am 31.12.2020.* Abruf unter https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/05-staedte.html;jsessionid=6F3203AA0B6FB60513108B9A0C0657A6.live732

Taylor, H., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *Journal of Applied Psychology*(23), 565–578. Abruf unter https://www.semanticscholar.org/paper/The-relationship-of-validity-coefficients-to-the-of-Taylor-Russell/6e00784fccb1d7302ce1b48319cdb9b664842f1c

Teunissen, J. W., Welle, S. S. ter, Platvoet, S. S., Faber, I. R., Pion, J., & Lenoir, M. (2021). Similarities and differences between sports subserving systematic talent transfer and development: The case of paddle sports. *Journal of Science and Medicine in Sport*, *24*(2), 200–205. https://doi.org/10.1016/j.jsams.2020.09.005

Thiess, G. (1979). Möglichkeiten der Eignungsdiagnostik. In H.-D. Harre (Hrsg.), *Trainingslehre* (S. 34-40). Sportverlag.

Till, K., & Baker, J. (2020). Challenges and Possible Solutions to Optimizing Talent Identification and Development in Sport. *Frontiers in psychology*, *11*, 664. https://doi.org/10.3389/fpsyg.2020.00664

Till, K., Cobley, S [Steve], O'Hara, J., Morley, D., Chapman, C., & Cooke, C. (2015). Retrospective analysis of anthropometric and fitness characteristics associated with long-term career progression in Rugby League. *Journal of science and medicine in sport*, *18*(3), 310–314. https://doi.org/10.1016/j.jsams.2014.05.003

Till, K., Jones, B. L., Cobley, S [Stephen], Morley, D., O'Hara, J., Chapman, C., Cooke, C., & Beggs, C. B. (2016). Identifying Talent in Youth Sport: A Novel Methodology Using Higher-Dimensional Analysis. *PloS one*, *11*(5), e0155047. https://doi.org/10.1371/journal.pone.0155047

Tomkinson, G., Carver, K., Atkinson, F., Daniell, N., Lewis, L., Fitzgerald, J., Lang, J., & Ortega, F. (2017). European normative values for physical fitness in children and adolescents aged 9-17 years: Results from 2 779 165 Eurofit performances representing 30 countries. *British Journal of Sports Medicine*, *52*, bjsports-2017. https://doi.org/10.1136/bjsports-2017-098253

True, L., Martin, E. M., Pfeiffer, K. A., Siegel, S. R., Branta, C. F., Haubenstricker, J., & Seefeldt, V. (2021). Tracking of Physical Fitness Components from Childhood to Adolescence: A Longitudinal Study. *Measurement in Physical Education and Exercise Science*, *25*(1), 22–34. https://doi.org/10.1080/1091367X.2020.1729767

Ulbricht, A., Fernandez-Fernández, J., Mendez-Villanueva, A., & Ferrauti, A. (2016). Impact of fitness characteristics on tennis performance in elite junior tennis players. *Journal of strength and conditioning research*, *30*(4), 989–998.

Unal, I. (2017). Defining an Optimal Cut-Point Value in ROC Analysis: An Alternative Approach. *Computational and mathematical methods in medicine*, *2017*, 3762651. https://doi.org/10.1155/2017/3762651

Utesch, T., Strauß, B., Tietjens, M., Büsch, D., Ghanbari, M.-C., & Seidel, I. (2015). Die Überprüfung der Konstruktvalidität des Deutschen Motorik-Tests 6-18 für 9- bis 10-Jährige. *Zeitschrift für Sportpsychologie*, *22*(2), 77–90. https://doi.org/10.1026/1612-5010/a000143

Vaeyens, R., Güllich, A., Warr, C. R., & Philippaerts, R. M. (2009). Talent identification and promotion programmes of Olympic athletes. *Journal of Sports Sciences*, *27*(13), 1367–1380. https://doi.org/10.1080/02640410903110974

Vaeyens, R., Lenoir, M., Williams, A. M., & Philippaerts, R. M. (2008). Talent identification and development programmes in sport : current models and future directions. *Sports medicine (Auckland, N.Z.)*, *38*(9), 703–714. https://doi.org/10.2165/00007256-200838090-00001

Vaeyens, R., Malina, R. M., 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), 928-34; discussion 934. https://doi.org/10.1136/bjsm.2006.029652

Vandorpe, B., Vandendriessche, J. B., Vaeyens, R., Pion, J., Lefevre, J., Philippaerts, R. M., & Lenoir, M. (2012). The value of a non-sport-specific motor test battery in predicting performance in young female gymnasts. *Journal of Sports Sciences*, *30*(5), 497–505. https://doi.org/10.1080/02640414.2012.654399

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

Willimczik, K. (1982). Determinanten der sportmotorischen Leistungsfähigkeit im Kindesalter. In H. Howald & E. Hahn (Hrsg.), *Kinder im Leistungssport.* (S. 141-153). Birkhäuser.

Wirtz, M. A. (Hrsg.). (2017). *Dorsch - Lexikon der Psychologie* (18., überarbeitete Auflage). Hogrefe.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*(3), 32–35.

Zhou, Q. M., Zhe, L., Brooke, R. J., Hudson, M. M., & Yuan, Y. (2021). A relationship between the incremental values of area under the ROC curve and of area under the precision-recall curve. *Diagnostic and prognostic research*, *5*(1), 13. https://doi.org/10.1186/s41512-021-00102-w

Zuber, C., Zibung, M., & Conzelmann, A. (2015). Motivational patterns as an instrument for predicting success in promising young football players. *Journal of Sports Sciences*, *33*(2), 160–168. https://doi.org/10.1080/02640414.2014.928827

Zuber, C., Zibung, M., & 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*. https://doi.org/10.3389/fpsyg.2016.01088

Appendix

Artikel 1

Hohmann*, A., Fehr, U., Siener, M., & Hochstein, S. (2017). Talentscreening und Talentorientierung: Zur Validität von Bewegungs-Checks im Grundschulalter. *Leistungssport*, *47*(4), 11–21.

Andreas Hohmann/Ulrich Fehr/Max Siener/Stefan Hochstein

TALENTSCREENING UND TALENTORIENTIERUNG[1]

Zur Validität von Bewegungs-Checks im Grundschulalter

1. Talentscreening und Talentorientierung

In der jüngeren Vergangenheit wurde eine intensive Debatte um die Zukunft des Leistungssports in Deutschland geführt. Im Rahmen der Förderung des Spitzensports wurden neben der Aufwertung des Trainerberufs und der Dualen Karriere der Spitzensportler insbesondere auch der langfristige Leistungsaufbau im Nachwuchstraining (Hoffmann & Pfützner, 2013/ 2014) und die damit in engem Zusammenhang stehende **vielseitige Sichtung und Sportartorientierung von Talenten** (DOSB, 2013) in Richtung einer individuell am besten geeigneten Zielsportart hervorgehoben.

In einem vorangegangenen Beitrag zu Talentscreening und Talentorientierung – als erste diagnostische Maßnahmen im Prozess der Talentidentifikation – haben Hohmann, Fehr und Voigt (2015) gefordert, dass bereits frühzeitig nach sportlich Hochbegabten gesucht werden sollte, um einen umfangreichen und leistungssportlich gehaltvollen Talentpool aufzubauen. Darauf aufbauend konnte am Beispiel bayerischer Kadersportler im Alter von 12 bis 16 Jahren gezeigt werden, dass eine mehrseitige motorische Talentdiagnose (*talent screening*) mit Hilfe von allgemeinen Testverfahren, wie z. B. dem Deutschen Motorik-Test 6-18 (Bös & Schlenker, 2016; plus Ballweitwurf), und die aus den individuellen Testergebnissen abgeleitete Sportartempfehlung (*talent orientation*) im Rahmen der Talentprognose aus forschungsmethodischer Sicht hinreichend treffsichere Maßnahmen darstellen.

Für eine umfassende Rechtfertigung der mittlerweile in mehreren Bundesländern und Kommunen sowie z. B. in Belgien (Pion et al., 2013), der Schweiz (Fuchslocher, Romann, Rüdisüli, Birrer & Hollenstein, 2011) oder Tschechien (Brezina, 2016) eingeführten Bewegungs-Checks auf der zweiten oder dritten Klassenstufe der Grundschule (Abb. 1) erscheint es jedoch erforderlich, den Validitätsnachweis aus konzeptioneller und methodischer Sicht zu vertiefen.

Das in diesem Beitrag vorgestellte, umfassende **Validitätskonzept** – einschließlich der dabei konkret vorgeschlagenen **Prüfschritte** – dient dem Ziel, die in den vergangenen Jahren intensiv geführte Diskussion um die Einführung flächendeckender Bewegungs-Checks in Deutschland (hierzu ausführlich Herrmann, Bartz, Lischka & Spahl, 2016) inhaltlich weiterzuentwickeln. Nachdem in den Workshops „Berlin hat Talent" 2015 in Berlin (Zinner, Werner, Mehler & Abraldes Rois, 2015) und „Internationale Bewegungs-Checks" 2016 in Frankfurt am Main Fragen der Testauswahl, Testauswertung und Teilnehmerpopulation im Mittelpunkt standen, geht es im vorliegenden Beitrag mehr um die generelle Werthaltigkeit im Sinne von Treffsicherheit, Wirksamkeit und Nützlichkeit von Bewegungs-Checks im Rahmen der Talentsichtung und Talentorientierung in

Eingegangen: 19.5.2017

[1] Gefördert durch das Bundesinstitut für Sportwissenschaft (BISp) unter Aktenzeichen ZMVI4 – 070504/15-16

Abbildung 1: Landkarte mit ausgewählten Bewegungs-Checks in Deutschland

der Grundschule. Der vorliegende Beitrag wurde zum frühestmöglichen Zeitpunkt verfasst, ab welchem die Auswertung der Längsschnittdaten aus den Jahren 2010 bis 2016 überhaupt statistisch gerechtfertigt und sinnvoll erscheint. Der Grund für die Inkaufnahme der damit unweigerlich verbundenen Auswertungsprobleme besteht darin, dass die in den zurückliegenden zwei Jahren begonnene, fruchtbare Diskussion rund um flächendeckende Bewegungs-Checks trotz des Abbruchs der deutschen Olympiabewerbung Hamburg 2024 nicht abreißen sollte. Darüber hinaus soll der Beitrag einen Übergang von der bislang dominierenden subjektiven und meist punktuellen Interpretation von Einzelfällen der Individualentwicklung talentierter Spitzensportler (Joch, 1992; Issurin & Lustig, 2008; Güllich, 2013) zu einer objektiven und längsschnittlichen, empirisch fundierten Befundinterpretation in Bezug auf eine gesamte Talententwicklungskampagne (z. B. Zibung & Conzelmann, 2010; Höner & Votteler, 2016) aufzeigen. Nicht zuletzt geht es auch darum, die Möglichkeiten und Grenzen der Prüfmethoden auszuloten und zur Diskussion zu stellen.

2. Das umfassende Validitätskonzept eines Bewegungs-Checks

Kommunale Träger und insbesondere die Eltern, Lehrer und Trainer sollen überzeugt werden, dass talentierten Kindern früh- oder zumindest rechtzeitig der Weg in eine individuell optimal passende Zielsportart geebnet werden sollte. Der Bewegungs-Check im Rahmen der Talentsichtung und Talentorientierung liefert dazu wertvolle Sportartempfehlungen. Dazu ist eine umfassende und komplex zu beurteilende **Validität der gesamten Kampagne eines Bewegungs-Checks** (Abb. 2) eine unabdingbare Voraussetzung. Dabei geht es nicht alleine im engeren Sinne um die Basisvoraussetzungen der inhaltlichen und kriterienbezogenen Validität der Testung an sich, sondern in einem umfassenderen Verständnis auch um die (mit der Validität eng verwandten) Nebengütekriterien Testökonomie, Zumutbarkeit, Nützlichkeit sowie schlussendlich der gesellschaftlichen Akzeptanz (Abb. 2). Die weit verstandene „Gesamtvalidität" der Kampagne resultiert aus der systemischen Bewertung aller genannten Validitätsfacetten und ist auf der politisch-administrativen Ebene entscheidend für die dauerhafte Etablierung eines Bewegungs-Checks.

Ein lokal, regional oder auch national flächendeckender Bewegungs-Check setzt als wichtigste Bedingung die **Akzeptanz** einer solchen Kampagne bei den Kindern selbst, aber mehr noch bei den Schulträgern, Lehrern, Eltern und Trainern voraus. Die gesellschaftliche Akzeptanz steht an der Spitze des umfassend verstandenen Validitätskonzepts und ergibt sich summarisch aus mehreren untergeordneten Validierungsschritten.

Die breite Akzeptanz von Bewegungs-Checks (mit den Maßnahmen Talentscreening und Talentorientierung) kann nur sichergestellt werden, wenn der Gesamtprozess sowohl vom ideellen Anliegen als auch von den konkret praktischen Einzelmaßnahmen her betrachtet bereits auf den ersten Blick einen deutlichen Mehrwert gegenüber der herkömmlichen Beratung und Betreuung der Kinder in den Bereichen Sport und Bewegung (im Sinne der *„Augenscheinvalidität"*; Moosbrugger & Kelava, 2012) bietet. Die alle Verantwortlichen und Teilnehmern überzeugende **Nützlichkeit** eines Bewegungs-Checks unter dem Dach der Grundschule sollte daher nicht nur im Hinblick auf die Talententwicklung evident werden, sondern sich grundsätzlich immer auch aus einem gleichwertigen Beitrag zur Gesundheitsförderung in der Region ergeben. Die meisten der in Deutschland durchgeführten Kampagnen sehen deshalb spezifische Förderangebote für jene Kinder vor, die aufgrund von körperlichen oder motorischen Defiziten keinen hinreichenden Fitnesszustand aufweisen und daher selbst breitensportlichen Angeboten (noch) reserviert gegenüberstehen.

Der Fuldaer Bewegungs-Check 2010-2016 (Hohmann, Hohmann, Scheuring & Zapp, 2015) besteht für alle Teilnehmer aus dem **Basis-Check**, der die allgemeinen Testverfahren des DMT 6-18 plus 1-3 Zusatztests sowie 2 Körperbaumaße umfasst, und einem etwa 8 Wochen später stattfindenden, sportartspezifischen **Talent-Check**, zu dem lediglich die motorische Besten eingeladen werden (Abb. 3). Diese ca. 15 % der Gesamtteilnehmer umfassende Gruppe der sportlich leistungsfähigsten Kinder (*best movers*) absolviert bei diesem Talent-Check gegenwärtig (Stand: 2016) die folgenden sechs sportartspezifischen Testverfahren: (1) 15-m-Schwimmen, (2) 10-m-Pedalofahren (Skisport), (3) Seitlicher Schwungwurf (Golf), (4) Fußballdribbling (mit einem Volleyball), (5) Tischtennisjonglieren, (6) Drop-Jump (Leichtathletik) sowie (7) schnelle Richtungssprünge (Kampf- und Spielsportarten). Im Anschluss an den Basis-Check wie auch den Talent-Check erhalten die Teilnehmer eine Urkunde mit einer aus den sportartspezifischen Testergebnissen abgeleiteten Sportartempfehlung in Form einer optimal passenden Erst-, Zweit- und Drittsportart. Gleichzeitig werden – bei Einverständnis der Eltern – die Teilnehmerdaten an die örtlichen Verbandsstützpunktleiter der empfohlenen Sportarten weitergegeben, damit diese die Talente und ihre Eltern direkt kontaktieren können. Als jährlicher Abschluss der Kampagne erhalten im Rahmen einer öffentlichen Siegerehrung im Hause des Landkreises Fulda die jeweils besten sechs Jungen und Mädchen einen Pokal und ein kleines Geschenk aus der Hand des Landrats.

Die Kampagne des Fuldaer Bewegungs-Checks 2010-2016 dient dem kommunalen Auftrag der *allgemeinen Daseinsvorsorge*. Aus diesem Grund wird im Anschluss an die erste Stufe des sogenannten Basis-Checks nicht nur den sportlich Hochbegabten am positiven Ende der Verteilungskurve der Leistungsfähigkeit (Abb. 3; Talententwicklung) eine zweite Teststufe angeboten, sondern auch der „Extremgruppe" der Kinder mit besonderen motorischen Defiziten ein weiterführendes und freiwilliges Förderangebot unterbreitet

Abbildung 2: Zur Struktur der Validität und damit verbundenen Nebengütekriterien bei der Bewertung einer Kampagne zur Talentsichtung und Talentorientierung

(Abb. 3; Gesundheitsförderung). Für diese 15 % der Kinder mit erkennbaren Bewegungsschwächen besteht die Fördermaßnahme aus einer wöchentlich stattfindenden **Bewegungsschulung** über den Zeitraum maximal eines halben Jahres. Da das Förderprogramm einerseits umfangreich und kostenfrei ist und andererseits von speziell ausgebildeten Fachkräften konzipiert und geleitet wird, erbringt die Teilnahme nicht nur für die Kinder selbst, sondern auch für die Eltern und Schulen einen erheblichen *finanziellen und gesundheitlichen Mehrwert*. Die Förderstunden finden einmal pro Woche in einer örtlichen Physiotherapiepraxis statt und beinhalten die Kernelemente Haltungsschulung, Koordinationsschulung und Bewegungsspiele sowie eine einmalige Ernährungsberatung für die Eltern der Kursteilnehmer.

In Anbetracht der begrenzten Gesamtteilnehmerzahl des gewählten Beispiels Fuldaer Bewegungs-Check 2010-2016 mit etwa 9.500 Zweitklässlern, sowie der bislang nur 6 Jahre Laufzeit der Talentsichtungskampagne und der im Landkreis Fulda vergleichsweise eingeschränkten spitzensportlichen Förderkapazitäten ist klar, dass die für die Validierung der Kampagne vorgeschlagenen Verfahrensschritte und Prüfmethoden erst in den kommenden Jahren sehr viel sicherer angewendet und besser interpretiert werden können. Da gegenwärtig nur für die 1. Stufe des Fuldaer Bewegungs-Checks (Basis-Check) genügend Daten für eine Längsschnittanalyse vorliegen, beschränkt sich der vorliegende Beitrag auf die Validierung der allgemeinen Testverfahren dieses **Basis-Checks**.

Die finanzielle und zeitliche **Ökonomie** der verschiedenen Bewegungs-Checks in Deutschland variiert unseres Wissens erheblich. Im vorliegenden Fall des Fuldaer Bewegungs-Checks handelt es sich um eine flächendeckende Vollerhebung auf der zweiten Klassenstufe der ca. 40 Grundschulen des Landkreises. Um die im Jahresdurchschnitt etwa 1.800 teilnehmenden Kinder zu testen, werden die einzelnen Grundschulklassen in die nächstliegende Mehrfachsporthalle eingeladen und dort von einem Testteam aus etwa 15 Fachkräften (meist Sportstudierende) betreut. Bei Nutzung der drei vormittäglichen Doppelstunden ergibt sich eine Gesamtdauer der Kampagne von vier Wochen, wobei die ausgewählten Sporthallen jeweils an 1-2 Vormittagen gesperrt bleiben. Die Teilnahme der Grundschulen geschieht im Klassenverband, wobei bis zu einer Höchstzahl von 50 Schülern auch zwei Klassen in einer Doppelstunde (90 Minuten) parallel getestet werden können. Die Anreise erfolgt entsprechend den örtlichen Gegebenheiten entweder zu Fuß oder wird von der Landkreisbehörde per Bus organisiert. In der Regel findet die letztere Variante meist in Form eines 2- bis 3-stündigen Ausflugs am Vormittag statt. Rechnet man zu den finanziellen Aufwendungen für die Testhelfer und Testmaterialien jene für den Bustransport und die Bewegungsförderung hinzu, entstehen Gesamtkosten in Höhe von etwa 15 Euro pro Teilnehmer. In Anbetracht des Umfangs der Kampagne kommt der Fuldaer Bewegungs-Check u. E. mit durchaus vertretbaren finanziellen und zeitlichen Ressourcen aus, was nicht zuletzt auch für die positive öffentliche Resonanz und den Zuspruch von Seiten der teilnehmenden Schulen mitverantwortlich ist. Hinzu kommt die auf ein Minimum beschränkte administrative Inanspruchnahme der Grundschulen, weil von deren Seiten lediglich der Vordruck mit der schriftlichen Einverständniserklärung der Eltern eingeholt und die Klassenlisten elektronisch (im Excel-Format) vorab übermittelt werden müssen. Alle weiteren Schritte der Testinstruktion und -durchführung, die EDV-basierte Dateneingabe durch die Tester an den einzelnen Teststationen sowie die Datenauswertung und Urkundenausgabe erfolgen direkt vor Ort durch das Testteam und erfordern keine operative Beteiligung der Lehrkräfte oder Begleitpersonen. Gerade die geringe organisatorische Belastung, speziell der Sport- und/oder Klassenlehrer, und die abschließende Übermittlung der klassenweisen Testergebnisse trägt entscheidend zur **Zumutbarkeit** der Teilnahme am Fuldaer Bewegungs-Check auf Seiten der Grundschulen bei.

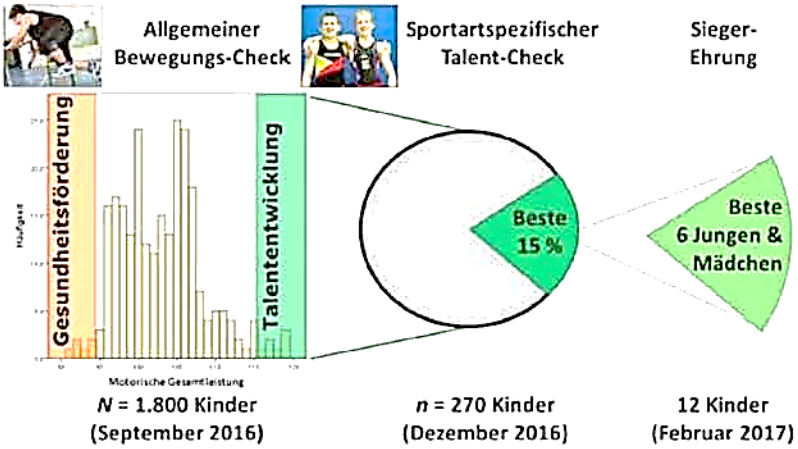

Abbildung 3: Jährliche Struktur und Verlauf des Fuldaer Bewegungs-Checks (Hohmann, Hohmann, Scheuring und Zapp, 2015; N steht jeweils für die jeweilige Grundgesamtheit und n für die daraus entnommenen Teilstichproben)

3. Zur kriterienbezogenen Validität der sportmotorischen Tests bei der Talentidentifikation

Die langfristige Leistungserwartung an ein sportliches Talent begründet den *prognostischen Charakter* der Talentsuche. Ausgehend vom inhaltlichen Messanspruch und der Zielgenauigkeit der eingesetzten Diagnoseverfahren (*inhaltliche Validität*) hängt die Qualität der Talentbestimmung und Sportartempfehlung somit nicht alleine von der sofortigen Aussagekraft des Testprofils (*konkurrente, d.h. unmittelbar zeitgleiche kriterienbezogene Validität*) ab, sondern mehr von der prognostischen, d.h. mittel- und langfristigen Aussagekraft (*prognostische kriterienbezogene Validität*) der Talentdiagnose, die sich im Hinblick auf die zukünftige Leistungsentwicklung ergibt.

Um die talentprognostische Validierung von Bewegungs-Checks im Grundschulalter zu erleichtern und auszudifferenzieren, sollte aufgrund des langen Prognosezeitraums bis hin zur finalen individuellen Bestleistung eine mindestens zweifach, besser jedoch dreifach gestufte **Etappenstruktur der Talentprognose** (Abb. 4 auf Seite 14) konzipiert werden. Mit Hilfe eines zeitlich gestuften Prozesses der Talentidentifikation kann auf der einen Seite die erforderliche Längsschnittuntersuchung in mehrere, parallel zu untersuchende gleitende Zeitfenster und Populationen aufgeteilt und damit der Untersuchungszeitraum erheblich verkürzt werden. Auf der anderen Seite hilft ein kombiniert quer-/längsschnittliches Studiendesign (*sliding populations approach*; Regnier, Salmela & Russell, 1993) auch, eine Reihe weiterer Untersuchungsprobleme

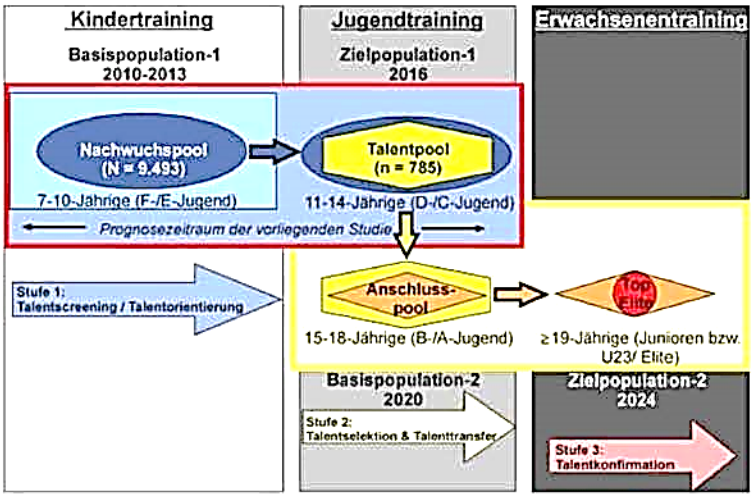

Abbildung 4: Dreifach gestufte Etappenstruktur der Talentprognose im (Nachwuchs)Leistungssport. Der bereits untersuchte Bereich der vorliegenden Studie ist rot umrahmt; die gelb umrahmte Untersuchung geplant.

zu verringern, wie z. B. *Stichprobenmortalität* (drop out), *Validitätsverschiebung* der Testverfahren oder unrealistisch *lange Finanzierungshorizonte der Forschungsförderung*.

Gleichzeitig erfordert die komplexe Leistungsstruktur der meisten Sportarten von Anfang an eine möglichst **vielseitige Merkmalserhebung** (*screening*), um ein hinreichend differenziertes Leistungsbild als Beurteilungsbasis zu schaffen.[2] Ferner sollte eine Konsequenz der natürlichen Merkmalskompensation zwischen den einzelnen Leistungsvoraussetzungen (z. B. Schnelligkeit und Ausdauer im Fußball oder Kraft und Koordination im Judo) darin bestehen, dass im Anschluss an die Datenerhebung auch auf der Auswertungsebene ein **mathematisch-statistischer Methodenverbund** aus herkömmlichen, *linearen Klassifikationsverfahren* und *nichtlinearen Mustererkennungsverfahren* zum Einsatz kommt. In der jüngeren Vergangenheit mehren sich die Belege, dass solche komplexen Auswertungsstrategien zu einer höheren Treffsicherheit bei der Identifikation von zukünftigen (Spitzen)Kadersportlern führen können (Hohmann, 2009; Zibung & Conzelmann, 2010; Hohmann, Fehr & Voigt, 2015; Pion, Hohmann, Liu, Vandorpe, Lenoir & Segers, 2016).

Speziell auf der ersten Etappe der Talentdiagnose besteht das Dilemma, dass zum Zeitpunkt des Trainingseinstiegs die sportartspezifischen Leistungsvoraussetzungen und hierunter insbesondere die sporttechnischen Voraussetzungen aufgrund der mangelnden Beherrschung der sportarttypischen Bewegungsfertigkeiten noch nicht überprüft werden können[3]. Daher stützt sich die Mehrzahl der in Deutschland bestehenden Kampagnen zur frühen Talentdiagnose beim sportmotorischen Screening der Nachwuchstalente auf **allgemeine Kontrollverfahren**, wie z. B. den *Deutschen Motorik-Test* 6-18 mit den acht Testverfahren (1) 20-m-Sprint, (2) Balancieren rückwärts, (3) Seitliches Hin- und Herspringen, (4) Rumpfbeuge, (5) Liegestütze, (6) Bauchaufzüge, (7) Standweitsprung und (8) 6-min-Lauf (Bös et al., 2009). Mit dem Ziel einer möglichst vielfältigen Merkmalserfassung wurde der DMT 6-18 im Fuldaer Bewegungs-Check 2010-2016 neben der Messung der Körperhöhe und des Körpergewichts von Anfang an (2010) durch einen (9) Schlagballweitwurf (80-Gramm-Ball) ergänzt[4].

Zur konkurrenten kriterienbezogenen Validität der Talentdiagnose

Um die beiden Modellebenen der kriterienbezogenen Validität für eine solche Testbatterie eines Bewegungs-Checks zu überprüfen, eignet sich die in Abbildung 5 im Überblick dargestellte **Validierungsstrategie** mit vier aufeinander aufbauenden Arbeitsschritten.Die vier Arbeitsschritte der kriterienbezogenen Validierung der Talentsichtung und Talentorientierung werden nachfolgend anhand einer früheren Studie von Hohmann, Fehr und Voigt (2015) sowie ausführlicher am Beispiel des Fuldaer Bewegungs-Checks erläutert.

Validität der Testauswahl

Zur Prüfung der konkurrenten kriterienbezogenen Validität werden die elf Talenttests zunächst *querschnittlich* bei älteren und aktuell bereits erfolgreichen Nachwuchs- oder Spitzensportlern eingesetzt. Auf der Basis der individuellen Muster der altersbereinigten Roh- oder normierten Testwerte[5] kann dann die Übereinstimmung der aus den Testprofilen abgeleiteten Sportlerklassifikation mit der tatsächlich betriebenen Sportart im Sinne ei-

[2] Mit diesem Ziel soll in naher Zukunft der Kanon der eingesetzten Körperbaumessungen und sportmotorischen Tests auch um sportpsychologische Diagnoseinventare und sportmedizinische Messgrößen erweitert werden.

[3] Das *Generalitäts-Spezifitäts-Dilemma* besagt, dass zur Messung bestimmter, in spezifischen Sportarten relevanten Merkmale auch spezifische Messinstrumente eingesetzt werden müssen, um eine aussagekräftige Diagnose zu erhalten. Mit der zunehmenden Spezifität des Messverfahrens geht jedoch einher, dass die damit erhaltenen Ergebnisse nur noch eingeschränkte Aussagen zur Eignung in anderen Sportarten gestatten. Je genereller ein Messverfahren konzipiert ist, desto geringer ist wiederum die sportartspezifische Aussagekraft. Das Dilemma besteht darin, dass ein bestimmtes Testergebnis, etwa ein fehlender Zusammenhang zum Kriterium, nicht eindeutig zu interpretieren ist. Einerseits könnte der Test nicht genügend valide sein und andererseits könnte aber auch tatsächlich kein Zusammenhang zwischen Prädiktor und Kriterium bestehen. Ein weitgehend akzeptierter Lösungsvorschlag besteht darin, in den frühen juvenilen Stadien eher eine allgemeine motorische Begabung zu testen und dann schrittweise auf sportartgruppen- und später auch sportartspezifische Verfahren umzustellen.

[4] Mittlerweile wurden zudem ein komplexer Wahlreaktionslauf (seit 2015) sowie ein Griffkrafttest (seit 2016) hinzugefügt

[5] Dazu wurden in einem ersten Arbeitsschritt die individuellen Testrohwerte mit Hilfe der Normwerttabellen des DMT 6-18 in *Prozentrangnormen* (PRN) überführt. Dabei wird auf regressionsanalytischer Basis das Lebensalter in Monaten berücksichtigt, um den Kalendarischen Altereffekt zu vermeiden (Lames et al., 2008). Für den Ballweitwurf, für den keine bundesweiten Normwerte existieren, wurden die Referenzwerte aus der Datenbasis des Fuldaer Bewegungs-Checks 2010-2014 (N > 6.000) berechnet. Als Referenzbasis für die Körperbaumaße Körperhöhe, Körpergewicht und BMI diente die für Deutschland repräsentative Datenbasis des Robert Koch-Instituts (Schienkiewitz, Neuhauser, Rosario, Dortschy, Kurth, Ellert & Stolzenberg, 2011).

[6] Um die Höhe der Gewichtungsfaktoren zu bestimmen, wurden sowohl *Expertenbefragungen* als auch eine umfassende *Literaturanalyse* durchgeführt.

ner **„unechten Quasi-Prognose"** geprüft werden. Die Treffsicherheit dieser ex-post-facto vorgenommenen Sportlerklassifikation informiert somit über die sportartbezogene Tauglichkeit der Testauswahl und die klassifikatorische Validität des gesamten Testprofils in Bezug auf die Güte der Sportartempfehlung.

Zu der vorgeschlagenen **Quasi-Prognose** haben Hohmann, Fehr und Voigt (2015) in dieser Zeitschrift über die Ergebnisse einer derartigen retrospektiven Analyse berichtet. In jener Studie wurden N = 56 bayerische D-Kaderathleten (oder vergleichbaren Leistungsniveaus) aus den sechs Sportarten Fußball, Leichtathletik-Sprint/Sprung, Ski-Alpin, Skilanglauf, Volleyball und Wasserball mit dem Deutschen Motoriktest plus Ballweitwurf getestet. Gemäß der Quasi-Prognose wurde aus den individuellen Testergebnissen der männlichen Nachwuchssportler im Alter von 12-16 Jahren (U13 bis U17) für jeden Kadersportler eine individuelle Empfehlung für eine Erst-, Zweit- und Drittsportart abgeleitet und mit seiner aktuellen Herkunftssportart verglichen.

Validität der sportartspezifischen Gewichtungsfaktoren

Um die kriterienbezogene Validität der Testbatterie weiter zu erhöhen, können in einem zweiten Arbeitsschritt die im Testprofil diagnostizierten Talentmerkmale *sportartspezifisch gewichtet* werden. Die Summe der für jede Sportart anforderungsspezifisch gewichteten Merkmalsausprägungen informiert dann über den Grad der sportartspezifischen Eignung des Testteilnehmers. Klassifiziert man anschließend die Sportler erneut, so erhält man über die kriterienbezogene Validität der Testwerte hinaus auch eine Information zur Validität der verwendeten **Gewichtungsfaktoren**[6]. Je besser die Gewichtungsfaktoren das jeweilige Anforderungsprofil einer Sportart abbilden, umso mehr sollte sich die Treffsicherheit der Sportlerklassifikation im Kontrast von ungewichteten und gewichteten Merkmalen erhöhen.

In der genannten Studie bei dem altersgemischten Kollektiv von Kaderathleten des Talent- und Anschlusspools konnten Hohmann, Fehr und Voigt (2015) die kriterienbezogene Validität des Testprofils bei der Empfehlung einer Zielsportart durch den Einsatz von sportartspezifischen **Gewichtungsfaktoren** deutlich steigern. So hatte sich bei den untersuchten Sportlern der sechs Sportarten gezeigt, dass durch den Einsatz von differenziert gewichteten Körperbaumerkmalen und motorischen Testleistungen die diskriminanzanalytische Klassifikationsgüte von 69,6 auf 76,8 % Treffer gegenüber dem ungewichteten Testprofil anstieg. Die Differenz zwischen der ungewichteten und gewichteten Prognosegüte in Höhe von d = 7,2 % bedeutet, dass neben dem individuellen Stärken-Schwächen-Profil der Kadersportler auch das konkrete Anforderungsprofil ihrer jeweiligen Herkunfts- bzw. der zu empfehlenden Zielsportart berücksichtigt werden muss. Dies geschieht am besten, indem die Einflusshöhe der in der jeweiligen Sportart führenden Leistungsvoraussetzungen bei der Talentsichtung durch eine entsprechende Gewichtung der überprüften Talentmerkmale abgebildet wird.

Zur prognostischen kriterienbezogenen Validität der Talentdiagnose

Bei Bewegungs-Checks muss aufgrund des grundsätzlich prognostischen Charakters jeder Form der Talentsichtung und Talentorientierung die prognostische Validität der Vorhersagen bestimmt werden. Bei der **„echten" Talentprognose** werden die jeweils am Ende des Kinder- und Jugendtrainings (s.o., Abb. 4) sowie am finalen Karrierehöhepunkt im Erwachsenenalter erzielten Bestleistungen der (Nachwuchs)Sportler abgewartet und dann die erreichten Zielleistungen retrospektiv mit den früheren individuellen Merkmalsprofilen und Sportartempfehlungen verglichen.[7]

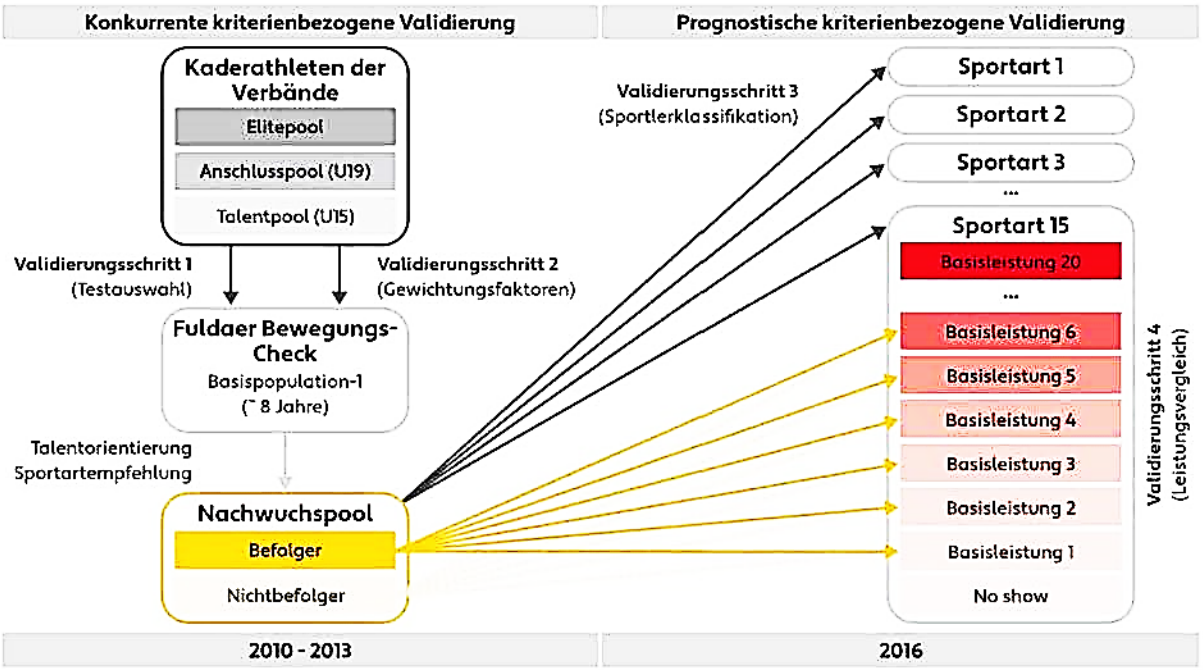

Abbildung 5: Vier aufeinander aufbauende Strategien/Studien zur konkurrenten und prognostischen Validierung der Talentsichtung und Talentorientierung im Rahmen eines Bewegungs-Checks

Die Teilnehmer des Fuldaer Bewegungs-Checks 2010-2016 bilden die 7-9-jährige **Basispopulation-1** (N_{Jungen} = 4.805; Alter: (94,93 ± 6,07) Monate; $N_{Mädchen}$ = 4.688; Alter: (92,43 ± 4,95) Monate. In dieser Gruppe wurde versucht, ab dem Teilnahmezeitpunkt am Bewegungs-Check möglichst jedes Wettkampfengagement in einer oder auch verschiedenen Sportarten mit Hilfe einer EDV-basierten, individuellen Namensrecherche im Internet und im örtlichen Pressezentrum (Intranet Fuldaer Zeitung) sowie im speziellen Fall Fußball anhand der verbandlich registrierten Spielerlizenzen zu ermitteln. Das Wettkampfengagement der erfolgreich recherchierten Nachwuchssportler der 11;0 bis 14;11-jährigen (Jahre; Monate) **Zielpopulation-1** (n_{Jungen} = 539; Alter: 149,44 ± 10,61 mon; $n_{Mädchen}$ = 246; Alter: 145,69 ± 9,64 mon) verteilt sich über insgesamt 36 verschiedene Sportarten, unter denen sich auch die 15 regional empfohlenen *Zielsportarten* mit Stützpunktausstattung befinden: Badminton, Basketball, Fußball (Empfehlungssportart - Codenummer 1) , Gerätturnen (#5), Golf, Handball, Judo (#7), Ju-Jutsu, Kanu-Classic (#24), Karate, Kickboxen, Kunstradfahren, Leichtathletik-Dauerlauf (#6), Leichtathletik-Sprint/Sprung (#2), Leichtathletik-Wurf, Minigolf, Motorsport, Radsport-Straße (#10), Radsport-Cross-Country (MTB), Reitsport, Schwimmen (#8), Ski-Alpin (#22), Skilanglauf (#23), Sportkegeln, Sportklettern, Sportschießen-Bogen, Sportschießen-Gewehr, Synchronschwimmen, Taekwondo, Tanz-Classic (#17), Tanz-Pop, Tennis, (#3) Tischtennis (#4), Volleyball (#12), Voltigieren und Wasserball (#15).

Validität der Sportartempfehlung (Sportlerklassifikation)

Nachfolgend soll für den Fuldaer Bewegungs-Check 2010-2016 die Güte einer **echten Talentprognose** prospektiv über den Zeitraum der in Abbildung 4 dargestellten ersten Prognoseetappe hinweg überprüft werden. Dazu wird zuerst die originale frühe Sportartempfehlung der Zweitklässler mit der 3 bis 6 Jahre später tatsächlich ausgeübten Zielsportart verglichen. Eine hohe prognostische Validität der Talentorientierung liegt dann vor, wenn die aktuell betriebene Zielsportart weitgehend mit der ursprünglichen individuellen Sportartempfehlung übereinstimmt. Darüber hinaus wird untersucht, welchen Einfluss die aus den führenden *Anforderungscharakteristiken* der einbezogenen Zielsportarten abgeleiteten Gewichtungsfaktoren beim Abgleich mit den *individuellen Stärken-Schwächen-Profilen* der Fuldaer Nachwuchssportler ausüben.

Ziel dieses **klassifikatorischen Validierungsschrittes** ist, herauszufinden, ob sich die im juvenilen Altersbereich der U12 bis zur U15 (11;0 bis 14;11 Jahre) ausgeübte Zielsportart treffsicher aus dem frühen Muster der Leistungsvoraussetzungen vorhersagen lässt. Die Hypothese besteht darin, dass diese Vorhersage dann optimal ausfällt, wenn die Teilnehmer am Fuldaer Bewegungs-Check bei der Auswahl ihrer Zielsportart ihre persönliche Eignung und Merkmalsausstattung berücksichtigt und eine der drei in der Urkunde empfohlenen Sportarten ab der Teilnahme am Bewegungs-Check wettkampfmäßig betrieben haben. Um diese Annahme zu überprüfen, wurde bei den bis zum Stichtag 31. Dezember 2016 im Wettkampfbetrieb Fuldaer Vereine angekommenen Nachwuchssportlern die später ausgeübte Zielsportart aus ihrem beim Bewegungs-Check ermittelten Stärken-Schwächen-Profil vorhergesagt. Dabei wurden wiederum jene 15 Zielsportarten analysiert, für die beim Bewegungs-Check Sportartempfehlungen formuliert worden waren. Als zweites Auswahlkriterium wurden bei dieser Analyse nur jene Nachwuchssportler einbezogen, deren Basisleistung auf der Erfolgsskala (gemäß Tab. 2) mindestens Level 3 erreichte, d. h. deren Wettkampferfolge ihre Zugehörigkeit zur sportartbezogenen Leistungsspitze des Landkreises Fulda belegen. Das Argument für diese Einschränkung besteht in der Annahme, dass die Sportartbetreiber mit nur mäßigem sportlichen Erfolg (Basisleistung 1 und 2) über die zurückliegenden 3 bis 6 Jahre hinweg vermutlich in weitaus geringerem Maße sportartspezifische Anpassungsleistungen erreicht haben. Somit wäre bei einer Einbeziehung auch dieser wenig leistungsfähigen Nachwuchssportler in das Klassifikationsverfahren zu erwarten, dass aufgrund ihrer geringer Leistungsentwicklung einerseits die sportartspezifische Eignung gering und andererseits die Heterogenität in den individuellen Stärken-Schwächen-Profilen so groß sein dürften, dass eine einheitliche Klassifizierung von vornherein aussichtslos wäre. Da die Mädchen deutlich in der Minderzahl sind und sich zudem in ihrem Stärken-Schwächen-Profil und auch in Bezug auf die Sportartwahl (z. B. Gerätturnen) erheblich von den Jungen unterscheiden, wird das Klassifikationsverfahren nur bei den männlichen Nachwuchssportlern angewendet. Insgesamt führen diese Einschränkungen zu einer Analysegruppe mit N = 190 Nachwuchssportlern, die sich aus den Sportarten Fußball (n = 54), Leichtathletik-Sprint/Sprung (n = 30), Tennis (n = 50), Tischtennis (n = 32) und Judo (n = 24) zusammensetzt.

Als Prognoseverfahren dienten die drei **Klassifikationsverfahren** lineare *Diskriminanzanalyse* (DA; mit Kreuzvalidierung durch individuelle Fallauslassung) und die beiden nichtlinearen Neuronalen Netzwerk-Varianten *Mehrschichtiges Perceptron* (Multilayer Perceptron, MLP; mit 10 % Holdout) und *Selbstorganisierende Kohonenkarte* (Self-organising Kohonen Feature Map, SOM; mit Kreuzvalidierung durch individuelle Fallauslassung, Leave-one-out[8]). Die *lineare Diskriminanzanalyse* und die Neuronale Netzwerk-Analyse (in der Variante des *mehrschichtigen Perceptrons* bei Verwendung der ungewichteten Profilmerkmale) bleiben deutlich hinter der Trefferquote zurück, die bei Berücksichtigung der sportartspezifischen Gewichtungsfaktoren erreicht wird (Tab. 1). So erhöht sich die Validität der Sportartprognose auf der Basis der **gewichteten Talentmerkmale** um d = 4,9 % von 33,6 auf 38,5 % Treffer. Im Methodenvergleich führt bei gewichteten Merkmalen das *mehrschichtige Perceptron* (MLP) mit insgesamt 44,7 % korrekten Zuordnungen der Sportler zu den jeweils von ihnen betriebenen Sportarten zum besten Klassifikationsergebnis.

Um die unterschiedliche Prognosequalität der drei Klassifikationsverfahren besser nachvollziehen zu können, wurden in einem zweiten Schritt die Klassifizierungsergebnisse auf der Basis der sportartspezifisch *gewichteten Profilmerkmale*

[7] Die im Abschnitt „Validität der Sportartempfehlung" beschriebene Variante der echten prognostischen Validitätsprüfung ist nicht immer und unter allen Umständen der Quasi-Prognose (s.o., Punkt 1) überlegen, denn die echte Prognose wird durch eine meist hohe Stichprobenmortalität (drop out) sowie eine ungesicherte langfristige Finanzierung stark beeinträchtigt. Auch drohen Validitätsverschiebungen bei den Testverfahren aufgrund von sportartspezifischen Regeländerungen mit Veränderung des Anforderungsprofils. Beispiele hierfür sind das Verbot auftriebsfördernder Schwimmbekleidung („Ganzkörperanzug") mit der nachfolgenden Bevorzugung eines leichteren Schwimmertyps oder die Anstoßregel („Schnelle Mitte") im Handball mit der anschließend erheblichen Steigerung der Schnelligkeitsanforderungen. Besonders schwer wiegt das mehr praktische Problem, dass die Strategie der echten Prognose das Problem des generell eher kurzfristigen Informationsbedarfs der Fachverbände nicht lösen kann und sie somit eher für die an theoretischen Erkenntnissen orientierte Grundlagenforschung taugt als für anwendungsorientierte und kurzfristige Problemlösungen bei der Talentsuche.

[8] Die mit erheblichem Arbeitsaufwand bei der Bedienung der Software (DataEngine; Fa. MIT, Aachen) verbundene Neuronale Netzwerk-Methode der Selbstorganisierenden Kohonenkarte wurde aus ökonomischen Gründen nur bei den gewichteten Profilmerkmalen eingesetzt.

Sportart	Diskriminanzanalyse		Mehrschichtiges Perceptron		Selbstorganisierende Kohonenkarte
	Ungewichtete Merkmale in %	Gewichtete Merkmale in %	Ungewichtete Merkmale in %	Gewichtete Merkmale: in %	Gewichtete Merkmale: in %
Fußball	40,5 (n = 42)	47,6 (n = 42)	27,5 (n = 52)	34,7 (n = 52)	15,4 (n = 52)
Leichtathletik-Sprint/ Sprung	31,3 (n = 32)	31,3 (n = 32)	44,0 (n = 32)	54,8 (n = 32)	34,4 (n = 32)
Tennis	15,4 (n = 26)	38,5 (n = 26)	7,7 (n = 26)	23,7 (n = 26)	0,0 (n = 26)
Tischtennis	35,3 (n = 34)	29,4 (n = 34)	30.6 (n = 34)	48,2 (n = 34)	39,7 (n = 34)
Judo	50,0 (n = 28)	50,0 (n = 28)	59,9 (n = 24)	49,6 (n = 24)	6,3 (n = 24)

Tabelle 1: Prognostische kriterienbezogene Validierung des DMT 6-18 plus Ballweitwurf sowie Körperhöhe und Körpergewicht anhand der Trefferzahl bei der Klassifikation der Bewegungs-Check-Teilnehmer in Bezug auf fünf Zielsportarten (162 < N < 168 männliche Kadersportler)

für jedes Individuum einzeln betrachtet (Abb. 6). Da einzelne Sportler in mehreren Sportarten wettkampfaktiv waren, wurden diese doppelten Untersuchungsfälle in der in Abbildung 6 grafisch veranschaulichten Analyse entsprechend mehrfach klassifiziert. Im Vergleich zu Tabelle 1 erhöht sich auf diese Weise die Fallzahl auf das Maximum von N = 190 Sportartfälle. Gleichzeitig reduziert sich aufgrund des identischen Stärken-Schwächen-Profils der „Doppelstarter", das nun der Zuordnung zu unterschiedlichen Sportarten dienen muss, naturgemäß auch die Trefferquote bei allen Methoden. Insgesamt wird erneut deutlich, dass das *mehrschichtige Perceptron* (MLP) knapp vor der *linearen Diskriminanzanalyse* (DA) mit jeweils etwa einem Drittel zutreffender Sportartprognosen die meisten korrekten Klassifikationstreffer erzielt.

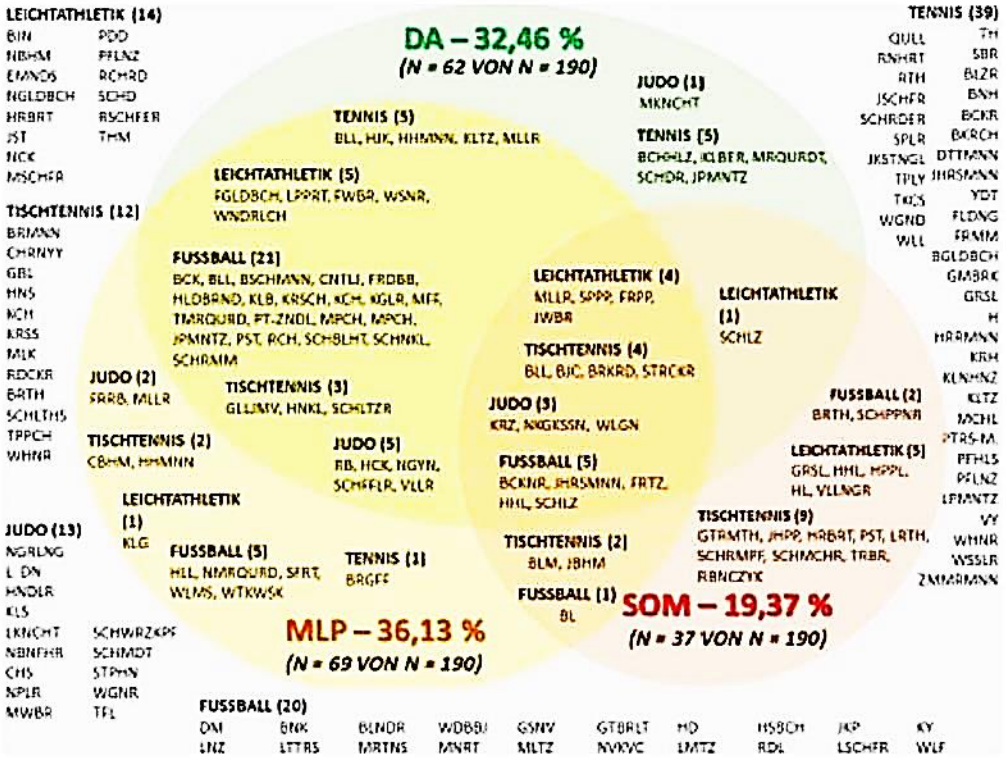

Abbildung 6: Individuelle Klassifikationsergebnisse mit Hilfe von Diskriminanzanalyse, Mehrschichtigem Perceptron und Selbstorganisierender Kohonenkarte bei N = 190 Nachwuchssportlern aus fünf einbezogenen Zielsportarten

Demgegenüber fällt der Anteil richtiger Vorhersagen bei der *selbstorganisierenden Kohonenkarte* (SOM) auf weniger als ein Fünftel ab.
Unabhängig von der durchaus unterschiedlichen Prognosequalität der drei Einzelverfahren ist zunächst auffällig, dass sich im Verbund der drei Klassifikationsmethoden (Abb. 6) bei insgesamt 92 der N = 190 Nachwuchssportler, d. h. bei 48,4 % mindestens eine der drei Klassifikationsmethoden eine korrekte Sportartvorhersage leistet. Unter diesen 92 Nachwuchssportlern wird für n = 43 die Zielsportart übereinstimmend von zwei der drei Methoden richtig vorhergesagt und bei n = 16 männlichen Nachwuchssportlern gelangen gleich alle drei Methoden zu identischen Sportartvorhersagen. Geht man bei der zufälligen (bzw. wahllosen) Zuordnung zu den fünf Sportarten von einer durchschnittlich 20-prozentigen (ein Fünftel) Zufallswahrscheinlichkeit aus, so kann eine korrekte Prognose der 3 bis 6 Jahre später betriebenen Zielsportart bei fast der Hälfte der Grundschüler durchaus als Erfolg interpretiert werden.
In Bezug auf das Klassifikationsergebnis in Abbildung 5 ist zu ergänzen, dass der Anteil der *Befolger*, d.h. jener Nachwuchssportler mit einer zur aktuell betriebenen Zielsportart passenden Sportartempfehlung, unter den 92 treffsicher identifizierten Mitgliedern der Zielsportarten (innerhalb der Ellipsen) höher ausfällt als in der Gruppe der 98 nicht korrekt klassifizierten Sportartbetreiber (siehe die Namenslisten an den Außenseiten). Mit 27,2 % befinden sich unter den anhand ihres frühen Stärken-Schwächen-Profils prognostisch valide identifizierten, späteren Mitgliedern der Zielsportarten deutlich mehr Befolger der ursprünglichen Talentorientierung als unter den nicht korrekt identifizierten Sportartvertretern, unter denen die falsch klassifizierten Befolger nur 23,5 % ausmachen.

Stufe	Erfolgsbasierte Leistungsskala
0	No show
1	Vereinsregistrierter Nachwuchssportler
2	Wettkampfteilnahme in der untersten lokalen altersbezogenen Leistungsklasse; lokal erfolgreiche Wettkampfteilnahme im Altersbereich bis zur U11
3	Wettkampfteilnahme in einer mittleren altersbezogenen Kreisklasse; lokal erfolgreiche Erfolg Wettkampfteilnahme im Altersbereich bis zur U13
4	Wettkampfteilnahme in der höchsten altersbezogenen Bezirks- oder Gauklasse; Finalteilnahme auf der Bezirksebene im Altersbereich bis zur U15
5	Wettkampfteilnahme in der höchsten altersbezogenen Landesklasse oder Finalteilnahme auf der Landesebene im Altersbereich bis zur U17
6	Teilnahme am Wettkampfbetrieb der höchsten altersbezogenen Landesgruppenklasse (z. B. Regionalliga) oder an der nationalen altersbezogenen Meisterschaft; Medaille auf Landesebene U17
7	Teilnahme am Wettkampfbetrieb der höchsten altersbezogenen nationalen Leistungsklasse (z. B. Jugend-Bundesliga) oder dem internationalen altersbezogenen Meisterschaftsbetrieb (z. B. Nominierung für U15-Nationalmannschaft); Gewinn der nationalen Meisterschaft im Altersbereich bis einschließlich U15 (Platz 1-3); Lehrgangsteilnahme Jugendnationalmannschaft (U17)
8	Teilnahme am Wettkampfbetrieb der höchsten altersbezogenen nationalen Leistungsklasse (z. B. Junioren-Bundesliga): Wettkampfteilnahme am internationalen Meisterschaftsbetrieb (z. B. Nominierung für U17-, U19- oder U21-Nationalmannschaft); Gewinn der nationalen Meisterschaft im Altersbereich ab U17 (Platz 1-3); Senioren: 2. Bundesliga
9	JEM-/JWM - Platz 9-16; Senioren: 1. Bundesliga / DM - Platz 13-20; Senioren: 1. Bundesliga / DM - Platz 9-12
10	Jugend-EM-/Jugend-WM - Platz 4-8; Senioren: 1. Bundesliga / DM - Platz 5-8
11	Medaille bei Jugend-EM/Jugend-WM; Senioren: 1. Bundesliga / DM – Platz 3-4
12	Senioren: 1. Bundesliga / DM – Platz 2
13	Medaille bei Junioren-EM/Junioren-WM; 1. Bundesliga / DM – Platz 1
14	B-Nationalkader / A-Länderkampf; Start bei EM / WM / OS
15	A-Nationalkader; EM - Platz 6-8 / WM/OS - Platz 9-16
16	EM - Platz 4-5 / WM/OS - Platz 5-8
17	EM - Platz 3 / WM/OS - Platz 4
18	EM - Platz 2 / WM/OS - Platz 3
19	Europameister / WM/OS - Platz 2
20	Weltmeister / Olympiasieger

Tabelle 2: Zielsportarten und 20- stufige erfolgsbasierte Leistungsskala zur Bewertung der Basisleistung, Anschlussleistung und Finalleistung der Teilnehmer am Fuldaer Bewegungs-Check 2010-2016

Validität der Sportartempfehlung (Leistungsvergleich)
Über die Treffsicherheit der Sportartempfehlung hinaus interessiert insbesondere deren Auswirkung auf die nachfolgende Leistungsentwicklung der Nachwuchssportler. Um die individuelle Leistung am Ende der verschiedenen Ausbildungsetappen bestimmen zu können, wurde für jeden Teilnehmer gemäß der lokalen, regionalen, nationalen oder auch internationalen Bedeutung der Wettkämpfe sowie der dort erzielten Leistungen und Erfolge das aktuelle Leistungsniveau anhand einer 20-stufigen Skala (Tab. 2) bewertet und als sogenannte „Basisleistung" registriert. Anhand des Wettkampfzeitpunkts und des Geburtsdatums wurde ferner das zugehörige „Basisalter" bei der Leistungserbringung bestimmt. Um den kalendarischen Alterseffekt (Lames, Augste, Dreckmann, Görsdorf & Schimanski, 2008) zu vermeiden, wurde das Testalter der Zweitklässler aus den Testwerten des Stärken-Schwächen-Profils ebenso regressionsstatistisch auspartialisiert wie später das Basisalter aus der Basisleistung im Altersbereich U12 bis U15. Sämtliche nachfolgend vorgestellten Untersuchungsergebnisse basieren somit bei Jungen und Mädchen auf altersbereinigten Leistungskennziffern.
Es liegt nahe, die **prognostische Validität** des allgemeinen Testprofils zunächst ganzheitlich zu betrachten und deshalb zuerst die Korrelation zwischen dem (ungewichtet) aufsummierten Profilwert des Bewegungs-Checks und der 3 bis 6 Jahre später erzielten Basisleistung zu prüfen.

Wie Tabelle 3 ausweist, besteht bei den Jungen in den Sportarten Fußball, Leichtathletik-Sprint/Sprung, Tennis, Judo und Schwimmen sowie bei den Mädchen im Fußball und im leichtathletischen Dauerlauf ein überzufälliger Zusammenhang ($p < 0.05$) zwischen der frühen sportmotorischen Gesamtleistung im Fuldaer Bewegungs-Check und der Höhe der 3 bis 6 Jahre später erreichten Leistungsstufe in der betriebenen Sportart. Da die sportartspezifische Gewichtung der Profilmerkmale lediglich vereinzelt zu einer geringfügigen Erhöhung dieses Zusammenhangs beiträgt, erscheint in den meisten Sportarten bereits das ungewichtete Testprofil prognostisch hinreichend aussagekräftig in Bezug auf das im Verlauf der ersten Prognoseetappe erreichte Leistungsniveau.

	Sportart		Ungewichtete Profilmerkmale		Sportartspezifisch gewichtete Profilmerkmale	
	Jungen	**n**	**r_{tc}**	**p**	**r_{tc}**	**p**
1	Fußball	266	.42	0.001	.42	0.001
2	Leichtathletik-Sprint/Sprung	30	.48	0.008	.50	0.005
3	Tennis	33	.52	0.001	.50	0.003
4	Tischtennis	41	.30	ns	.15	ns
6	Leichtathletik-Dauerlauf	15	.03	ns	.37	ns
7	Judo	23	.35	ns	.46	0.029
8	Schwimmen	12	.57	ns	.62	0.033
10	Radsport	4	.03	ns	-.06	ns
12	Volleyball	8	.50	ns	.36	ns
15	Wasserball	7	-.13	ns	.09	ns
22	Skilanglauf	4	.32	ns	.11	ns
	Mädchen	**n**	**r_{tc}**	**p**	**r_{tc}**	**p**
1	Fußball	12	.63	0.027	.59	0.042
2	Leichtathletik-Sprint/Sprung	40	.02	ns	.12	ns
3	Tennis	22	.31	ns	.34	ns
4	Tischtennis	26	.13	ns	.04	ns
5	Gerätturnen	27	.07	ns	.06	ns
6	Leichtathletik-Dauerlauf	8	.77	0.024	.50	ns
7	Judo	10	.63	ns	.43	ns
8	Schwimmen	5	.06	ns	.55	ns

Tabelle 3: Korrelationskoeffizienten der ungewichteten und sportartspezifisch gewichteten Summenscores aus dem sportmotorischen Merkmalsprofil des DMT 6-8 (plus Ballweitwurf) mit der 3-6 Jahre später erreichten (altersbereinigten und z-standardisierten) sportartspezifischen Basisleistung (gemäß Tabelle 1) in den 15 empfohlenen Zielsportarten (n > 3)

Ist die kriterienbezogene Validität des Talentprofils auf diese Weise gesichert, schließt sich unweigerlich die weitergehende Frage nach der **ökologischen Validität**, d.h. der *lebenswirklichen Effektivität* der Gesamtkampagne eines Bewegungs-Checks im *natürlichen Feldsetting* des Nachwuchs(leistungs)sports an. Hierbei wird geprüft, ob die „Befolger" der ursprünglichen Sportartempfehlung (also die zum Testzeitpunkt bereits existierenden Sportartbetreiber plus die spontan reagierenden Sportartwechsler) im Nachgang des Bewegungs-Checks höhere Leistungen in der empfohlenen Zielsportart erreichen als die „Nichtbefolger". Der retrospektive Vergleich der im Anschluss an den Bewegungs-Check eingetretenen Leistungsentwicklung zwischen den beiden Teilnehmergruppen soll als Maß für die ökologische Validität dienen, also über die Wirksamkeit der Kampagne unter den natürlichen Feldbedingungen des Nachwuchsleistungssports und somit zugleich über das Nebengütekriterium der *Nützlichkeit* eines Bewegungs-Checks informieren.

Die **ökologische Validität und Nützlichkeit** der Sportartempfehlungen bei der Talentorientierung wird anhand der am Ende des Prognosezeitraums erreichten und dann als altersbereinigte z-Werte aufbereiteten *Basisleistungen* beurteilt. Dazu wurden die von den Teilnehmern des Fuldaer Bewegungs-Checks im Altersbereich U12 bis U15 erzielten Wettkampfergebnisse registriert, gemäß der erfolgsbasierten Leistungsskala (Tab. 1) bewertet und dem entsprechend die Basisleistung fortlaufend aktualisiert. Die Hypothese besteht nun darin, dass die Gruppe der Befolger der Talentorientierung am Ende der ersten Ausbildungsetappe ihren Begabungsvorsprung konserviert haben und der Gruppe der Nichtbefolger leistungsbezogen überlegen ist. Der statistische Mittelwertvergleich (T-Test für unabhängige Stichproben bzw. U-Test) zeigte jedoch aufgrund der begrenzten Fallzahlen und geringen Mittelwertsdifferenzen in keiner Sportart signifikante Unterschiede. In Tabelle 4 (Seite 20) werden deshalb die Gruppenunterschiede mit Hilfe der **Effektgröße Cohens's d** (Cohen, 1988) beurteilt. Dabei werden Werte unterhalb von 0,2 als trivialer Effekt, zwischen 0,2 und 0,5 als kleiner, Werte zwischen 0,5 und 0,8 als mittlerer und Werte darüber als großer Effekt bewertet (Albrecht, Hanssen-Doose, Bös, Schlenker, Schmidt, Wagner, Will & Worth, 2016). Während für drei Sportarten keine Befolger ermittelt werden konnten, erreichen in acht der verbleibenden zwölf Zielsportarten die Befolger einer der drei Sportartempfehlungen höhere Basisleistungen, wobei das Ausmaß des Leistungsvorsprungs von trivial bis mittel reicht. In den restlichen vier Sportarten Tischtennis, Gerätturnen, Schwimmen und Wasserball zeigen hingegen nach 3 bis 6 Jahren die Nichtbefolger Leistungsvorsprünge. Die Effektstärken sind dabei trivial oder klein, wobei die Sportart Wasserball eine Ausnahme bildet und der Leistungsvorsprung der Nichtbefolger als groß einzustufen ist. Weit überwiegend erreichen die Effektgrößen der Leistungsdifferenzen zwischen Befolgern und Nichtbefolgern jedoch nur eine höchstens mittlere Höhe, was aufgrund der gegenwärtig geringen

Größe der Teilstichproben und der in diesem frühen Altersabschnitt noch nicht sehr ausgeprägten Unterschiede in der Basisleistung nicht überraschen kann.

4. Diskussion und Fazit

Die beispielhaft analysierte Kampagne des Fuldaer Bewegungs-Checks zur frühen Talentsichtung und Talentorientierung auf der Basis des DMT 6-18 plus Ballweitwurf darf zusammengefasst als praktikabel sowie zeitgleich prognostisch und aufgrund der Ökonomie, Zumutbarkeit und Nützlichkeit auch als ökologisch valide angesehen werden. Die in der Gesamttendenz **positive Gesamtvalidität** hat daher über die bisherige Laufzeit von insgesamt 6 Jahren hinweg zu einer hohen Akzeptanz dieses Verfahrens zur Talent- und Bewegungsförderung in der Bildungsregion Fulda geführt.

Die vorgestellten Untersuchungsergebnisse sind aufgrund der begrenzten Gesamtteilnehmerzahl von bislang nur etwa 9.500 Zweitklässlern, der lediglich 6 Jahre Laufzeit der Längsschnittanalyse sowie der im Landkreis Fulda in vielen Sportarten eher begrenzten spitzensportlichen Förderkapazitäten gegenwärtig in ihrer Aussagkraft noch stark limitiert. Neben der geringen Gesamtteilnehmerzahl tragen zum noch als vorläufig zu bezeichnenden Charakter der Auswertungsergebnisse noch folgende **Limitationen** bei:

- Mit rund 10 % (Mädchen) bzw. 20 % (Jungen) der Grundgesamtheit ist die *Stichprobengröße* der über die örtliche Presse erfolgreich recherchierbaren Sportartbetreiber noch relativ gering.
- Nach nur 3-6 Jahren Wartezeit bis zum gewählten Stichtag (31. Dezember 2016) besteht eine noch zu breite *Altersspanne* der analysierten Zielpopulation, die von 11;0 Jahre (U12) bis 14;11 Jahre (U15) reicht und erst zu einem Stichtag am 31. Dezember 2018 die sinnvolle Einengung auf den talentbezogen aussagekräftigeren sowie alters- und ausbildungsbezogen homogeneren Altersbereich U14/U15 gestattet.
- Nach maximal 3 bis 6 Jahren Training ist die *Leistungsentwicklung* vieler talentierter Nachwuchssportler in den meisten Sportarten noch nicht hinreichend ausgebildet und reicht speziell in den frühen Jahren des untersuchten Altersbereichs U12 bis U15 nur in wenigen Einzelfällen über das lokale und regionale Niveau hinaus.
- Die Vergleichbarkeit von *Kaderstatus* und *Kaderniveau* in den verschiedenen Sportarten ist gerade auf den unteren Leistungsstufen sehr eingeschränkt. Während das sportmotorische Niveau der Nachwuchssportler im Fußball, Gerätturnen, Judo, Schwimmen, Tennis und Tischtennis aufgrund des bereits frühzeitig leistungsorientierten Trainings und der hohen Wettkampfqualität bereits im Grundschulalter vergleichsweise hoch ist, bleibt das sportmotorische Leistungsniveau in anderen Sportarten (Kanu, Leichtathletik, Radsport, Ski-Alpin und Skilanglauf, Tanz-Classic, Volleyball, Wasserball) aufgrund eines betont grundlagenorientierten Trainings oder eines auf dieser Altersstufe noch deutlich reduzierten Trainingsvolumens deutlich zurück. Dabei ist auch zu berücksichtigen, dass sich die Anforderungsprofile in bestimmten Sportarten (Leichtathletik-Sprint, -Sprung, -Wurf, Mittel- und Langstreckenlauf sowie -Mehrkampf) alters- und entwicklungsbedingt erst in späteren Altersabschnitten als der U12 und U13 ausdifferenzieren.
- Schlussendlich kommt hinzu, dass die Anforderungen in bestimmten Disziplinen besonders auf den frühen Ausbildungsstufen eng verwandt sein oder übereinstimmen können. So hat das Anforderungsprofil des Skilanglaufs einerseits die Gleichgewichtsfähigkeit mit dem Alpinen Skilauf gemeinsam und im Volleyball und Wasserball wird gleichermaßen eine überdurchschnittliche Körperhöhe und hohe Wurf- bzw. Schlagkraft gefordert. Ebenso ist plausibel, dass es in den Sportspielen Volleyball und Wasserball deshalb zu Einbußen bei der Sportlerklassifikation kommen kann, weil beide Sportarten vergleichsweise spät in ein systematisches Training einsteigen und aufgrund der bis zum Erhebungszeitpunkt nur kurzen Dauer und des geringen Umfangs des vorangegangenen Trainings die bis dahin hervorgerufenen Adaptationswirkungen sich im individuellen Stärken-Schwächen-Profil kaum abbilden.

Sportart	Befolger Leistungspunkte MWz-Wert+ SD Fallzahl (n)	Nichtbefolger Leistungspunkte MWz-Wert + SD Fallzahl (n)	Effektgroße Cohen´s d (Bewertung nach Hopkins, 2002)
Fußball	0,039 ± 0,80 (32)	-0,015 ± 0,87 (246)	+ 0,06 (Trivial)
LA – Sprint/Sprung	-0,081 ± 0,51 (5)	-0,172 ± 0,66 (65)	+ 0,14 (Trivial)
Tennis	0,468 ± 0,95 (17)	0,228 ± 0,99 (38)	+ 0,25 (Klein)
Tischtennis	-0,285 ± 0,23 (9)	0,008 ± 0,98 (57)	- 0,31 (Klein)
Gerätturnen	0,298 ± 0,93 (16)	0,483 ± 0,99 (12)	- 0,20 (Klein)
LA - Dauerlauf	-0,979 ± 0,36 (6)	-1,017 ± 0,72 (15)	+ 0,06 (Trivial)
Judo	0,869 ± 1,38 (14)	0,421 ± 1,45 (19)	+ 0,32 (Klein)
Schwimmen	-0,032 ± 1,54 (4)	0,080 ± 1,24 (13)	- 0,03 (Trivial)
Radsport	-0,714 ± 1,43 (7)	-1,318 ± 0,08 (3)	+ 0,54 (Mittel)
Volleyball	0,550 ± 0,49 (3)	0,263 ± + 0,60 (6)	+ 0,52 (Mittel)
Wasserball	-0,485 ± 1,69 (2)	0,528 ± 0,97 (5)	- 1,01 (Groß)
Tanz - Classic	0,64 ± 0.60 (2)	-0,098 ± 0,53 (2)	+ 0,36 (Klein)
Ski-Langlauf	---	0,764 ± 0,47 (6)	---
Ski-Alpin	---	0,213 ± 0,62 (3)	---
Kanu - Classic	---	1,815 ± 1,67 (2)	---

Tabelle 4: Vergleich der 3-6 Jahre später von den Befolgern und Nicht-Befolgern unter den männlichen Teilnehmern am Fuldaer Bewegungs-Check 2010-2016 erreichten (altersbereinigten und z-transformierten) Basisleistungen (gemäß Tabelle 1)

Insgesamt münden diese Einschränkungen gerade bei Talenten mit einem intraindividuell homogenen Stärken-Schwächen-Profil bei der frühen Talentorientierung auf der Basis lediglich allgemeiner Testverfahren in ein erhöhtes Fehlerrisiko. Gerade in Sportarten mit hochspezifischen Anforderungen, wie z. B. Tischtennis (Auge-Hand-Koordination) oder Schwimmen und Wasserball (Wassergefühl), könnte die Treffsicherheit der Talentprognose unter Umständen durch die Einbeziehung alternativer allgemeiner oder auch zusätzlicher sportartspezifischer Tests möglicherweise erhöht werden.

Neben diesen Limitationen bestehen die **Stärken** des Fuldaer Bewegungs-Checks darin, dass die Kampagne

• als **Totalerhebung** bei der Grundgesamtheit aller Zweitklässler in der Testregion konzipiert ist und auf diese Weise eine maximal breite Abdeckung und ausgleichende „Mittelung" der vielfältigen und unterschiedlichen *Feldbedingungen* (Sportvereins- und Schulsportsettings) in der Bildungs- und Sportregion gewährleistet.
• durch die Ergänzung des Deutschen Motoriktest DMT 6-18 um die Körperbaumaße Körperhöhe und Körpergewicht sowie insbesondere die *Zusatztests* Ballweitwurf (seit 2010), komplexer Wahlreaktionslauf (seit 2015) und Griffkraft (seit 2016) bereits im Rahmen der ersten Stufe des Basis-Checks zu einem motorisch sehr breit angelegten **Talentscreening** und damit einer im Hinblick auf die Vielfalt der Sportarten deutlich differenzierteren **Talentorientierung** führt. In diesem Zusammenhang ist zu betonen, dass aufgrund unterschiedlicher individueller Sportartpräferenzen und verschiedenartiger örtlicher Angebotsstrukturen den Teilnehmern am Fuldaer Bewegungs-Check grundsätzlich nicht nur eine, sondern mindestens 2-3 geeignete Sportarten empfohlen werden.
• für die Gruppe der motorisch Besten eine zweite Stufe *Talent-Check* vorgesehen ist, die aus sportart- oder zumindest semispezifischen Tests für bestimmte Sportarten mit regional vorhandener Förderinfrastruktur besteht und die zur Präzisierung des ersten Empfehlungsdurchgangs beiträgt. Auch wenn die zweite Stufe des Fuldaer Bewegungs-Checks im vorliegenden Beitrag noch nicht Gegenstand einer profunden empirisch-statistischen Analyse sein konnte, dürfte in der Gesamtschau des **zweistufigen Gesamtkonzepts** die Treffsicherheit der Sportartempfehlung auch für jene Sportarten deutlich steigen, die beim hier untersuchten Basis-Check noch Schwierigkeiten bereitet haben. Darüber hinaus bietet zukünftig die Einbeziehung von weiteren oder alternativen allgemeinen und sportartspezifischen Zusatztests sowie die laufende Optimierung der sportartspezifischen *Gewichtungsfaktoren* durchaus noch Potenzial, um die Qualität der Talentorientierung weiter zu verbessern.

Abschließend ist es für die Akzeptanz und Wirksamkeit von Bewegungs-Checks wichtig, dass durch eine geeignete kampagnenbegleitende **Informationspolitik** bei den Eltern und Lehrern dafür geworben wird, dass Nachwuchssportler aus Sportarten, in denen sie nicht länger verbleiben möchten (Talenttransfer), und insbesondere solche Zweitklässler, die noch keine Sportart aufgenommen haben, sich gezielt an den Sportartempfehlungen orientieren mögen. Die überwiegend positiven Befunde in Sportarten mit einem vergleichsweise höheren Befolgeranteil deuten darauf hin, dass sich durch eine vermehrte Nutzung der Sportartempfehlungen die ökologische Validität der Talentorientierung sicherlich weiter steigern ließe. Gleichzeitig zeigt die ansatzweise erkennbare Überlegenheit der Befolger der
Talentorientierung, dass die festgestellten sportartbezogenen Begabungsunterschiede zwischen den Grundschulkindern zumindest über die frühe Etappe des Kindertrainings fortbestehen. So bleiben die Nichtbefolger trotz eines identischen Trainingsinput und der allgemein guten Adaptationsvoraussetzungen in diesem Altersbereich hinter den begabungsgerecht trainierenden Nachwuchssportlern zurück. Über die Zumutbarkeit der Testdurchführung hinaus spricht somit auch die Nützlichkeit der Sportartempfehlungen für die Durchführung von Bewegungs-Checks in der Grundschule.

Literatur

Albrecht, C., Hanssen-Doose, A., Bös, K., Schlenker, L., Schmidt, S., Wagner, M., Will, N. & Worth, A. (2016). Motorische Leistungsfähigkeit von Kindern und Jugendlichen in Deutschland. Eine 6-Jahres-Kohortenstudie im Rahmen des Motorik-Moduls (MoMo). *Sportwissenschaft, 46* (4), 294-304.

Bös, K. & Schlenker, L. (2016). *Der Deutsche Motorik-Test 6-18* (2. Auflage). Hamburg: Feldhaus.

Brezina, H. (2016). *Start-Up "Sport-Analytik" - Find the right sport for your child.* (Vortrag im "Internationalen Experten-Talk zum Einsatz von sportartübergreifenden sportmotorischen Testverfahren" am 14. Juni 2016 im Haus des DOSB; Frankfurt/Main)

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Deutscher Olympischer Sport-Bund (DOSB) (2013). *Nachwuchsleistungssportkonzept 2020.* Frankfurt am Main: Deutscher Olympischer Sport-Bund (DOSB).

Fuchslocher, J., Romann, M., Rüdisüli, L.R., Birrer, D & Hollenstein, C. (2011). Das Talentselektionsinstrument PISTE: wie die Schweiz Nachwuchsathleten auswählt. *Leistungssport, 41* (4), 22-27.

Güllich, A. (2013). Talente im Sport. In A. Güllich & M. Krüger (Hrsg.), *Sport. Das Lehrbuch für das Sportstudium* (S. 623-653). Berlin, Heidelberg: Springer.

Herrmann, S., Bartz, E. Lischka, B. & Spahl, O. (2016). Bewegungs-Checks an Deutschlands Grundschulen. Eine Bestandsaufnahme. *Leistungssport, 46* (3), 52-55.

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).

Hoffmann, A. & Pfützner, A. (2013/2014). Wege an die Spitze. Herausforderungen und Schwerpunkte im deutschen Nachwuchsleistungssport. *Zeitschrift für Angewandte Trainingswissenschaft, 20* (2) und *21* (1), 1-283.

Hohmann, A. (2009). *Entwicklung sportlicher Talente an Sportbetonten Schulen.* Petersberg: Imhof.

Hohmann, A., Fehr U. & Voigt, L. (2015). Heute im Talentpool - In Hamburg auf dem Podium. *Leistungssport, 45* (5), 5-11.

Hohmann, A., Hohmann, L., Scheuring, L. und Zapp, M. (2015). *Abschlussbericht Fuldaer Bewegungs-Check 2010-2015.* Bayreuth: Institut für Sportwissenschaft der Universität Bayreuth (http://www.sport.uni-bayreuth.de/spo_wiss_I/de/research/b_projects/ Bewegungscheck_Fulda/index.html; Zugriff am 17. Januar 2017).

Hopkins, (2002). *A New View of Statistics: A Scale of Magnitudes for Effect Statistics.* (http://www.sportsci.org/resource/stats; Zugriff am 17. Januar 2017).

Issurin, V. & Lustig, G. (2008). Zur Identifikation talentierter Sportler: allgemeine Ansätze und praktische Einsichten. *Leistungssport, 38* (4), 4-7.

Joch, W. (1992). *Das sportliche Talent.* Aachen: Meyer & Meyer.

Lames, M., Augste, C., Dreckmann, C., Görsdorf, K. & Schimanski, M. (2008). Der „Relative Age Effect" (RAE): neue Hausaufgaben für den Sport. *Leistungssport, 38* (6), 4-9.

Marconi, M. (2013). *Die Identifikation von Talenten im Sport aus entwicklungstheoretischer Perspektive* (unv. Dissertation an der Universität Bern). Bern: Universität Bern.

Moosbrugger, H. & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion.* Berlin, Heidelberg: Springer.

Pion, J., Lenoir, M. & Segers, V. (2013). *Talent prognosis in different sports.* Ghent: Department of Sport Science.

Pion, J., Hohmann, A., Liu, T., Vandorpe, B., Lenoir, M. & Segers, V. (2016). Predictive models reduce talent development costs in female gymnastics. *Journal of Sports Sciences,* [0264-0414] 2016, 1 -6.

Regnier, G., Salmela, J. & Russell, S.J. (1993). Talent detection and development in sport. In R.N. Singer, M. Murphy & L.K. Tennant (eds.), *Handbook of Research in Sport Psychology* (pp. 290-313). New York: Wiley.

Schienkiewitz, A., Neuhauser, H., Rosario, A.S., Dortschy, R., Kurth, B.-M., Ellert, U. & Stolzenberg, H. (2011). *Beiträge zur Gesundheitsberichterstattung des Bundes - Referenzperzentile für anthropometrische Maßzahlen und Blutdruck aus der Studie zur Gesundheit von Kinder und Jugendlichen in Deutschland (KiGGS) 2003-2006:* Robert Koch-Institut.

Zibung, M. & Conzelmann, A. (2012). The role of specialisation in the promotion of young football talents: A person-oriented study. *European Journal of Sport Science, (12),* 1-9 (dx.doi.org/10.1080/17461391.2012.749947).

Zinner, J., Werner, C., Mehler, S., Abraldes Rois, L.M. (2015). Berlin hat Talent" Ergebnisse der Untersuchungen in mehreren Bezirken von Berlin 2014. *Zeitschrift für Gesundheit und Sport,* 5 (1), 59-66.

Korrespondenzadresse

Prof. Dr. Andreas Hohmann, Universität Bayreuth, Institut für Sportwissenschaft, Lehrstuhl für Trainings- und Bewegungswissenschaft, 95440 Bayreuth
E-Mail: andreas.hohmann@uni-bayreuth.de

Artikel 2

Hohmann, A., Siener*, M., & He, R. (2018). Prognostic validity of talent orientation in soccer. *German Journal of Exercise and Sport Research*, *48*(4), 478–488. https://doi.org/10.1007/s12662-018-0549-5

Main article

Ger J Exerc Sport Res 2018 · 48:478–488
https://doi.org/10.1007/s12662-018-0549-5
Received: 30 January 2018
Accepted: 3 October 2018
Published online: 14 November 2018
© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature 2018

Andreas Hohmann · Maximilian Siener · Renye He
Institute of Sports Science, University of Bayreuth, Bayreuth, Germany

Prognostic validity of talent orientation in soccer

Introduction

In long-term talent development programs, early talent orientation includes diagnostic talent screening measures as well as sports recommendation; these are the first steps of talent identification. According to Pion (2015, 2017), talent orientation is related to early talent detection in still-heterogeneous samples and "aims at motivating youngsters to choose a sport that matches the individual talent characteristics" (Pion, 2015, p. 22). Some sports require early talent orientation, given the very young age of athletes at the highest level (Papic, Rogulj, & Plestina, 2009). Following this idea, several talent screening programs in elementary schools have implemented motor diagnostics (e.g., Stemper, Bachmann, Diehlmann, & Kemper, 2009; Pion, 2015; Fuchslocher, Romann, Rüdisüli, Birrer, & Hollenstein, 2011; Golle, Muehlbauer, Wick, & Granacher, 2015) to orientate children into groups of sports; these could be either game sports, such as soccer (Pion, 2015) or specific sports, such as alpine skiing (Mueller, Mueller, Kornexl, & Raschner, 2015). The more accurately recommended sport that best suits the individual's talent make-up, the greater the chances become that a beginner will achieve success and satisfaction within his or her chosen sport. Furthermore, early talent orientation leads to a higher overall quality of the talent pool preparing for the second step: talent selection (Hohmann & Seidel, 2003). On one hand, early talent screening and sports orientation could promote fairness in the uneven race between competing sport federations for the diminishing numbers of talented sport beginners; the most professional sports, like soccer and tennis, attract the majority of children, while less popular sports receive candidates from the pool of remaining beginners. On the other hand, those children selecting popular sports—although their motor competence profile suits another discipline much more closely—might sacrifice important learning time necessary for developing the specific technical skills of their best-fitting sport. This not only limits their chances of reaching their individual performance potential, but also hinders children's opportunity to "exploit their talents to the fullest" (Pion, 2017, p. 5).

In general, there is a lack of research that investigates the prognostic value of different performance prerequisites over the entire long-term period from child to adult training. Most research is concentrated on the middle stage of juvenile performance development, but even at the advanced stage of late adolescence some studies question juvenile success as an appropriate indicator for soccer success into adulthood (Guellich, 2014). With regard to the early part of this training stage between early and late adolescence, only a few studies (Höner & Votteler, 2016; Carling, Le Gall, & Malina, 2012; Le Gall, Carling, Williams, & Reilly, 2010) comprised a prognostic period of more than three years. Doing so, Höner and Votteler (2016; see also Höner et al., 2015, and Höner, Leyhr, & Kelava, 2017) reveal that, even on the homogeneous level of the German soccer competence centres, a soccer-specific test battery (Deutscher Fußball-Bund, 2009) provides prognostic, valid, and also practically worthwhile information about the 7.2 times greater odds of the fastest and technically best third of preselected players to reach the junior national team (U15).

In soccer, early talent orientation is of great relevance, as elite soccer demands high technical proficiency, and therefore requires an early start for systematic long-term athletic development (Hohmann & Pietzonka, 2017; Papic et al., 2009). In addition to that, early talent orientation along with soccer education beginning at the elementary school level give coaches a longer observation period, which thus reduces selection errors during early adolescence.

Although reliable and valid information about the future potential of talented players on the basis of motor abilities and technical skills diagnostics is a valuable tool in talent development programs for clubs and sport federations, several studies question the long-term predictability of future success (Lidor, Cote, & Hackfort, 2009; Pankhurst & Collins, 2013; Carling & Collins, 2014). The main reasons for scientific concerns are evident in the often-undifferentiated mixture of general as well as sport-specific tests in talent identification campaigns; these concerns are also apparent from the unsystematic timing of cross-sectional diagnostics at single points in time during the long-term athletic development process. It henceforth comes as no surprise that the great variety of study design parameters have led to inconsistent research results, providing an inconsistent picture with regard to the prognostic validity of motor tests as they address general motor abilities and sport-specific technical skills. Some studies verified the prognostic validity of motor tests (Höner & Votteler, 2016; Figueiredo, Gonçalves, Coelho, Silva, & Malina, 2009; Zuber, Zibung, & Conzelmann, 2016), whereas

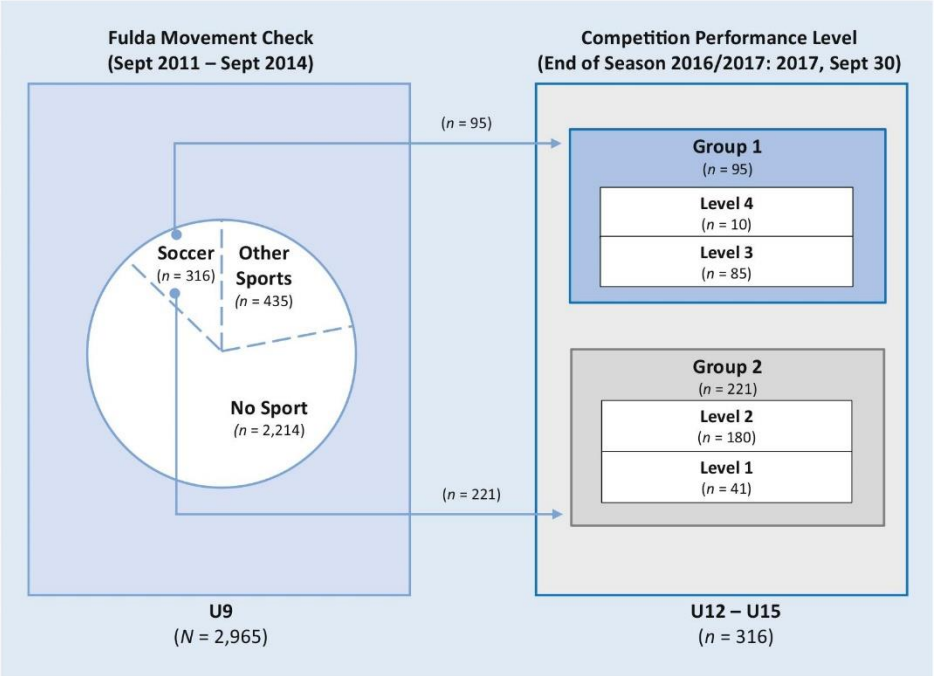

Fig. 1 ◄ Scheme of the different participants' groups of the Fulda Movement Check (2011–2014) included in this study (U9= under the age of 9 years)

others did not find significant associations between test results and later success at youth, junior, or adult soccer levels (Carling et al., 2012; Le Gall et al., 2010). Recent talent research in soccer by Höner and Votteler (2016) and Höner et al. (2017) has highlighted that a longitudinal investigation of motor predictors' prognostic relevance for long-term success is not only a key topic in talent research (Höner & Votteler, 2016; Gonaus & Mueller, 2012), but also an indispensable prerequisite for a sophisticated understanding of the changing prognostic validity of the various diagnostics at different stages of elite soccer players' long-term development.

The aim of this prospective cohort study was to evaluate the mid-term prognostic validity of general motor tests as well as that of a soccer-specific score calculated from a test battery. Therefore, the true players' success in soccer is compared with the tested performance prerequisites of young athletes as well as with the given recommendations. It was expected that children with better test results und a higher sport-specific test score in soccer would reach higher performance levels after a period of 2–6 six years. In addition, it raises the question of whether soccer players stand out in certain motor tests compared to athletes participating in other sports at the elementary school level.

Methods

General study design

According to the model of long-term talent development (see Hohmann, Fehr, & Voigt, 2015), the relevance of the predictors from the motor competence profile must be investigated in a stepwise fashion for at least three prognostic periods from the beginning of a talent development program until the full reach at the professional level. As talent development in soccer mostly starts in the U9 age group (U9= under the age of 9 years) and the elite adult level is reached at age 21, the long-term perspective should cover at least 12 years. This perspective also corresponds to the average time span of 15 years that is necessary for a player to potentially reach soccer excellence (Leite, Baker, & Sampaio, 2009).

In this study, a prognosis was calculated from the beginning of soccer training at the age of 8 years (generally second grade at the elementary level) over a middle-term time span of 2–6 years. Furthermore, this study reveals statistics regarding general anthropometric, physical, and physiological abilities. Therefore, the predictive validity of the anthropometric variables was investigated, which include body height and weight, the eight general tests of the German Motor Test (GMT) 6–18 (Boes & Schlenker, 2016), and an additional ball throw (ball weight: 80 g).

These diagnostics were administered in the FMC campaign. The FMC is a talent and health promotion campaign for children in the Fulda district (with a total of about 250,000 inhabitants), which was implemented in the year 2010 and comprises the second grade classes of all 40 elementary schools belonging to the Fulda district (Hohmann et al., 2015; Hohmann, Hohmann, Scheuring, & Zapp, 2016; see also Hohmann, Fehr, Siener, & Hochstein, 2017a; Hohmann, Fehr, Siener, & Hochstein, 2017b). For the ten most successful sport disciplines of the Fulda region (swimming, track and field, cross-country skiing, soccer, handball, water polo, tennis, table ten-

nis, judo, canoeing), a sport-specific recommendation score (SSRS) was calculated (Hohmann et al., 2015). At the end of the FMC, each participant received a diploma displaying his or her SSRS, which revealed and explained the sport disciplines that best suit each participant's individual talent make-up.

Participants

All individuals participated in motor diagnostics at the age of 7.89 ± 0.47 years during the particular time points of the respective FMC in the years 2011–2014. Regarding the current prospective cohort study, we investigated the data from $N = 2935$ male second grade school children with no more than one sample missing in the test battery. All study members participated in the FMC between the years 2011 to 2014 and were found to engage later on in soccer (soccer players: $n = 316$), other sports (other athletes: $n = 435$) or no sport at all (non-athletes: $n = 2214$) by the end of the sport season 2016/17 (September 30, 2017; ◘ **Fig. 1**).

University staff members and students conducted the motor tests. The performance development of the children in regard to the four cohorts (2011–2014) were recorded from the beginning of the season 2011/12 (October 1, 2011) through the end of the season 2016/17 (September 30, 2017) from the regional print, online media, (tournament and soccer match lists; www.torgranate.de) and the official DFB website (www.fussball.de; soccer team lists).

The U9 motor test results served as predictors for participants' later success in soccer at the early adolescent level with the U12 to U15 age group, which is between the minimum age of 11 years (U12) and the maximum age of 14 years and 11 months (U15). Due to the fact that the soccer player sample consisted of four successive cohorts, the prognostic period varied between 32 and 78 months with an average of $M = 55.4 \pm 9.4$ months.

Before entering the FMC, all children's parents provided written, informed consent for the recording and scientific use of the data collected for both the anthropometric and motor tests. The university's

Abstract · Zusammenfassung

Ger J Exerc Sport Res 2018 · 48:478–488 https://doi.org/10.1007/s12662-018-0549-5
© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature 2018

A. Hohmann · M. Siener · R. He

Prognostic validity of talent orientation in soccer

Abstract

Several talent-screening programs conducted within elementary schools have implemented motor diagnostics to introduce children to groups of sports including both game and specific sports, the latter of which includes gymnastics, skiing, and tennis. However, as in most other sports, the predictive value for soccer regarding such early testing is unclear. This prospective cohort study evaluated the mid-term prognostic validity of general motor tests as well as a soccer-specific score calculated from a test battery. The sample consisted of $N = 2965$ U9 age group of children from the Fulda Movement Check (FMC). The FMC is a basic check comprised of two anthropometric parameters and eight general items of the German Motor Test 6–18 by Boes and Schlenker with the addition of a ball throw. The test data were collected from the second grade classes of 2011–2014. The soccer competition performance of those children who chose soccer ($n = 316$) up to the end of the season 2016/17 (September 30, 2017) was recorded. This group of U12–U15 players was then assigned individually to four different competition levels. The prognostic validity of the motor tests was determined using analysis of variance, odds ratios, and a discriminant analysis. All diagnostic methods exhibited medium to high prognostic validity over the 2- to 6-year time span from the talent screening and talent orientation period to the later soccer competitions in the early adolescent U12–U15 age groups. The prognostic relevance of the nine general motor tests was confirmed.

Keywords

Talent · Talent screening · Sport orientation · Motor tests · Aptitude · Motor skills · Athletic performance · Soccer

Prognostische Validität der Talentorientierung im Fußball

Zusammenfassung

Üblicherweise nutzen Talentsuchprogramme, die im Grundschulalter durchgeführt werden, motorische Tests, um Kindern eine ihrem Eignungsprofil entsprechende Sportart zu zuzuweisen; darunter fallen beispielsweise Mannschaftssportarten oder Individualsportarten wie Turnen, Skilauf oder Tennis. Hierbei scheint allerdings der Erfolgswert dieser frühen Talentorientierung in den meisten Sportarten – so auch im Fußball – ungeklärt. In der vorliegenden prospektiven Kohortenstudie wurde die mittelfristige prognostische Validität allgemeiner motorischer Tests sowie eines, aus einer Testbatterie abgeleiteten, fußballspezifischen Testwertes untersucht. Die vorliegende Stichprobe besteht aus 2965 Kindern der Altersstufe U9 des Fuldaer Bewegungschecks. Beim Fuldaer Bewegungscheck handelt es sich um eine Testbatterie, bestehend aus zwei anthropometrischen Parametern und acht allgemeinen Bewegungsaufgaben des Deutschen Motorik-Tests 6–18, zusätzlich erweitert um einen Ballweitwurftest. Die Testdaten wurden in den Grundschulklassen der Jahre 2011–2014 erhoben. Für die Studie wurde nachfolgend die Wettkampfleitung zum Ende der Saison 2016/2017 (30. September) derjenigen Kinder betrachtet, die an der Testung teilnahmen und im Nachhinein die Sportart Fußball ausübten ($n = 316$). Diese Gruppe von Spielern der Altersstufe U12–U15 wurde dann individuell vier verschiedenen Wettbewerbsstufen zugeteilt. Die prognostische Validität des Motoriktests (Fuldaer Bewegungscheck) wurde anhand einer Varianzanalyse, Odds Ratios und einer Diskriminanzanalyse ermittelt. Alle diagnostischen Verfahren zeigten eine mittlere bis hohe prognostische Validität über die Zeitspanne von vier bis sechs Jahren – von der Talentsuche und Talentorientierungsphase bis zu späteren Fußballturnieren in der Altersgruppe der frühen Jugend (U12–U15). Die prognostische Relevanz der neun allgemeinen motorischen Tests wurde somit bestätigt.

Schlüsselwörter

Talent · Talentsuche · Motorische Tests · Begabung · Motorische Fertigkeiten · Sportliche Leistungsfähigkeit · Fußball

Table 1 The 4-level scale recording sport-specific participants' competition results of the Fulda Movement Check (FMC) taking part in official competitions for age groups U12 to U15 ($N = 316$)

4-level scale of U12 to U15 age group competition results		
Low soccer performance group ($N = 221$)		
Level 1	$N = 41$ (2 + 39)	Holder of an official club license in a specific sport or competition participation on the lowest local level ("Kreisklasse") at the age of U12 to U15
Level 2	$N = 180$	Competition participation at the district level ("Kreisliga") at the age of U12 to U15
High soccer performance group ($N = 95$)		
Level 3	$N = 85$	Competition participation at the county level ("Bezirksliga") at the age of U12 to U15
Level 4	$N = 10$ (8 + 2)	Competition participation at the regional level ("Verbandsliga", "Hessenliga") or national level ("Regionalliga", "Bundesliga") at the age of U12 and U15

ethics department approved the implementation of this study.

Measurements

Anthropometric characteristics and motor abilities

Aside from the motor tests of the FMC, each player's height, weight, and calendar age (measured by the month of birth within a calendar year) were registered. The motor test battery consisted of nine tests, of which one (20-m sprint) was negatively coded so that a lower value indicated a better performance. Players were tested according to the GMT 6–18 (Boes & Schlenker, 2016), including sprint[1], coordination[2], balance[3], flexibility[4], arm and upper body strength[5], leg power[6], and endurance[7]. All tests were executed according to the test description provided by Boes and Schlenker (2016). In addition to the GMT 6–18, a ball throw[8] was added to the FMC to contribute to judging talent for track and field as well as game sports like handball, water polo, volleyball, tennis etc. In the sit-up and push-up tests as well as in the 6-min run, only one trial was executed, whereas in the other tests two attempts were allowed. For these tests, the better result between the two trials was recorded except for the sideward jumping test, where the mean value of the two trials was recorded. Players were provided with sufficient time for recovery between each attempt.

Boes and Schlenker (2016) analyzed the test battery's psychometric properties for a sample consisting of nearly 50,000 school children and adolescents. The authors found good internal consistencies in terms of an average test-retest correlation coefficient of $r_{tt} = 0.82$ at the elementary school level (7–11 years), ranging from $r_{tt} = 0.52$ for balancing backward to $r_{tt} = 0.94$ for bending forward. The ball throw that was added in the FMC showed test–retest reliability of $r_{tt} = 0.82$ ($p < 0.001$; $n = 3193$) up until the end of 2015.

For the evaluation of motor predictors' prognostic validity, the age influence on the test performances should be considered (Höner et al., 2017; e.g., Meylan, Cronin, Oliver, & Hughes, 2010; Höner & Votteler, 2016; Carling & Collins, 2014). Univariate ANOVAs (analysis of variance) were conducted to check the data set for significant differences between youth athletes in regard to calendar age. As age was growing with the birth quartiles ($F(3; 3186) = 7.61$, $p < 0.05$), the calendar age (in months) was partialized out of these results through all predictors by bivariate regression analysis to avoid confounding effects in the following analysis (Willimczik, 1982; Hohmann et al., 2015; Hohmann et al., 2017a, 2017b). In the bivariate regression analysis, the test results served as dependent, and the age (in months) as independent variables (Willimczik, 1982). To allow for comparisons between the different predictors, the residuals of the bivariate regression were standardized by z-values. In all procedures, data were analyzed using SPSS version 25.0 (IBM Corp., Chicago, IL, USA), and the minimum level of significance was set at $p < 0.05$.

Aside from the single motor tests, the prognostic validity of a soccer-specific total score was derived from the general test battery being evaluated. Therefore, the mean value of a selection of only five of the general tests was calculated. Each of these five tests was hierarchically weighted by a weight factor (WF) between 1.2 and 2.0 according to their estimated validity for soccer performance. Thus, the 20-m sprint (WF 2.0), sideward jump (WF 1.8), 6-min run (WF 1.6), standing long jump (WF 1.4), and sit-up tests (WF 1.2) were used because they had proved, in a previous study, to be more soccer-related than the other four tests (see Hohmann et al., 2015). Based on this weighting, the soccer recommendation score (SRS) was calculated.

Level of sport-specific competition

The competition performance level (CPL; **Table 1**) reached by soccer players as well as by athletes of all other sports until the end of the season 2016/17 was utilized to quantify all athletes' success in early adolescence as a criterion variable. In all sports, athletes who participated in the U9 motor diagnostics, participated at minimum in one official sport-specific competition until the end of the season 2016/17, or at least held an official club license without taking part in competitions (level 1), were recorded. Based on their success in single as well as team competition events, the results reached in these competitions were ranked from level 2 up to level 4 if the athlete was playing on the national level. The performance levels for game sport athletes, however, are difficult to judge (Höner & Votteler, 2016; Go-

[1] Time for a 20 m linear running sprint; measurement by means of light gates (Brower Timing Systems; Draper, UT, USA); error correction of timing was 0.26 s; starting position was 0.3 m behind the start line.

[2] 15-s sideward jumping.

[3] 3 × 2 trials stepping backward on a 3.0, 4.5 and 6.0 cm wide beam.

[4] Bend forward.

[5] Timed push-ups and sit-ups with number of repetitions within 40 s.

[6] Standing long jump.

[7] 6-min run around a volleyball pitch.

[8] Ball weight: 80 g; both feet on ground behind a line; no step or run-up; measurement perpendicular to measuring tape; accuracy: 0.1 m.

Table 2 Descriptive statistics for the two anthropometric and nine motor diagnostics of second-grade students taking part in the Fulda Movement Check (FMC) 2011–2014 and playing soccer later on in age groups U12 to U15 until the end of season 2016/17

Variables	Groups	*N*	M	SD	SE	95%-CL		Min	Max
						LL	UL		
Calendar age (months)(U9)	Soccer players[a,b]	316	95.8	5.6	0.31	95.2	96.4	84	*124*
	Other athletes	435	94.3	4.9	0.23	94.0	95.0	80	*120*
	Non-athletes	2214	94.3	5.4	0.12	94.3	94.7	75	119
Calendar age (months)(U12–U15)	Soccer players	316	151.5	10.3	0.57	150.1	152.3	132	174
	Other athletes	435	149.3	10.3	0.50	148.1	150.0	131	178
Performance (level)(U12–U15)	Soccer players	316	2.1	0.5	0.03	2.0	2.2	1.0	4.0
	Other athletes	435	1.8	0.8	0.03	1.7	1.9	1.0	5.0
General tests of the Fulda Movement Check									
Body height (cm)	Soccer players	316	129.4	5.5	0.30	128.9	130.1	110	147
	Other athletes	434	129.4	6.2	0.30	128.8	130.0	110	146
	Non-athletes	2214	129.0	6.0	0.13	128.7	129.2	108	153
Body weight (kg)	Soccer players	316	28.0	5.0	0.28	27.5	28.6	18.6	54.8
	Other athletes	434	27.9	5.2	0.25	27.4	28.4	18.0	54.6
	Non-athletes	2214	28.1	5.7	0.12	27.9	28.3	16.4	70.0
Sideward jumping (repeats)	Soccer players[b]	316	25.2	5.8	0.32	24.6	25.9	11.0	44.0
	Other athletes[c]	434	24.5	6.2	0.30	24.0	25.1	6.5	40.0
	Non-athletes	2207	23.3	6.3	0.13	22.9	23.5	0.5	45.0
Balance backwards (steps)	Soccer players[b]	316	29.1	8.6	0.48	28.2	30.1	7.0	41.0
	Other athletes[c]	434	29.0	9.0	0.43	28.1	29.8	7.0	48.0
	Non-athletes	2212	26.7	8.9	0.19	26.3	27.0	3.0	48.0
Standing long jump (cm)	Soccer players[b]	316	136.0	16.4	0.92	134.4	137.9	70.0	173.0
	Other athletes[c]	430	134.0	16.8	0.81	132.4	135.6	79.0	185.0
	Non-athletes	2194	126.4	18.6	0.4	125.5	127.1	57.0	190.0
20-m sprint (seconds)	Soccer players	316	4.50	0.31	0.02	4.46	4.53	3.77	5.52
	Other athletes[c]	435	4.54	0.33	0.02	4.51	4.57	3.50	5.87
	Non-athletes	2214	4.57	0.38	0.01	4.56	4.59	3.50	7.15
Push-ups (repeats)	Soccer players[b]	316	14.2	3.7	0.21	13.9	14.7	1	25
	Other athletes[c]	434	14.0	3.6	0.17	13.6	14.3	4	25
	Non-athletes	2208	13.5	3.7	0.08	13.4	13.7	0	30
Sit-ups (repeats)	Soccer players[a,b]	316	20.8	4.9	0.27	20.3	21.4	0	35
	Other athletes[c]	434	19.3	5.8	0.28	18.8	19.9	0	37
	Non-athletes	2208	17.9	5.6	0.12	17.7	18.1	0	36
Bend forward (cm)	Soccer players[b]	316	1.2	5.4	0.30	0.6	1.8	−15.0	15.0
	Other athletes[c]	430	0.7	5.8	0.28	0.1	1.2	−16.0	15.0
	Non-athletes	2195	−0.1	5.9	0.13	−0.4	0.1	−23.0	20.0
6-min run (m)	Soccer players[a,b]	316	1014	122	6.87	1002	1029	558	1332
	Other athletes[c]	431	967	136	6.56	954	980	299	1332
	Non-athletes	2180	913	129	2.76	907	918	216	1269
Ball throw (m)	Soccer players[a,b]	316	16.2	4.0	0.22	15.8	16.6	5.1	27.5
	Other athletes[c]	432	14.6	4.2	0.20	14.2	15.0	6.0	29.0
	Non-athletes	2208	12.9	3.7	0.08	12.8	13.1	4.5	28.1

N number, *M* mean, *SD* standard deviation, *SE* standard error, *LL* lower level, *UL* upper level, *CL* confidence level, *Min* minimum, *Max* maximum
[a]Sign. between Soccer players and Other athletes
[b]Sign. between Soccer players and Non-athletes
[c]Sign. between Other athletes and Non-athletes

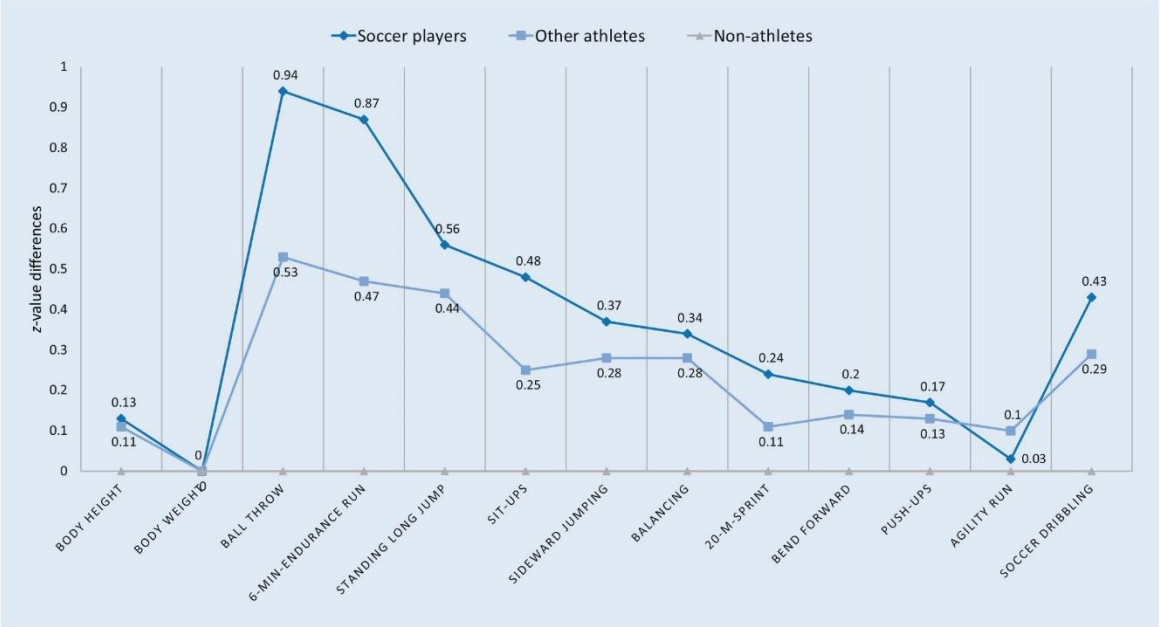

Fig. 2 ▲ Test performances (*z*-value) for later soccer players, other athletes, and non-athletes of the Fulda Movement Check (FMC) held in second grade at the elementary level

naus & Mueller, 2012; Roescher, Elferink-Gemser, Huijgen, & Visscher, 2010). To enhance the reliability of soccer players' assignments to the different competition performance levels, the records of the soccer performance development of the U12 to U15 age groups were checked by the head coach of the local competence centre; this individual was in charge of the nomination, selection, and education of soccer players promoted by the German Soccer Federation in the Fulda region during the investigation period.

Statistical analysis

All data were analyzed with SPSS (Version 25.0; SPSS Inc., Chicago, IL, USA) and statistical significance was set at $p < 0.05$. In order to obtain solid results in regard to the prognostic relevance of the predictors, the data sets of four successive second grade cohorts (2011, 2012, 2013, and 2014) were collected so that the samples of the soccer players and other athletes also achieved a sufficient number at higher competition levels. In order to achieve a greater comparability between the test values, the calendar age (in months) of the participants was partialized from the test results by means of a bivariate linear regression analysis, and then *z*-standardized residuals were used for all further analysis.

To gain insight into the prognostic relevance of the predictors of the FMC (two anthropometric variables and nine motor tests) in the soccer group, univariate ANOVAs were conducted analyzing mean differences between the different soccer performance levels. By means of a post hoc test (Bonferroni), the multiple differences between the groups were tested.

In addition, cross-tabulations were used to analyze the achievements of county level (level 3) or higher in terms of the single test results for 16 % of the best participants in the FMC. For a clearer indication of a player's relative chances to reach a higher performance level in the future (group 1), odds ratios (on the basis of the *z*-standardized residuals) were computed.

To obtain a mid-term talent prognosis in both a linear discriminant analysis and a nonlinear neural network (multilayer perceptron, MLP), five out of nine juvenile performance prerequisites of the FMC were used to predict two final groups of soccer players. The two performance groups were formed according to the performance level achieved at early adolescence in the age groups U12 to U15 (**Table 1**). Group 2 consisted of $n = 221$ players from lower levels 1 and level 2, and group 1 contained the $n = 95$ better-performing players from county level 3 up to regional and national level 4. To obtain a "true" prognosis, the talent forecasts on the basis of the stepwise discriminant analysis have to follow a cross-validation classification ("leave-one-out"). Thus, the total number of all cases minus one was used to compute the discriminant functions that were then used to determine the later-on performance group of the remaining single case. Similarly, for the multilayer perceptron (MLP) analysis, three subsets were created for (i) training and (ii) testing the predictive model as well as determining (iii) the final classification of the left-out cases. Therefore, the MLP was trained with 80 % of all cases, whereas 10 % was used to test the trained network. Finally, the classification was calculated for the hold-out of the remaining 10 % of cases. This specific kind of leave-out strategy was repeated eleven times so that each case should at

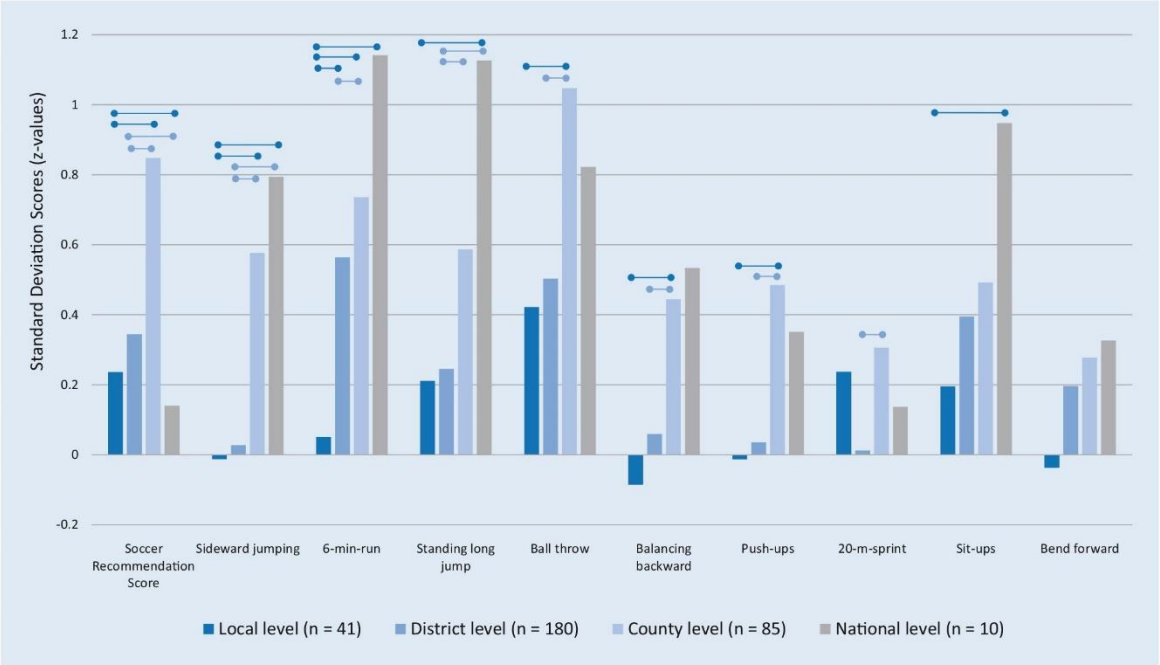

Fig. 3 ▲ Prognostic validity of the test performances of the U9 Fulda Movement Check (FMC) participants in regard to the later competition performance levels at early adolescence (U12 to U15). (Legend: Data of the 20-m sprint were transformed in a positive direction. Variances of subgroups were homogeneous [Levene; $p > 0.05$]. Significant group differences [Bonferroni; $p < 0.05$] are indicated)

least belong once to the left-out athletes who were finally classified. To quantify the validity of this talent identification strategy, the percentage of correct hits of the neural network classification was averaged over the ten trials, and the mean value was used from there on. The classification qualities of both methods were expressed by the proportion of correct hits, referring to the percentage of athletes that were assigned as true positives to their own sport. An athlete was defined as false positive if he was classified as a participant of a specific sport for which he did not practice. On the basis of the ratio of correct to erroneous hits, the prognostic validity of the test results was calculated.

Results

Test performance results for soccer players, other athletes, and non-athletes

Comparing the test results of later soccer players, other athletes, and non-athletes, it is noticeable that the tested soccer players in most of the nine tests showed the greatest conclusive performance (excluding the 6-min run). Regarding the field of anthropometric tests, there are no major differences in the results. The descriptive performance characteristics of the FMC participants are presented in ◘ **Table 2**.

The early talent screening results for the *z*-standardized predictors in ◘ **Fig. 2** demonstrate—in ten of the eleven motor diagnostics—better performances for future soccer players compared to athletes from 31 different sports (middle line) and non-athletes that served as reference groups.

Prediction of general motor tests

Comparison of different sport-specific competition levels

Within the group of soccer players ($n = 316$), the ANOVAs demonstrated significant differences in the predictors corresponding with later performance levels reached by the U12 to U15 children. Soccer players who reached the regional or national level 4 performed better in eight of nine motor tests than players who made it only to the county level 3, district level 2, or local level 1: total score[9], soccer recommendation score[10], sideward jumping[11], 6-min endurance run[12], standing long jump[13], ball throw[14], balancing backward[15], push-ups[16], 20-m sprint[17], and sit-ups[18]. In the row of general tests, only the results in the bend forward flexibility test[19] do not go hand-in-hand with the later achieved soccer performance level in the U12 to U15 age

[9] GMT in addition to Ball throw; $F(3;312) = 12.99$; $p < 0.001$; $\eta^2 = 0.111$.

[10] $F(3;312) = 10.03$; $p < 0.001$; $\eta^2 = 0.087$.

[11] $F(3;312) = 10.02$; $p < 0.001$; $\eta^2 = 0.086$.

[12] $F(3;312) = 6.83$; $p < 0.001$; $\eta^2 = 0.061$.

[13] $F(3;312) = 6.51$; $p < 0.001$; $\eta^2 = 0.057$.

[14] $F(3;312) = 6.00$; $p < 0.01$; $\eta^2 = 0.053$.

[15] $F(3;312) = 4.85$; $p < 0.01$; $\eta^2 = 0.044$.

[16] $F(3;312) = 4.81$; $p < 0.01$; $\eta^2 = 0.043$.

[17] $F(3;312) = 3.74$; $p < 0.05$; $\eta^2 = 0.034$.

[18] $F(3;312) = 2.73$; $p < 0.05$; $\eta^2 = 0.025$.

[19] $F(3;312) = 1.29$; $p = 0.278$.

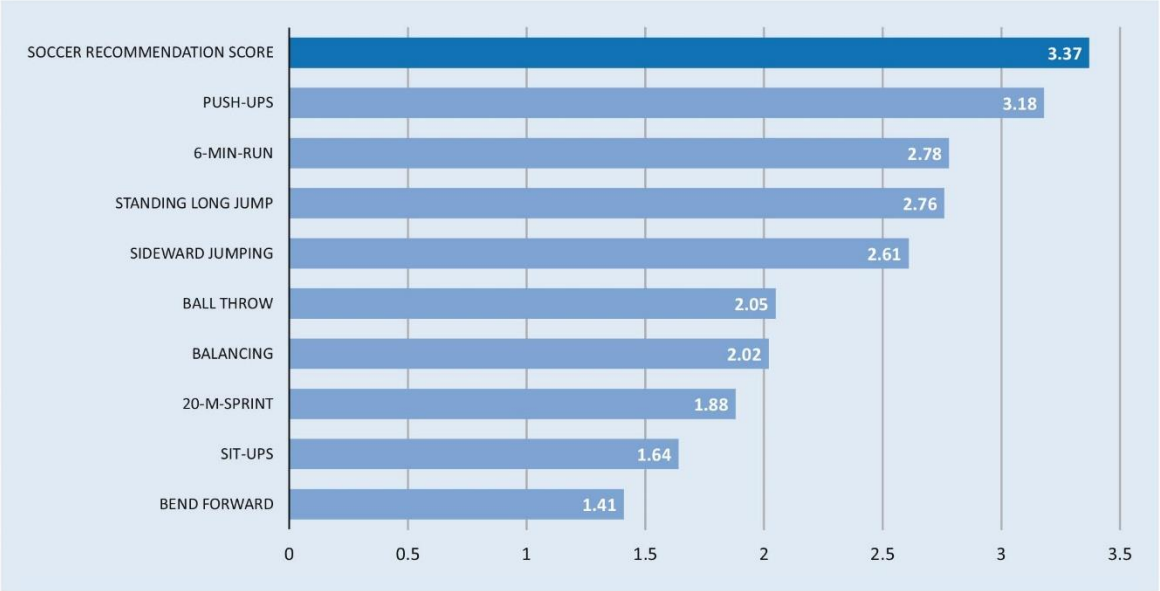

Fig. 4 ▲ Odds ratios (at the $z = 1.0$ and percentage rank [*PR*] = 84 threshold) of the nine general motor tests of the Fulda Movement Check's (FMC) basic test and the more specific SRS predict the relative chances of the U9 participants of the FMC to reach at least the county level (levels 3 or 4) of soccer competition performance by an early adolescent age (U12 to U15)

categories (**Fig. 3**). As the anthropometric characteristics body height and weight did not relate significantly to the later soccer performance, those data were not presented in the figure.

The post hoc comparison of the four subgroups showed significant mean differences in nearly all predictors, including the two overall measures of total score for the GMT added by ball throw and weighted SRS; the bend forward results are an exception (**Fig. 3**). Of the general GMT items, only the 6-min run discriminated significantly within the majority of those U9 children that later on formed the lowest U12 to U15 performance groups at both local and district levels.

Odds ratios for each test and the soccer recommendation score

Odds ratios (ORs) for each single test as well as for the SRS represent the prognostic validity of the investigated predictors and make the sport-specific relevance of the different tests comparable (**Fig. 4**). In the context of this study, the ORs quantify the relative chances of a U9 FMC participant in reaching a certain soccer performance level in early adolescence (U12 to U15). In **Fig. 4**, the OR for the SRS and any of the nine single tests were calculated for children who had achieved a total score among the best 16 % of the total group ($z \geq 1.0$ or percentage rank (PR) ≥ 84). The OR ranking in **Fig. 4** reveals the relative chances of a participant to reach at least the county soccer performance level (3 or 4) in the future.

Specificity and sensitivity

Since soccer is characterized by a great number of beginners at the elementary school level, as well as by a broad variety of demands and necessary abilities due to the complexity of the game, the quality of sports recommendation in the talent identification model is of great importance (Pion, 2015). Therefore, factors including how many non-talented athletes will be correctly classified (specificity of the testing) and how many talented youngsters will be successfully identified (sensitivity of the testing) by the test battery of the FMC are of great interest (see also Höner & Votteler, 2016). **Fig. 5** illustrates that at a test performance limit of $z = 1.0$ ($PR = 84$ respectively) in the SRS, according to the sensitivity curve, 53.6 % of the FMC participants were correctly identified as future successful soccer players (true positives), whereas in regard to specificity, 75.1 % of the non-talented players (true negatives) were excluded from talent promotion if the campaign aimed for county soccer performance level (level 3 and higher). Alternatively, even better players that will perform a minimum at the regional level 4 in the future could be sought after. In such cases, 83.3 % of the later groups of successful soccer players were correctly predicted, and 70.7 % of the non-talent individuals were sorted out.

The "true" performance prognosis by the cross-validated discriminant analysis on the basis of the five SRS tests led to a correct classification of 73.7 % of U9 soccer players into their later lower or higher performance groups at early adolescence (**Fig. 6**). In all, 205 of the 221 weaker (92.8 %) and 28 of the 95 stronger (29.5 %) soccer players were classified correctly into the performance groups that reached them more than 4 years later. The more efficient prognosis of future low-performers remaining on the local level underlines the high specificity of SRS derived from the FMC.

If an Artificial Neural Network (Multilayer Perceptron; SPSS 25.0, IBM Corp.,

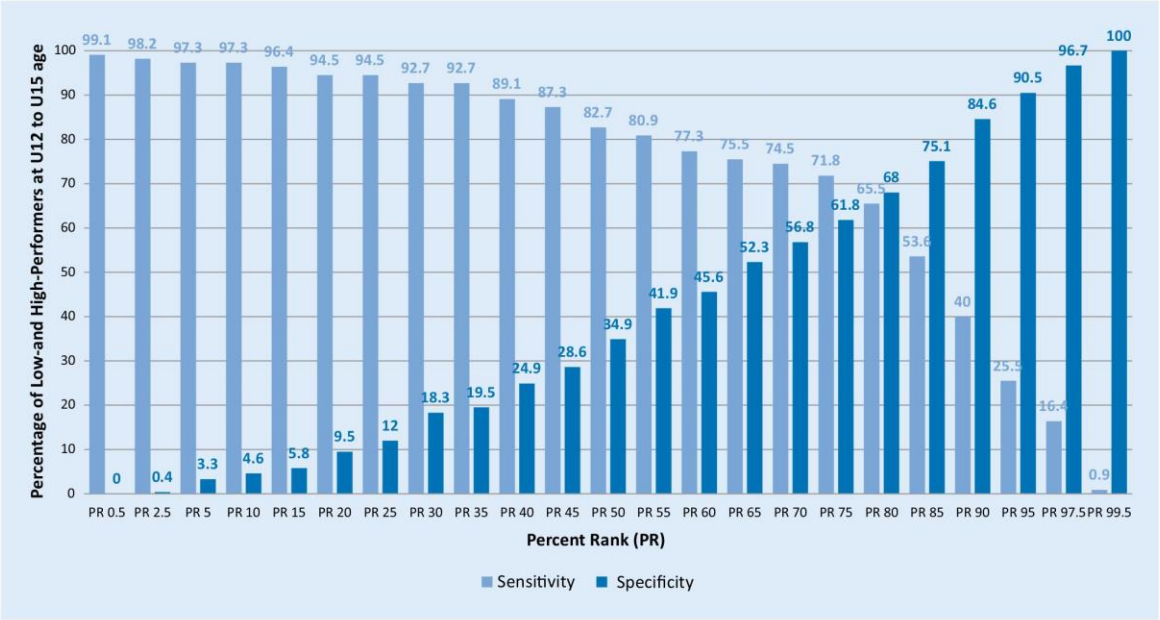

Fig. 5 ▲ Specificity and sensitivity of soccer recommendation score (SRS) of the Fulda Movement Check (FMC) U9 participants to predict a future soccer performance of at least county, regional, and national levels (3 or 4) at early adolescence (U12 to U15). *PR* percentage rank

Chicago, IL, USA) is used, the cross-validated procedure led to a somewhat lower prediction of 70.7 % correct hits. Especially evident in the group of non-talented players, the number of correct low performance predictions (88.2 % true negatives) was smaller, whereas the predictions of the later soccer players with higher performances (29.7 % true positives) remained more or less at the same value.

Discussion

It is imperative that talent screening and sports orientation campaigns based on the assessment of the performance predictors and competition performance development are closely related to a long-term talent development process, which includes proper allocation to defined stages of a complex talent development model. Compared to recent studies (Höner & Votteler, 2016; Höner et al., 2017) investigating the prognostic relevance of talent predictors in youth soccer, this study assessed a very young sample over a mid-term period from the early talent screening in second grade (U9) until the period of the talent promotion program of the German Soccer Federation at an early adolescent age (U12–U15). As a multidimensional talent screening campaign, the FMC is based on the GMT 6–18. Thus, a broad variety of physiological (e.g., endurance run), psychomotor (e.g., sideward jumping), and anthropometric (e.g., body height) predictors relevant for the talent identification in soccer were included in the study.

The present results contradict the later importance of the anthropometric variables body height and weight found by Höner et al. (2017) in early adolescence for more or less successful adult players. These findings could have resulted from two different factors. On one hand, the elimination of the calendar age from our test data by partializing out the age (in months) from all test performance data could have led to a more homogeneous "age group". On the other hand, there could certainly be a considerably lower relative age effect in the early talent development stage compared to middle or late adolescence. Due to the early stage of the talent screening campaign, the participants are still far from prepubertal acceleration regarding the development of body dimensions; this factor might reduce the impact of the anthropometric predictors on the soccer performance of younger individuals. To fully understand the relationship between biological maturity (e.g., skeletal age; Lago-Peñas, Casais, Dellal, Rey, & Domínguez, 2011) and soccer performance development from childhood to early adolescence, more differentiated analyses would be necessary. The five test results of the 20-m sprint, sideward jumping, 6-min run, agility run, standing long jump, and sit-ups form the basis of an SRS. Thus, it is not surprising that the later soccer group already exhibited in the talent screening campaign at age U9 contributed much better soccer-related performance prerequisites compared to non-athletes. Even more interesting are the leads of soccer players compared to the group of athletes from 31 other sports. The great differences in 20-m sprint, 6-min run, and the sit-ups, which are also part of the specific and highly valid SRS, might be indicators of early soccer specific athleticism that is relevant for soccer talents. One

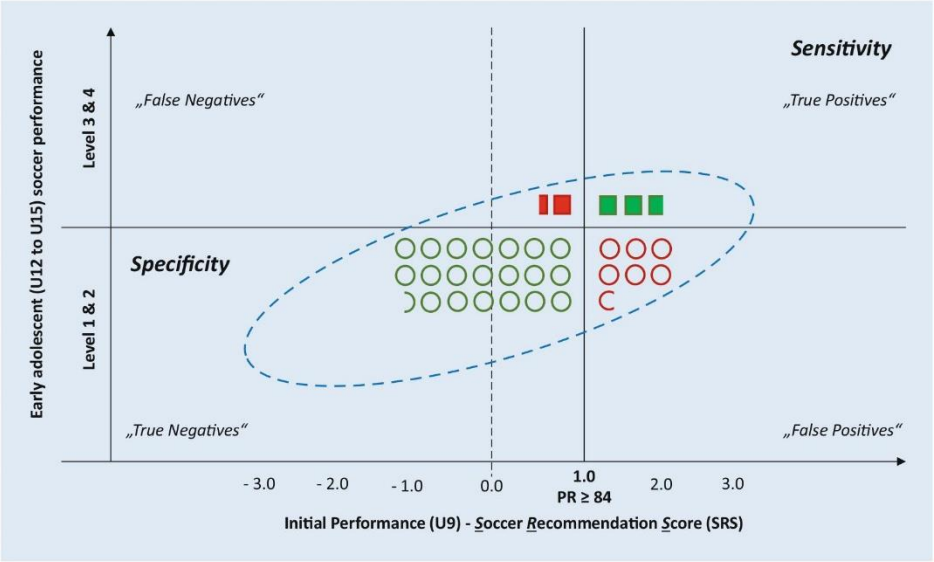

Fig. 6 ◄ Discriminant analysis to predict two later U12 to U15 soccer performance groups based on U9 performances in the five tests of the soccer recommendation scort (SRS; 20-m sprint, sideward jumping, 6-min run, standing long jump, and sit-ups) of the Fulda Movement Check (FMC) (at the limit of percentage rank [*PR*] = 84 or *z* = 1.0; each full symbol represents ten children)

reason for the somewhat surprising excellent throwing performance of the soccer group could lie in the fact that the best movers get drafted by soccer clubs at a very young age, as the soccer federation recommends and promotes a very early recruitment of promising youngsters. This also underlines that the general athleticism should be regarded in the search for soccer talents.

According to the descriptive data, the ANOVA results, and especially the odds ratios of the motor tests, this mid-term longitudinal study verified the prognostic validity of the nine general motor tests of the FMC. Although the investigation period was almost the earliest possible in children's participation, the findings for the predictive validity of the general tests of the GMT 6–18 in general were not surprising; these results were expected to be fairly high because the best movers (Gulbin, Croser, Morley, & Weissensteiner, 2013) at the elementary level generally become involved with soccer at a very early age, as it is the most popular sport in Germany.

The prognostic period of our study ends exactly when the talent selection of the German Soccer Federation (Deutscher Fußball-Bund, 2009) starts. This study may complete the picture from early talent screening in very heterogeneous samples until the talent selection stages begin focusing on very homogeneous soccer populations striving for top performance levels at junior and adult age.

Conclusion

The study provides reliable empirical knowledge on the prognostic relevance of general and soccer-related coordinative and technical skill tests in a regional talent screening and sports orientation campaign. The study shows that regional talent screening and sports orientation can be used to make valid statements in regard to young soccer players' future success. The results demonstrated motor predictors' prognostic validity over a mid-term period (on average about 4.5 years) after controlling all test data for calendar age. The specificity of a general testing in the second grade is very high and the majority of non-talented individuals are advised to pursue other sports for which they are more closely suited. However, due to the low sensitivity of the FMC for talent orientation, in the future an additional soccer-specific testing should be applied to obtain more reliable information for the sports recommendation regarding soccer. Thus, introducing elementary school children to soccer still remains a very complex practical and theoretical problem.

With a focus on practical utility of motor tests, the GMT 6–18 (Boes & Schlenker, 2016) assessing speed, endurance, and coordinative abilities turned out to be a useful tool for talent orientation as both a combination of talent screening and recommendation of suitable sports. The specificity of the general testing in the second grade is very high and the majority of non-talented individuals are advised to pursue other sports for which they are better suited. However, due to the low sensitivity of the general motor test, further sport-specific additional tests seem to be necessary for this purpose.

Further studies over longer periods of time (see also Höner et al., 2017), as well as other promising methodological approaches such as person-oriented pattern analyses (Zuber et al., 2016) and a combination of linear and nonlinear tools (Pion et al., 2016) should be examined and compared with one another to identify corresponding strengths and weaknesses. In doing so, the research on talent identification may provide coaches with more scientifically sound tools for supporting their talent identification strategies as well as offer a deeper understanding of the long-term development of talented soccer players in talent promotion.

Corresponding address

Maximilian Siener
Institute of Sports Science,
University of Bayreuth
Universitätsstraße 30,
95440 Bayreuth, Germany
maximilian.siener@uni-bayreuth.de

Acknowledgements. This study is part of the research project "talent identification" which was supported by the German Institute of Sport Science (Bonn, Germany). Also, we would like to thank the authorities of the city of Fulda, the Fulda district authorities and the participating elementary schools for supporting the Fulda Movement campaign since 2010.

Compliance with ethical guidelines

Conflict of interest. A. Hohmann, M. Siener and R. He declare that they have no competing interests.

The author did not conduct any studies with human or animal participants for this article. As for other studies cited in this article, information on ethical guidelines may be found in the respective sources.

References

Boes, K., & Schlenker, L. (2016). *Der Deutsche Motorik-Test 6–18* (2nd edn.). Hamburg: Feldhaus.

Carling, C., & Collins, D. (2014). Comment on "Football-specific fitness testing: adding value or confirming the evidence?". *Journal of Sports Science*, *32*(13), 1206–1208.

Carling, C., Le Gall, F., & Malina, R. M. (2012). Body size, skeletal maturity, and functional characteristics of elite academy soccer players on entry between 1992 and 2003. *Journal of Sports Sciences*, *30*(15), 1683–1693. https://doi.org/10.1080/02640414.2011.637950.

Deutscher Fußball-Bund (2009). *Talentförderprogramm. Leitfaden für die Ausbildung*. Münster: Philippka. Talent development program. Training guideline

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

Fuchslocher, J., Romann, M., Rüdisüli, L. R., Birrer, D., & Hollenstein, C. (2011). Das Talentselektionsinstrument PISTE: wie die Schweiz Nachwuchsathleten auswählt. *Leistungssport*, *41*(4), 22–27.

Le Gall, F., Carling, C., Williams, M., & Reilly, T. (2010). Anthropometric and fitness characteristics of international, professional and amateur male graduate soccer players from an elite youth academy. *Journal of Science and Medicine in Sport*, *13*(1), 90–95.

Golle, K., Muehlbauer, T., Wick, D., & Granacher, U. (2015). Physical fitness percentiles of German children aged 9–12 years: findings from a longitudinal study. *PLoS ONE*, *10*(11), e142393. https://doi.org/10.1371/journal.pone.0142393.

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. https://doi.org/10.1080/02640414.2012.713980.

Gulbin, J. P., Croser, M. J., Morley, E. J., & Weissensteiner, J. R. (2013). A closer look at the FTEM framework. Response to "More of the same? Comment on 'An integrated framework for the optimisation of sport and athlete development: a practitioner approach'". *Journal of Sports Sciences*, *32*(8), 796–800. https://doi.org/10.1080/02640414.2013.855806.

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

Hohmann, A., & Pietzonka, M. (2017). *Techniktraining zur Entwicklung der Spielfähigkeit im Fußball, Handball und Basketball*. Köln: Sportverlag Strauss. Technique training to develop play ability in soccer, handball, and basketball

Hohmann, A., & Seidel, I. (2003). Scientific aspects of talent development. *International Journal of Physical Education*, *40*(1), 9–20.

Hohmann, A., Fehr, U., & Voigt, L. (2015). Heute im Talentpool – In Hamburg auf dem Podium. *Leistungssport*, *45*(5), 5–11.

Hohmann, A., Hohmann, L., Scheuring, L., & Zapp, M. (2016). *Fuldaer Bewegungs-Check 2010-2015 – Vorläufiger Ergebnisbericht*. Bayreuth: Institute of Sport Science. Fulda movement check – preliminary results

Hohmann, A., Fehr, U., Siener, M., & Hochstein, S. (2017a). Talentscreening und Talentorientierung. *Leistungssport*, *47*(4), 11–13.

Hohmann, A., Fehr, U., Siener, M., & Hochstein, S. (2017b). Validity of early talent screening and talent orientation. In P. Platen, A. Ferrauti, E. Grimminger-Seidensticker & T. Jaitner (Eds.), *Sport science in a metropolitan area* (p. 590). Bochum: University Press.

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.

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. https://doi.org/10.1080/02640414.2014.928416.

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*(8), e182211. https://doi.org/10.1371/journal.pone.0182211.

Lago-Peñas, C., Casais, L., Dellal, A., Rey, E., & Domínguez, E. (2011). Anthropometric and physiological characteristics of young soccer players according to their playing positions: relevance for competition success. *Journal of Strength and Conditioning Research/National Strength & Conditioning Association*, *25*(12), 3358–3367. https://doi.org/10.1519/JSC.0b013e318216305d.

Leite, N., Baker, J., & Sampaio, J. (2009). Paths to expertise in Portuguese national team athletes. *Journal of Sports Science and Medicine*, *8*(4), 560–566.

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*(2), 131–146.

Meylan, C., Cronin, J., Oliver, J., & 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.

Mueller, L., Mueller, E., Kornexl, E., & Raschner, C. (2015). The relationship between physical motor skills, gender and relative age effects in young austrian alpine ski racers. *International Journal of Sports Science and Coaching*, *10*(1), 69–86.

Pankhurst, A., & Collins, D. (2013). Talent identification and development: the need for coherence between research, system, and process. *Quest*, *65*(1), 83–97.

Papic, V., Rogulj, N., & Plestina, V. (2009). Identification of sport talents using a web-oriented expert system with a fuzzy module. *Expert Systems with Applications*, *36*, 8830–8838. https://doi.org/10.1016/j.eswa.2008.11.031.

Pion, J. (2015). *The Flemish sports compass: from sports orientation to elite performance prediction*. Ghent: Ghent University Press.

Pion, J. (2017). *Sustainable investment in sports talent: the path to the podium through the school and the sports club*. Arnhem: HAN University of Applied Sciences Press.

Pion, J., Hohmann, A., Liu, T., Vandorpe, B., Lenoir, M., & Segers, V. (2016). Predictive models reduce talent development costs in female gymnastics. *Journal of Sports Sciences*, *35*(8), 806–811.

Roescher, C., Elferink-Gemser, M., Huijgen, B., & Visscher, C. (2010). Soccer endurance development in professionals. *International Journal of Sports Medicine*, *31*, 174–179. https://doi.org/10.1055/s-0029-1243254.

Stemper, T., Bachmann, C., Diehlmann, K., & Kemper, B. (2009). Das Düsseldorfer Modell der Bewegungs-, Sport- und Talentförderung (DüMo). In Bundesinstitut für Sportwissenschaft (Ed.), *Talentdiagnose und Talentprognose*. 2. BISp-Symposium: Theorie trifft Praxis. (pp. 139–142). Köln: Strauss. The Duesseldorf model of movement, sport and talent promotion.

Willimczik, K. (1982). Determinanten der sportmotorischen Leistungsfähigkeit im Kindesalter. In H. Howald & E. Hahn (Eds.), *Kinder im Leistungssport* (pp. 141–153). Stuttgart: Birkhäuser. Chapter: determinants of the motor performance at child age. Title: children in elite sport.

Zuber, C., Zibung, M., & 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. https://doi.org/10.3389/fpsyg.2016.01088.

Artikel 3

Siener*, M., & Hohmann, A. (2019). Talent orientation: the impact of motor abilities on future success in table tennis. *German Journal of Exercise and Sport Research*, *49*(3), 232–243. https://doi.org/10.1007/s12662-019-00594-1

Main Article

Ger J Exerc Sport Res 2019 · 49:232–243
https://doi.org/10.1007/s12662-019-00594-1
Received: 30 November 2018
Accepted: 19 May 2019
Published online: 12 June 2019
© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature 2019

Maximilian Siener · **Andreas Hohmann**
Institute of Sports Science, University of Bayreuth, Bayreuth, Germany

Talent orientation: the impact of motor abilities on future success in table tennis

Introduction

Three Chinese and two German players (International Table Tennis Federation, 2018) are in the top five of the current table tennis world rankings (November 2018). To reach the world's top athletes' performance level, future generations must be optimally educated and trained in sports. In this process, the search for table tennis talents can be seen as a basic step. Some long-term talent development programs use motor tests as part of a talent screening campaign to assess the characteristics profiles of young athletes (Hohmann, Siener, & He, 2018; Pion, 2015). In a second step, based on the individual level of skills and abilities, a personal recommendation for a certain sport that best matches the athlete's performance is given to each participant (Pion, 2015, 2017). Since many sports, including table tennis, are accompanied by a high level of technical, coordination and tactical demands (Faber, Bustin, Oosterveld, Elferink-Gemser, & Nijhuis-Van der Sanden, 2016; Munivrana, Furjan-Mandić, & Kondrič, 2015a) that need to be developed early on, many talent search programs start in primary school (e.g., Stemper, Bachmann, Diehlmann, & Kemper, 2009; Pion, 2015; Fuchslocher, Romann, Rüdisüli, Birrer, & Hollenstein, 2011; Golle, Muehlbauer, Wick, & Granacher, 2015). On the one hand, early recognition of talents goes hand in hand with a possible early entry into a suitable sport and, according to Doherty, Martinent, Martindale, and Faber (2018), there is also a higher chance of competing in professional competitions for a greater number of training years. On the other hand, some sports require a very early start, since the age in the international field is also very young (Papic, Rogulj, & Plestina, 2009). According to Hohmann and Seidel (2003), sport recommendations can lead to a higher overall quality in sport clubs. This can be explained by the fact that talents do not "waste" their potential in a sport that is inappropriate for them, but can develop their qualities early in their most suitable sport (Pion, 2015). These sports recommendations made in the context of talent orientation can focus on groups of sports such as game sports (Pion, 2015) or more precisely on single sports disciplines such as alpine skiing (Mueller, Mueller, Kornexl, & Raschner, 2015).

It is sometimes difficult to make a reliable table tennis recommendation as part of the talent search programs. Table tennis is characterized by outstanding motor control (perceptuo-motor skills and coordination) (Faber, Elferink-Gemser, Oosterveld, Twisk, Nijhuis-Van der Sanden, 2017a; Munivrana, Petrinović, & Kondrič, 2015b), fast footwork (Nikolic, Furjan-Mandić, & Kondrič, 2014), timing, as well as anticipatory (Akpinar, Devrilmes, & Kiraczi, 2012) and cognitive skills (Chu, Chen, Chen, Huang, & Hung, 2012; Lopez & Santelices, 2012). A large number of these factors cannot be adequately assessed by the usual motor tests (Bös & Schlenker, 2016). With regard to these difficulties, some authors try to optimize the test selection (Faber, Pion, Munivrana, Faber, Nijhuis-Van der Sanden, 2017b; Faber, Oosterveld, & Nijhuis-Van der Sanden, 2014). The risk here lies in the generality–specificity continuum. If one wants to reproduce the characteristics of certain sports well, the tests become more and more specific and the test battery larger. Furthermore, there is a danger that more techniques than abilities will be tested.

Additionally, to optimize the overall test result, there is also the possibility of weighting the skills in an analysis to obtain a clearer picture of the characteristic test profile (Hohmann, Fehr, Siener, & Hochstein, 2017a) and thus allow a more valid recommendation. By means of differentiated weighting, individual recommendation scores can be determined for each sport and can then be compared. Finally, the highest recommendation value serves as the basis of a future performance forecast. In addition to the "positive" prerequisites, various skills can also seem unimportant or have a negative influence on later success (Kondrič, Zagatto, & Sekulic, 2013). Perhaps it can be seen from these complex recommendation structures why, with a few exceptions (Faber, Elferink-Gemser, Oosterveld, Twisk, & Nijhuis-Van der Sanden, 2017a; Hohmann et al., 2018), there is a large gap in the long-term studies on the prognostic value of childhood performance prerequisites for performance in adolescence. As many studies only consider short or medium time intervals (Höner & Votteler, 2016; Höner, Leyhr, & Kelava, 2017; Le Gall, Carling, Williams, & Reilly, 2010), the results are sometimes ambivalent. While some of these studies could prove the validity of sport-specific motor tests (Höner & Votteler, 2016; Figueiredo, Concalves, Coelho, Silva, & Malina, 2009; Zuber, Zibung, & Conzelmann, 2016), other studies showed no

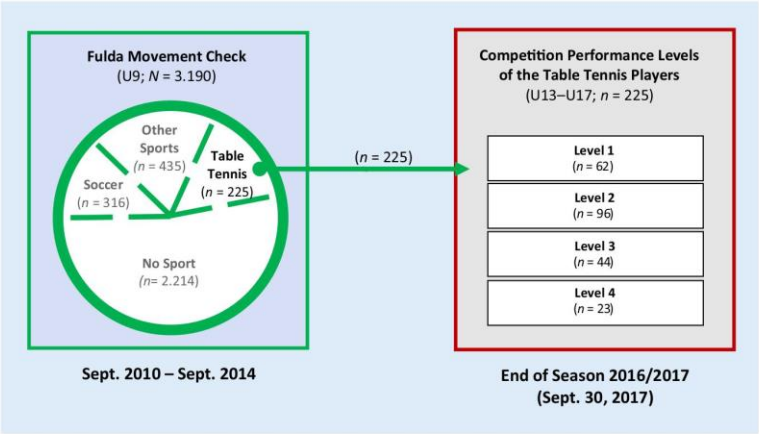

Fig. 1 ▲ Scheme of the different participants' groups of the Fulda Movement Check (2010–2014) included in this study and their performance levels in adolescence (2016/2017)

significant correlation between initial test performance and future sport success (Carling, Le Gall, & Malina, 2012; Le Gall et al., 2010). So far, it remains questionable, not least due to the many short-term investigations, whether there is a direct connection between initial performance and later success in certain sports.

The aim of this study is to narrow the gap and look at the prognostic validity of talent orientation using table tennis as an example. The midterm prognostic validity of a general motor test battery and a calculated table tennis recommendation score will be evaluated. To ensure this, competitive players at the table tennis performance levels of U13–U17 (age range from 11.0 to 17.0 years) in the 2016/2017 season (closing date was September 30, 2017) were compared, with their skill profiles collected in a test battery at the U9 age (age range from about 7.0 to 9.0 years) from the years 2010–2014. It was hypothesized that those athletes that performed better in early testing and received a higher table tennis recommendation score (TTRS) should, 4–9 years later (U13–U17), play on a comparatively higher competition level than athletes with lower test scores. In addition, the situation arises where, in single tests of the initial talent screening campaigns, these later table tennis players performed better than their counterparts performed in other sports.

Methods

General study design

The study is based on the results of the Fulda Movement Check (FMC) campaign between 2010 and 2014. The FMC was first launched as a pilot project in 2010 and is a talent and health campaign of the city and district of Fulda together with the University of Bayreuth (Hohmann, Fehr, & Voigt, 2015; see also Hohmann et al., 2017a; Hohmann, Fehr, Siener, & Hochstein, 2017b). This talent screening campaign tested almost all second classes of the Fulda district and city primary schools (more than 13,500 participants) to date by means of a general motor test battery. In addition to general anthropometric parameters such as height and weight, eight general tests of the German Motor Test (GMT) 6–18 (Bös & Schlenker, 2016) and a ball-throw are carried out. Based on the individual results, a recommendation score (RS) is calculated for each of 25 different sports disciplines. The three highest scores then result in the individual sports recommendation, which consequently consists of three sports suitable for each child. At the end of the tests, each participant is handed a certificate showing the different test results and listing the three individual sports recommendations derived from the individual ability profile. The sport aptitudes are documented using a percentage scale. The recommendations focus on medium-to-high performance sports disciplines available in the Fulda region. In Fulda these include, for example, swimming, track and field, gymnastics, cross-country skiing, soccer, handball, water polo, tennis, table tennis, judo, and canoeing. Other sports that do not provide high performance–oriented training (e.g., basketball, badminton, golf, etc.) or where training only begins at a higher age (e.g., baseball, rugby, boxing, etc.) are not recommended. Those athletes that have arrived at a table tennis club after the testing, regardless of their initial recommendation, were determined by daily print media and table tennis website analysis up to the deadline of this study (September 30, 2017). The initial individual test performances in the FMC at age U9 were then compared with the maximum table tennis competition performance up to ages U13–U17, to provide information about the prognostic validity of the early talent orientation.

Participants

A total of 3190 participants of a test campaign between 2010 and 2014 were found to take part in later sports competitions. Of them, 225 students played table tennis ($n_{♀} = 84$ and $n_{♂} = 141$) until the cut-off date (September 30, 2017). Their mean age at the time of initial FMC testing was $M = 94.1 \pm 5.2$ months. **Fig. 1** shows the exact distribution of the participants tested in the FMC 2010–2014. Soccer was explicitly mentioned here, as this sport is the most widespread in Germany, leading to a comparably good data basis ($n = 316$). Performance until the cut-off date in the 2016/2017 season was individually evaluated for each player using a four-level scale (see **Table 1**). Level 4 represents the highest performance group of the determined table tennis players.

To generate a larger sample, the test values for the years 2010, 2011, 2012, 2013, and 2014 were analyzed. However, as a result, the players' ages at the end of the 2016/17 season varies widely between $Min = 132$ months (11.0 years) and $Max = 200$ months (16.7 years). For the lowest age difference at the time when the

performance levels were collected, this means that the test person had reached his or her highest performance level more than 3 years (3.6 years) after testing in the FMC. The average age difference between the initial performance shown in U9 and the related performance shown in U13–U17 is 5.1 ± 1.0 years.

Within the scope of the FMC campaign, all parents of participating children were fully informed about the evaluations and objectives of the study and provided written consent. University and school ethical guidelines were strictly adhered to.

Measurements

Anthropometric characteristics and motor abilities

In the FMC campaign, anthropometric data were collected in addition to calendar age (in months) and gender. Body mass index (BMI) and Broca index were later calculated from height and weight. The motor tests were based on the German Motor Test 6–18 (GMT) (Bös & Schlenker, 2016). In the test battery, the abilities sprint[1], coordination[2], balance[3], flexibility[4], arm and upper body strength[5], leg power[6], and endurance[7] are tested. The test was conducted in accordance with the execution methods of the corresponding test manual. In addition to this general assessment of the GMT 6–18, a ball-throw[8] was performed to show the importance of throwing for many sports (water polo,

[1] Time for a 20 m linear running sprint; measurement by means of light gates (Brower Timing Systems; Draper, USA); error correction of timing was 0.26 s; starting position was 0.3 m behind the start line.

[2] 15 s of sideward jumping.

[3] 3 × 2 trials stepping backward on a 3.0, 4.5, and 6.0 cm wide beam.

[4] Bend forward.

[5] Timed push-ups and sit-ups with number of reps within 40 s.

[6] Standing long jump.

[7] 6-min run around a volleyball pitch.

[8] Ball weight: 80 gr; both feet on ground behind a line; no step or run-up; measurement perpendicular to measuring tape; accuracy: 0.1 m.

Abstract · Zusammenfassung

Ger J Exerc Sport Res 2019 · 49:232–243 https://doi.org/10.1007/s12662-019-00594-1
© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature 2019

M. Siener · A. Hohmann

Talent orientation: the impact of motor abilities on future success in table tennis

Abstract

In talent screening campaigns of primary school children, the aim is to identify talents and assign the children to a sport suitable to their characteristics. Here, the question arises whether these tests can help find talents at the elementary school age. Therefore, this study on table tennis players examined whether second-grade children tested with a battery of nine motor tests (Fulda Movement Check) were able to reach the predicted performance level in later adolescence. For this prognostic validity study, $N = 225$ U9 table tennis players were tested with two anthropometric and nine general motor tests between 2010 and 2014. These characteristics were then related to the success that table tennis players achieved by 2017. The comparison of table tennis players with soccer players, athletes of other sports, and nonathletes by means of a univariate analysis of variance (ANOVA) showed that table tennis players were clearly superior to other sports groups in the test tasks of sideward jumping and push-ups. Odds ratios for certain test performance thresholds were calculated to determine probabilities of later success. For a special table tennis recommendation score based on a selection of six of the nine tests, the identification and classification chances (sensitivity and specificity) were determined using a linear discriminant analysis and a neural network (multilayer perceptron). All in all, a medium to high prognostic validity could be proven with the complete motor test battery as well as with the table tennis recommendation score.

Keywords

Talent · Talent screening · Sport orientation · Validity · Motor skills · Table tennis

Talentorientierung: Einfluss der sportmotorischen Fähigkeiten auf den zukünftigen Erfolg im Tischtennis

Zusammenfassung

In Talentscreeningkampagnen für Grundschulkinder sollen Talente erkannt und Kinder Sportarten zugeführt werden, die zu deren Eigenschaften passen. Hierbei stellt sich mitunter die Frage, ob diese Testungen überhaupt ihr Ziel erfüllen und tatsächlich Talente in dieser Altersgruppe aufspüren können. In der vorliegenden Studie zu Tischtennisspielern wurde daher geprüft, ob die in einem sportmotorischen Test (Fuldaer Bewegungscheck) getesteten Grundschulkinder der zweiten Klasse im späteren Jugendalter das ihnen prognostizierte Leistungsniveau erreichen konnten. Für diese prognostische Validitätsstudie wurden in den Jahren 2010–2014 $n = 225$ Tischtennisspieler der U9 auf ihre sportmotorischen Fähigkeiten in zwei anthropometrischen und neun allgemeinen Motoriktests getestet. Die erhobenen Merkmalsausprägungen wurden anschließend mit den im Jahr 2017 erreichten Erfolgen in Bezug gesetzt. In einem zusätzlichen Vergleich von Tischtennisspielern mit Fußballspielern, anderen Sportlern und Nichtsportlern zeigte sich, dass Tischtennisspieler vor allem in der Testaufgabe des seitlichen Hin- und Herspringens und im Liegestütztest den anderen Sportgruppen deutlich überlegen waren. Dies konnte auch durch eine univariate Varianzanalyse (ANOVA) bestätigt werden. Für bestimmte Testleistungsschwellenwerte wurden Odds Ratios berechnet, um die Wahrscheinlichkeiten des späteren Erfolgs zu bestimmen. Für einen eigens entwickelten Tischtennisempfehlungswert, basierend auf einer Auswahl von sechs der neun Tests, wurden zudem anhand einer linearen Diskriminanzanalyse und eines neuronalen Netzes (mehrlagiges Perzeptron) die Identifikations- und Klassifikationschancen (Sensitivität und Spezifität) bestimmt. Insgesamt konnte sowohl mit den sportmotorischen Tests als auch mit dem speziell berechneten Tischtennisempfehlungswert eine mittlere bis hohe prognostische Validität nachgewiesen werden.

Schlüsselwörter

Talent · Talentscreening · Sportartorientierung · Validität · Sportmotorische Fähigkeiten · Tischtennis

Table 1 Four-level scale recording the participants' sport-specific competition results of the FMC participants taking part in official table tennis competitions in the age groups of U13 to U17 ($N = 225$)

Four-level scale of the U13 to U17 age groups' competition results		
Level 1	$N = 62$	Holder of an official club license
Level 2	$N = 96$	Competition participation at the local level [a] (*"Kreisebene"*)
Level 3	$N = 44$	Competition participation at the regional level [a] (*"Stadtmeister"* or *"Kreismeister"*)
Level 4	$N = 23$	Competition participation at the national level [a] (*"Bezirksmeister," "Gaumeister,"* finals of *"Landesmeisterschaft,"* or even higher)

[a]Term is based on the highest participating performance class

handball, track and field) and thus to improve the accuracy of the recommendations for these sports. In sit-ups, push-ups, and the 6-min run, only one attempt was allowed, whereas two attempts were possible in all other disciplines. With the exception of sideward jumping (the average value was counted here), only the best trial was used for further data analysis.

In Bös and Schlenker's (2016) reliability investigations, nearly 50,000 children in the 7–11 age groups of the GMT 6–18 showed comparatively good values in the test-retest correlation. The coefficient was $r_{tt} = 0.82$. Balancing backward achieved the lowest value with $r_{tt} = 0.52$ and bend forward the highest value with $r_{tt} = 0.94$. The low reliability of the balance test has many causes. One reason lies in the concentration and motivation of the test persons—this is not always completely given in children with repeated tests (see Bös und Schlenker, 2016). Supplemented ball-throwing showed a test-retest reliability of $r_{tt} = 0.82$ ($p < 0.001$; $n = 3193$) in the authors' own study.

Since age might have an influence on the evaluation of the predictors' prognostic validity (Meylan, Cronin, Oliver, & Hughes, 2010; Carling & Collins, 2014; Höner et al., 2017; Höner & Votteler, 2016) for all variables, a univariate ANOVA was used to test the data set for significant differences in age. As age was rising with the birth quartiles ($F(3; 3186) = 7.61$, $p < 0.05$; Hohmann et al., 2018), the calendar age (in months) was partialized out of the test results by bivariate regression analysis to avoid confounding effects in the following analysis (Willimczik, 1982; Hohmann et al., 2015, 2017a, b). Subsequently, the residuals of this analysis were z-standardized.

The TTRS was composed of six weighted test results calculated by means of weight factors (WF) in descending order of importance from the sideward jumping (WF = 2.0), balancing backward (WF = 1.8), push-ups (WF = 1.6), standing long jump (WF = 1.4), ball-throw (WF = 1.1), and bend forward (WF = 1.1) values. The weighting allows a better representation of table tennis talent, which is in line with former investigations of elite sport players (see Hohmann et al., 2015). An RS calculated in this way is determined for each of the sports of interest, and thus the children's primary talents are identified.

Level of sport-specific competition

To compare the performances of the table tennis players with one another and also with athletes of other sports, different competition levels were defined. According to this principle, all players that participated in an official competition before the deadline of the 2016/17 season received a rating from level 1 to level 4 (**Table 1**). The lowest class were assigned to level 1. These athletes did not receive any notable awards and tended to practice the sport as a hobby ($n = 62$). The second level indicates that the athletes have played official matches beyond the lowest local level. This performance range contains $n = 96$ of the tested students. If an athlete performed at regional competitions or championships, he was assigned to level 3 ($n = 44$). All higher performances above this level were assigned to level 4 ($n = 23$). This highest level also comprises a female member of the German youth national team in table tennis, thus underlining the high quality of this subgroup. Nevertheless, it is difficult in table tennis to correctly determine a player's performance despite the existing ranking systems (Gonaus & Mueller, 2012; Roescher, Elferink-Gemser, Huijgen, & Visscher, 2010). To reduce the general difficulties in level assignment, regional talent scouts and table tennis coaches were asked to confirm the ratings.

To investigate the performance of table tennis players, surveys of coaches and active sports clubs' regional websites were conducted. A large data set was taken from the archives of the local daily newspaper and current local reporting. In addition, the results of the official cooperation partners of the "German Table Tennis Federation" (DTTB, www.tischtennis.de) were retrieved.

Statistical analysis

All analyses and evaluations were performed using SPSS software (version 25.0; SPSS Inc., Chicago, IL, USA), and statistical significance was set at $p < 0.05$. The participants' data collected in the years 2010–2014 were related to the test age in months. For this purpose, the calendar monthly age was determined for each student test value, and the age profiles (mean values, standard deviations) were analyzed using a bivariate linear regression. The z-standardized residuals of these regression calculations of the individual test values were then used for all subsequent evaluations (Hohmann et al., 2015).

To show significant group differences, a univariate ANOVA was performed for levels 1–4. This indicates the prognostic validity of the nine individual tests and the calculated TTRSs. Group differences were analyzed with a post-hoc test. The Bonferroni test was used for analyses of variance equivalence and the Dunnett T3 test for unequal variances.

On the basis of cross-tabulations, odds ratios (ORs) were determined. The OR informs about the chances of reaching the national performance level (level 4) when the test results exceeded a z-value of 1.0. The z-value of 1.0 determines the limit of one standard deviation and thus

Table 2 Descriptive statistics for the two anthropometric and nine motor diagnostics of second-grade students taking part in the FMC 2010–2014 and playing table tennis later on in age groups U13–U17 until the end of season 2016/17

Variables	Groups	*N*	M	SD	SE	95% CI		Min	Max
						LL	UL		
Calendar age (month) (Time of testing)	Table tennis players	225	94.1	5.2	0.38	93.1	94.9	80	114
	Soccer players	316	95.8	5.6	0.31	95.2	96.4	84	124
	Other athletes	435	94.3	4.9	0.23	94.0	95.0	80	120
	Nonathletes	2214	94.3	5.4	0.12	94.3	94.7	75	119
Calendar age (month) (U13–U17)	Table tennis players	225	154.4	12.4	0.83	152.7	156.0	132	200
	Soccer players	316	151.5	10.3	0.57	150.1	152.3	132	174
	Other athletes	435	149.3	10.3	0.50	148.1	150.0	131	178
Performance level (U13–U17)	Table tennis players	225	2.2	1.0	0.07	2.0	2.3	1.0	4.0
	Soccer players	316	2.1	0.5	0.03	2.0	2.2	1.0	4.0
	Other athletes	435	1.8	0.8	0.03	1.7	1.9	1.0	4.0
General tests in the Fulda Movement Check									
Body height (cm)	Table tennis players	225	128.2	5.9	0.40	127.4	128.9	112	146
	Soccer players	316	129.4	5.5	0.30	128.9	130.1	110	147
	Other athletes	434	129.4	6.2	0.30	128.8	130.0	110	146
	Nonathletes	2214	129.0	6.0	0.13	128.7	129.2	108	153
Body weight (kg)	Table tennis players	225	27.2	4.7	0.31	26.7	27.9	18.0	48.9
	Soccer players	316	28.0	5.0	0.28	27.5	28.6	18.6	54.8
	Other athletes	434	27.9	5.2	0.25	27.4	28.4	18.0	54.6
	Nonathletes	2214	28.1	5.7	0.12	27.9	28.3	16.4	70.0
Sideward jumping (repeats)	Table tennis players	225	24.7	6.2	0.41	23.9	25.5	8.0	39.0
	Soccer players	316	25.2	5.8	0.32	24.6	25.9	11.0	44.0
	Other athletes	434	24.5	6.2	0.30	24.0	25.1	6.5	40.0
	Nonathletes	2207	23.3	6.3	0.13	22.9	23.5	0.5	45.0
Balance backward (steps)	Table tennis players	225	28.7	8.9	0.59	27.5	29.8	8.0	48.0
	Soccer players	316	29.1	8.6	0.48	28.2	30.1	7.0	41.0
	Other athletes	434	29.0	9.0	0.43	28.1	29.8	7.0	48.0
	Nonathletes	2212	26.7	8.9	0.19	26.3	27.0	3.0	48.0
Standing long jump (cm)	Table tennis players	225	127.8	18.8	1.25	125.4	130.3	73.0	190.0
	Soccer players	316	136.0	16.4	0.92	134.4	137.9	70.0	173.0
	Other athletes	430	134.0	16.8	0.81	132.4	135.6	79.0	185.0
	Nonathletes	2194	126.4	18.6	0.4	125.5	127.1	57.0	190.0

corresponds to approximately the best 16% of the students in the individual test results. For each of the nine individual tests and the TTRS, the probability of reaching level 4 was calculated for the best performers compared to the 84% of the weaker children.

To obtain a midterm talent prognosis by means of both a linear discriminant analysis and a nonlinear neural network (multilayer perceptron, MLP), six of the nine FMC juvenile performance prerequisites (according to the TTRS: sideward jumping, push-ups, balancing backward, ball-throw, standing long jump, and bend forward) were chosen to predict two final groups of table tennis players. For this analysis, only the six TTRS tests were used to compare the three results (TTRS, discriminant analysis, and neural network). The two performance groups were formed according to the performance level achieved at early adolescence in the age groups U13 to U17 (see **Table 1**). Group 1 contained the $n = 23$ better players from the national level 4, while group 2 consisted of $n = 202$ players from the lower levels 1–3. To obtain a "true" prognosis, the talent forecasts on the basis of the stepwise discriminant analysis must follow a cross-validation classification ("leave-one-out"). Thus, the total number of all cases minus one was used to compute the discriminant functions that were then used to determine the later performance group of the remaining single case. Similarly, for the MLP analysis, three subsets were created for: (i) training and (ii) testing the predictive model, as well as determining (iii) the final classification of the left-out cases. The architecture of the feed-forward net consisted of the seven input neurons (the six z-transformed test values: balancing backward, sideward jumping, bend forward, standing long jump, ball-throw, and push-ups and a distortion neuron), one hidden layer with a maximum of four neurons independently selected by the system, and the two output neurons (high performer and low performer). The number of iterations was automatically selected and the process stopped when no further reduction of the error was detected. Since the sample size of the high performers (level 4) was too small for a general statement, these cases were doubled and, together with a group of low performers (levels 1–3) of the same size, used for the analysis. The sample alignment prevented the otherwise larger group of weaker athletes from distorting the neuronal network due to the greater a-priori probability. From the resulting total of 80 cases, 80% (approximately 64 cases) were used to train the neural network and another 10% (approximately eight cases) to test the resulting network (Hohmann et al., 2018). Finally, the classification was calculated for the hold-out of the

Table 2 (Continued)

Variables	Groups	*N*	M	SD	SE	95% CI		Min	Max
						LL	UL		
20-m Sprint (s)	Table tennis players	225	4.61	0.36	0.02	4.56	4.66	3.50	5.87
	Soccer players	316	4.50	0.31	0.02	4.46	4.53	3.77	5.52
	Other athletes	435	4.54	0.33	0.02	4.51	4.57	3.50	5.87
	Nonathletes	2214	4.57	0.38	0.01	4.56	4.59	3.50	7.15
Push-ups (repeats)	Table tennis players	225	14.0	3.8	0.26	13.5	14.5	3	25
	Soccer players	316	14.2	3.7	0.21	13.9	14.7	1	25
	Other athletes	434	14.0	3.6	0.17	13.6	14.3	4	25
	Nonathletes	2208	13.5	3.7	0.08	13.4	13.7	0	30
Sit-ups (repeats)	Table tennis players	225	19.0	5.4	0.36	18.3	19.8	0	37
	Soccer players	316	20.8	4.9	0.27	20.3	21.4	0	35
	Other athletes	434	19.3	5.8	0.28	18.8	19.9	0	37
	Nonathletes	2208	17.9	5.6	0.12	17.7	18.1	0	36
Bend forward (cm)	Table tennis players	223	1.3	6.0	0.40	0.5	2.1	−16.0	15.0
	Soccer players	316	1.2	5.4	0.30	0.6	1.8	−15.0	15.0
	Other athletes	430	0.7	5.8	0.28	0.1	1.2	−16.0	15.0
	Nonathletes	2195	−0.1	5.9	0.13	−0.4	0.1	−23.0	20.0
6-Min run (m)	Table tennis players	224	943	124	8.37	927	960	540	1242
	Soccer players	316	1014	122	6.87	1002	1029	558	1332
	Other athletes	431	967	136	6.56	954	980	299	1332
	Nonathletes	2180	913	129	2.76	907	918	216	1269
Ball-throw (m)	Table tennis players	221	12.7	4.3	0.29	12.1	13.2	3.1	24.9
	Soccer players	316	16.2	4.0	0.22	15.8	16.6	5.1	27.5
	Other athletes	432	14.6	4.2	0.20	14.2	15.0	6.0	29.0
	Nonathletes	2208	12.9	3.7	0.08	12.8	13.1	4.5	28.1

SD standard deviation, *SE* standard error, *CI* confidence interval, *LL* lower limit, *UL* upper limit

remaining 10% of cases (approximately eight cases) that were not used for training and testing the network. This specific type of hold-out strategy was repeated ten times, so that each of the 80 cases should, at least once, be among the left-out athletes finally classified. For each of these ten trials, a new network structure was identified, and independent classification results were produced. To include the remaining 160 cases of low performers in the study, this procedure was carried out on five different randomly selected groups (80 participants each) so that, in the end, about 50 different network structures and thus also 50 different classification results were available. To quantify the validity of this talent identification strategy, the percentage of correct hits of the neural network classifications was averaged over the 50 trials, and the mean value was then used. The classification qualities of both methods (discriminant analysis and MLP) were expressed by the proportion of correct hits, which refers to the percentage of athletes who were correctly assigned as true positives to the group of high performers. An athlete was defined as a false positive if he was classified as a talent but did not reach the best-performing groups. On the basis of the ratio of correct to erroneous hits, the prognostic validity of the test results was calculated.

Results

Test performance results for table tennis players, soccer players, other athletes, and nonathletes

Comparing the test results of later table tennis players, soccer players, other athletes, and nonathletes, it is noteworthy that the tested table tennis players could not deliver outstanding results in any of the tests. It should be noted, however, that this group was the youngest of all the groups tested. Regarding the anthropometric tests, there are no major differences in the results. The descriptive performance characteristics of the FMC participants are presented in **Table 2**.

The information from the table is not sufficient to elucidate the individual sports' different characteristics. Differences in gender and test age (55.1% of the tested students are under the age of 95 months) require consideration of the corresponding z-values to obtain a clearer picture. **Fig. 2** demonstrates the early talent screening results for z-standardized predictors. Compared to soccer players, other athletes from 30 different sports, and nonathletes, the table tennis players reached the best results in sideward jumping and push-ups. The mean z-values of these two predictors show that table tennis players belong to the best 15% of all tested athletes in these tests. In the sit-up test, the table tennis players reached the lowest value of all tested predictors and were just as good as the nonathletes. Overall, this shows that the table tennis players are weaker in athletic abilities such as endurance and strength compared to the much better soccer players and even to the average of the other athletes. Only in the coordination skills (which also include push-ups) do they show obviously better test results than the other groups.

Prediction of general motor tests

Comparison of different sport-specific competition levels

The comparison of the four groups of table tennis players (**Fig. 3**) shows that the highest group representing the national level scores best in five of the nine

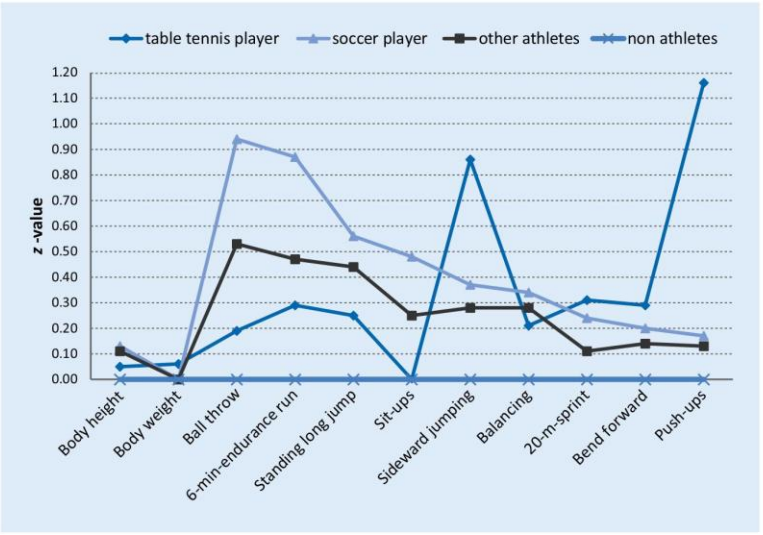

Fig. 2 ▲ Fulda Movement Check test performances (*z*-values) for later table tennis players, soccer players, other athletes, and nonathletes held in second grade at elementary level

tests. This performance increase could not be detected only in balancing backward, sprint, standing long jump, and endurance run. Within the groups of table tennis players, the ANOVA demonstrated significant differences ($p < 0.05$) in sit-ups[9], standing long jump[10], push-ups[11], and endurance run[12]. In push-ups, the post-hoc test (Bonferroni; $p < 0.05$) pointed out that there are significant group differences between group four (national level) and all other groups. In the endurance run, the Dunnett T3 test (Levene; $p = 0.41$) also shows significant differences between the lower and higher groups (levels: 1–3; 1–4; 2–3; 2–4). The results for sideward jumping[13] and TTRS[14] showed a tendency to be significant. As the anthropometric characteristics body height and weight did not relate significantly to the later table tennis performance, those data were not presented in the figure.

[9] $F(3;221) = 2.86; p = 0.038$.

[10] $F(3;221) = 2.83; p = 0.021$.

[11] $F(3;221) = 5.20; p = 0.006$.

[12] $F(3;220) = 5.71; p = 0.001$.

[13] $F(3;221) = 2.51; p = 0.058$.

[14] $F(3;221) = 2.45; p = 0.065$.

Odds ratios for each test and the table tennis recommendation score

ORs can be an effective way of discovering which tests provide a good prediction of young athletes' future success. The purpose of this study is to use the ORs to determine whether an athlete with a good test score in one of the nine initial tests has a higher chance of reaching the national level (level 4) in adolescence compared to an athlete with a lower test score. For each test, we looked at the chances of reaching the national level at the threshold of a z-value of over 1.0. For example, considering the ball-throwing test, one out of five students with a z-score greater than 1.0 reached competition level 4, whereas only one of eleven of the students with a z-score less than 1.0 played in this league. The chance for a good ball-throwing student to reach the national level is therefore more than twice as high ($OR = 2.4$) as the chance of a less talented student. In **Fig. 4**, the ORs for the nine single tests and the TTRSs for those children belonging to the best 16% in this single predictor, which corresponds to a *z*-value of more than 1.0, were calculated. Thus, the most meaningful test is the push-up test. It shows the highest score with a value of $OR = 3.8$, whereas (as also shown in **Fig. 3**) the two predicators sit-ups and balancing backward show values of less than 1.0, which indicates a nonexistent or even negative effect on future success.

Looking at the distribution of students within the national-level 4 group, it can be seen that 82.6% of the athletes have a z-score greater than 1.0 in push-ups (**Fig. 5**). Furthermore, in the sideward jumping test, the percentage of high performers at level 4 is very high (78.3%). From a statistical point of view, the chance of any player reaching a z-value above 1.0 in a test should be about 16%. In this study, the percentages of the national-level players exceed this number in almost all predictors: push-ups (82.6%), sideward jumping (78.3%), ball-throw (39.1%), standing long jump and endurance run (30.4%), bend forward (22.7%), and balancing (17.4%). An exception is the sprint and sit-ups tests, where the percentage is just 8.7% and well below the statistical share.

Specificity and sensitivity

In the context of talent orientation, it is of great interest to know which threshold in the TTRS most likely separates the talents (e.g., national level 4) from the nontalents (e.g., levels 1–3). Depending on the selected limit, the sensitivity and specificity of the TTRS can be determined. In **Fig. 6**, the respective sensitivities and specificities (U13–U17) for the TTRS score are documented. If, for example, a threshold value of 80% is selected for the U13–U17 players in the TTRS, 16 of the 23 players at the national level can be correctly identified (sensitivity is 69.7%). With the same limit value, 101 of the 202 later low performers can also be classified correctly (specificity is 50%). Based on these results, the threshold at which the aim of the sports recommendation is done best can be judged. For example, the largest sum of the percentages of both values is 121.4% (78.3% versus 43.1%), which is found at a TTRS threshold of 77.5%. For comparison, if every second child is generally recommended as a talent, a 50%–50% result should be obtained. The threshold value found is therefore 21.4% more accurate than a random result. A balanced value

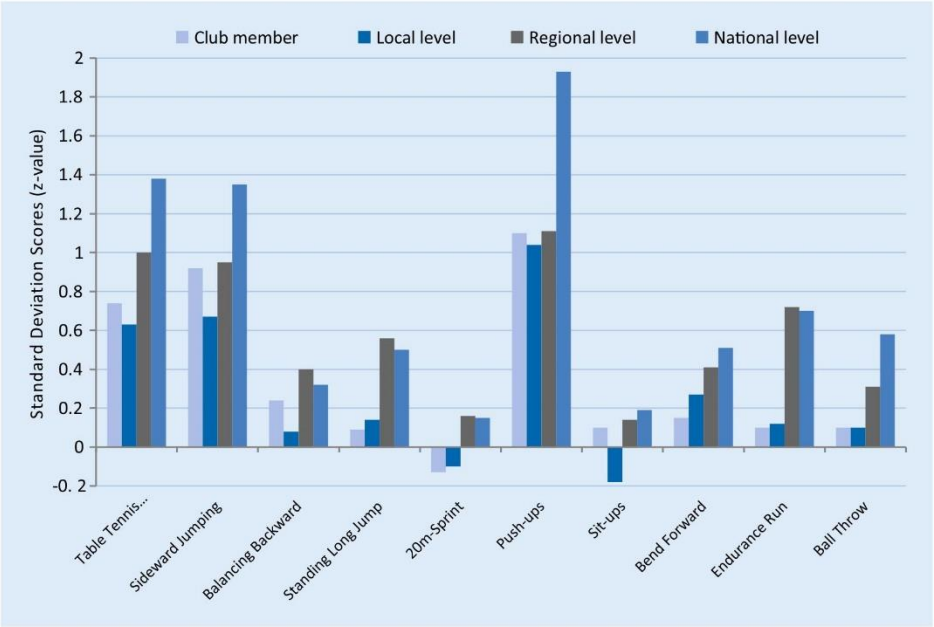

Fig. 3 ◀ Prognostic validity of the test performances of the Fulda Movement Check participants in regard to the later competition performance levels at adolescence (U13–U17). *Table Tennis …* stands for the table tennis recommendation score. Data on the 20-m sprint were transformed in a positive direction

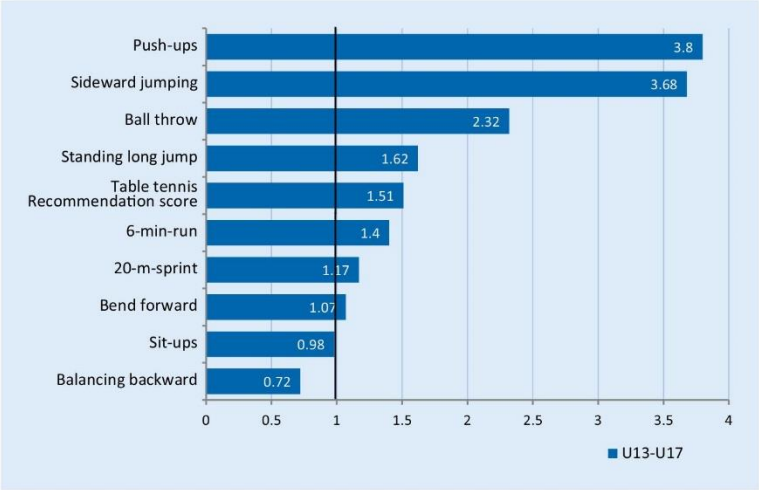

Fig. 4 ▲ Odds ratios (ORs) of the nine Fulda Movement Check (FMC) general motor tests (at the $z \geq 1.0$ and $PR \geq 84$ threshold) and the more specific recommendation score predict the relative chances of the U9 FMC participants reaching at least the national level (level 4) of table tennis competition performance by their competition age (U13–U17)

of both percentages (59% versus 59%) can be found at a TTRS limit of 81.6%.

A cross-validated discriminant analysis can be used to determine a "true" prognosis of future success. Based on the six test values composing the TTRS (sideward jumping, push-ups, bend forward, standing long jump, ball-throw, and balancing backward), 68.2% of the later top performers (level 4) could be identified at the age of about 8 years (**Fig. 7**; $n = 219$[15]). The overall prognosis leads to a percentage of 69.9%. The best test in this regard was the sideward jumping, correctly identifying 69.6% of the future national-level players in advance. However, only 57.4% of the subsequently weaker athletes could be classified correctly by this single test. **Fig. 6** shows the respective totals of the cross-validated discriminant analysis. Other test combinations than the six above-mentioned tests led to partly weaker overall forecasts and fluctuated between a total of 74.8%–50.7% correct hits. Thus, the discriminant analysis provided the best overall result with a combination of seven test values: sprint, balancing backward, sideward jumping, bend forward, endurance run, push-ups, and sit-ups with 74.8% (75% specificity and 72.7% sensitivity).

If an artificial neural network (Multilayer Perceptron; SPSS 25.0) was used, the cross-validated procedure led to a much higher prediction of 78.4% correct hits. The better talent forecast of this nonlinear mathematical tool is made clear particularly in the talented players group. Here, the MLP significantly increased the number of correct high-performance predictions (86.2% true positives). The number of correct predictions of the table tennis players (67.9% true negative predictions) is comparable with the result of the linear discriminant analysis.

[15] Six children could not perform at least one test, so they were excluded from this analysis.

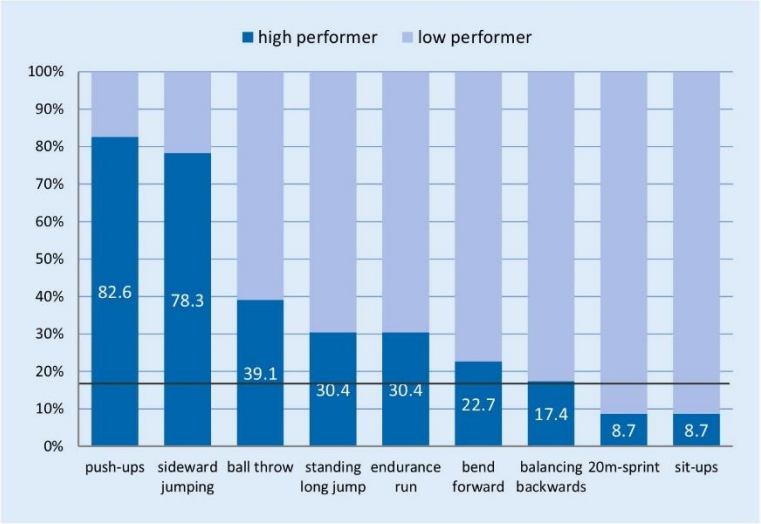

Fig. 5 ▲ Percentage of high performers (z-value ≥1) and low performers (z-value <1) in the nine single tests within the group (U13–U17) of the national level (level 4) table tennis players (the *line* indicates the statistical average)

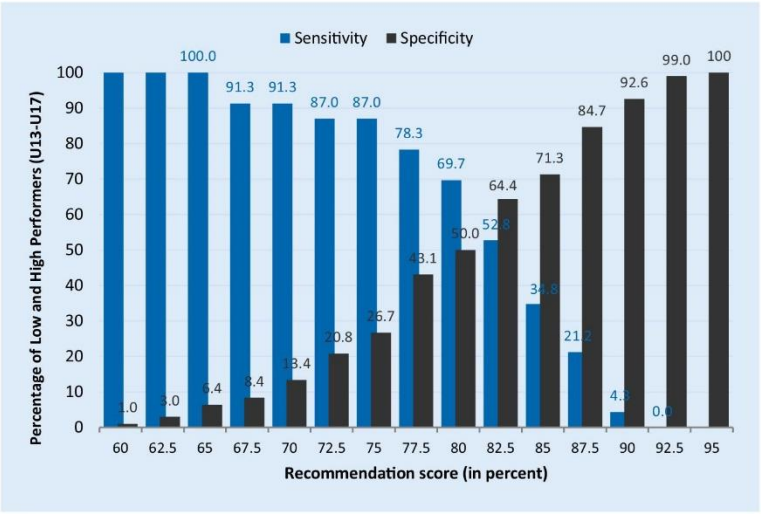

Fig. 6 ▲ Specificity and sensitivity of Fulda Movement Check participants' table tennis recommendation scores to predict future table tennis performance at national levels (level 4) in adolescence (U13–U17)

Discussion

Although table tennis is a very complex and coordinative sport (Robertson et al., 2018) that requires visual perception, control of arm and head movements, as well as ball control and visual perception (Faber et al., 2014), it turns out that certain skills, which can be assessed in a general motor test set-up, accurately represent the talent make-up of a table tennis player. If one compares the initial performance of the later table tennis players with that of football players (see also Hohmann et al., 2018), it is notable that table tennis players perform weaker in the tests with high athletic demands (e.g., endurance and strength). Robertson et al. (2018) showed similar results in studies between tennis, badminton, and table tennis players. With the exception of four test results, the average of all performance prerequisites was always lower than that of the footballers and sometimes even lower than the scores of the other athletes. It is generally known that a high level of athleticism and the associated strength training are related to muscular shortening. So, it can be assumed that the better mobility of table tennis players is due to the low proportion of strength training in this sport. This assumption is reinforced by the low mean value in the sit-up test. In the field of sit-up performance, the table tennis players showed the weakest values of all tested athletes and only equaled the "non-sport" participants. If the highest youth national level is considered, in our study, good core muscle strength was associated with a negative outcome in the later performance development (ORs).

The anthropometric values (height and weight) did not differ significantly from the values of other athletes. These findings partly correspond to those of Robertson et al. (2018). In test tasks primarily related to general and gross-motor body coordination, however, the later table tennis players achieved very good scores. This included sideward jumping (Nikolic et al., 2014) as well as push-ups. The push-up test is not fundamentally dependent on arm and upper body strength due to the still good weight–force ratio of the participating children and the complex upper-body coordination required in this test (see Bös & Schlenker, 2016). Rather, there seems to be an upper limitation in the coordinative speed the children exhibit. Surveys of table tennis players showed that many top athletes find learning coordinative movements easy: "I … just did it and it was correct" (Faber et al., 2015, p. 9). The push-up test could perhaps represent this fast learning of coordinative movements. The test tasks push-ups and sideward jumping led to significant differences between the sports groups on the one hand and the table tennis performance groups on the other hand. Also, it was shown at the national level that mainly athletes (more than 75%) with a high z-value ($z \geq 1.0$) have reached this performance level.

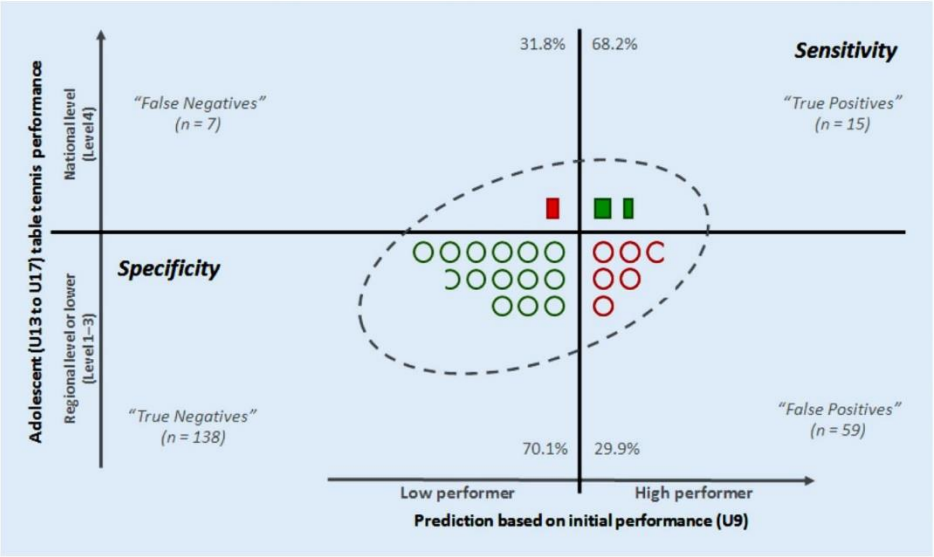

Fig. 7 ◀ Discriminant analysis to predict the later U13–U17 table tennis national performance group based on initial Fulda Movement Check performances (each *full symbol* represents 10 children)

It is not surprising that the push-up test showed the highest OR values overall. The probability that one of the best 16% of the participants in this test will reach the national level is 3.8 times higher than for the test participants with a lower test value (z-value <1.0). Other studies with elite table tennis athletes have shown similar results (Mähner, 2018). It is astonishing, however, that the push-up test in the older subgroup of U15–U17 did not provide comparable results and only came second in the ORs ($OR_{push\text{-}up} = 2.18$). Nevertheless, this can be explained by the fact that, in all four performance groups, the z-value is on average higher than $z = 1.0$. The distinction between a high-performance group (z-value ≥1) and a low-performance group (z-value <1) could not cover these differences, since almost all groups performed above that limit value. The situation would be different with a z-value limit of for example $z \geq 1.3$. The differentiation now would be much stronger, but this feature must be analyzed more thoroughly on the basis of a greater sample size in further studies. Overall, however, this study found that a higher performance level also increases the share of the best performers in a particular test. Only four tests were identified where higher athletic performances did not have a statistically significant impact on later table tennis performance (U13–U17). In addition to the previously mentioned sit-up values, the bend forward, sprint, and balancing backward performances also did not influence later table tennis success. As the U13 sample investigated in this study was much younger than the much older adolescent table tennis athletes investigated by Hohmann et al. (2015; see also Pion et al., 2015), the different relevance of the change in performance in these four motor tests should be investigated more deeply in future studies (Mähner, 2018).

A sport-specific RS is a useful way to better reflect table tennis performance. In this test, too, the TTRS at least distinguished between top performers and hobby athletes. The probability of reaching the national level for athletes with a high TTRS, however, was comparatively low ($OR_{TTRS} = 1.51$). The previous TTRS calculation, consisting of sideward jumping, balancing backward, push-ups, standing long jump, ball throw, and bend forward, could not be confirmed in the discriminant analysis. Here, the best results (74.8%) were calculated by the seven predictors: sprint, balancing backward, sideward jumping, bend forward, endurance run, push-ups, and sit-ups. While in the post-hoc test the three variables sit-ups, push-ups, and endurance run significantly separated the performance groups, the sideward jumping values in the ORs showed better quality ($OR_{sideward_jumping} = 3.68$). This goes hand in hand with the results of Doherty et al. (2018). In spite of the potentially expandable recommendation calculation, the previous set-up in sensitivity and specificity led to a good overview in a wide range of border areas.

Conclusion

Prognostic validity studies on talent screening and orientation campaigns, such as this one, build on the evolving long-term performance development of the tested participants. In addition to the physiological prerequisites (Kondrič, Furjan-Mandić, Kondrič, & Gabaglio, 2010), many studies (Lopez & Santelices, 2012; Chu, Chen, Chen, Huang, & Hung, 2012; Rodrigues, Vickers, & Williams, 2002) focus on mental, cognitive, and perceptual abilities. These studies indirectly demand that talent orientation campaigns must also look at the psychological features of the table tennis sport. Irrespective of such new developments, the present study has already shown that only with the help of test-based and long-term talent test campaigns, such as the Flemish Sport Compass or the FMC, prognostically valid statements can be made about the future success of sports talents. The results of this study con-

tribute to the table tennis talent search and orientation, as they explicitly show the characteristic abilities of successful youth table tennis players at an early age and thus offer a valid strategy to find and identify later talents. For more precise statements in the recommendations, however, it is advisable to integrate further perceptuo-motor tests into the test battery (Faber et al., 2017a) to fairly depict such a complex sport. From this point of view, it may also be necessary to add advanced statistical methods such as neural networks or other nonlinear tools to current talent identification procedures (Pion et al., 2016).

Corresponding address

Maximilian Siener
Institute of Sports Science,
University of Bayreuth
Universitätsstraße 30,
95440 Bayreuth, Germany
Maximilian.Siener@
uni-bayreuth.de

Acknowledgements. The authors would also like to thank the authorities of the city of Fulda, the Fulda district authorities, and the participating elementary schools for supporting the Fulda Movement campaign since 2010.

Funding. This study is part of the research project "talent identification," which was supported by the German Institute of Sport Science (Bonn, Germany).

Compliance with ethical guidelines

Conflict of interest M. Siener and A. Hohmann declare that they have no competing interests.

For this article no studies with human participants or animals were performed by any of the authors. All studies performed were in accordance with the ethical standards indicated in each case.

References

Akpinar, S., Devrilmez, E., & Kirazci, S. (2012). Coincidence-Anticipation timing requirements are different in racket sports. *Perceptual and Motor Skills: Exercise & Sport,*, *115*(2), 1–13.

Bös, K., & Schlenker, L. (2016). Der Deutsche Motorik-Test6–18 (2nd edn.). Hamburg: Feldhaus.

Carling, C., & Collins, D. (2014). Comment on "football-specific fitness testing: adding value or confirming the evidence?". *Journal of Sports Science*, *32*(13), 1206–1208.

Carling, C., Le Gall, F., & Malina, R. M. (2012). Body size. skeletal maturity. and functional characteristics of elite academy soccer players on entry between 1992 and 2003. *Journal of Sports Sciences*, *30*(15), 1683–1693. https://doi.org/10.1080/02640414.2011.637950.

Chu, C. Y., Chen, I. T., Chen, L. C., Huang, C. J., & Hung, T. M. (2012). Sources of psychological states related to peak performance in elite table tennis players. *International Journal of Table Tennis Sciences*, *7*, 86–90.

Doherty, S. A. P., Martinent, G., Martindale, A., & Faber, I. R. (2018). Determinants for table tennis performance in elite scottish youth players using a multidimensional approach: a pilot study. *High Ability Studies*, *29*, 241–254. https://doi.org/10.1080/13598139.2018.1496069.

Faber, I. R., Oosterveld, F. G. J., & Nijhuis-Van der Sanden, M. W. (2014). Does an eye-hand coordination test have added value as part of talent identification in table tennis? A validity and reproducibility study. *PLoS one*. https://doi.org/10.1371/journal.pone.0182211.

Faber, I. R., Oosterveld, F. G., Van den Heuvel, S., Bustin, P., Elferink-Gemser, M., & Nijhuis-Van der Sanden, M. W. (2015). *High potential in table tennis from the perspectives of elite players and their youth trainers: an explorative qualitative study*. 5th World Racket Sport Congress, China.

Faber, I. R., Bustin, P. M., Oosterveld, F. G., Elferink-Gemser, M. T., & Nijhuis-Van der Sanden, M. W. (2016). Assessing personal talent determinants in young racquet sport players: a systematic review. *Journal of Sports Sciences*, *34*(5), 395–410. https://doi.org/10.1080/02640414.2015.1061201.

Faber, I. R., Elferink-Gemser, M. T., Oosterveld, F. G., Twisk, J. W., & Nijhuis-Van der Sanden, M. W. (2017a). Can an early perceptuo-motor skills assessment predict future performance in youth table tennis players?: An observational study (1998–2013). *Journal of sports sciences*, *35*(6), 593–601.

Faber, I. R., Pion, J., Munivrana, G., Faber, N. R., & Nijhuis-Van der Sanden, M. W. (2017b). Does a perceptuomotor skills assessment have added value to detect talent for table tennis in primary school children? *Journal of Sports Science*, *6*, 593–601.

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

Fuchslocher, J., Romann, M., Rüdisüli, L. R., Birrer, D., & Hollenstein, C. (2011). Das Talentselektionsinstrument PISTE: wie die Schweiz Nachwuchsathleten auswählt. *Leistungssport*, *41*(4), 22–27.

Le Gall, F., Carling, C., Williams, M., & Reilly, T. (2010). Anthropometric and fitness characteristics of international. professional and amateur male graduate soccer players from an elite youth academy. *Journal of Science and Medicine in Sport*, *13*(1), 90–95.

Golle, K., Muehlbauer, T., Wick, D., & Granacher, U. (2015). Physical fitness percentiles of German children aged 9–12 years: findings from a longitudinal study. *PLoS ONE*, *10*(11), e142393. https://doi.org/10.1371/journal.pone.0142393.

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. https://doi.org/10.1080/02640414.2012.713980.

Hohmann, A., & Seidel, I. (2003). Scientific aspects of talent development. *International Journal of Physical Education*, *40*(1), 9–20.

Hohmann, A., Fehr, U., Siener, M., & Hochstein, S. (2017a). Talentscreening und Talentorientierung. *Leistungssport*, *47*(4), 11–13.

Hohmann, A., Fehr, U., Siener, M., & Hochstein, S. (2017b). Validity of early talent screening and talent orientation. In P. Platen, A. Ferrauti, E. Grimminger-Seidensticker & T. Jaitner (Eds.), *Sport science in a metropolitan area* (p. 590). Bochum: University Press.

Hohmann, A., Fehr, U., & Voigt, L. (2015). Heute im Talentpool – In Hamburg auf dem Podium. *Leistungssport*, *45*(5), 5–11.

Hohmann, A., Siener, M., & He, R. (2018). Prognostic validity of talent orientation in soccer. *German Journal of Exercise and Sport Research*, *48*, 478–488. https://doi.org/10.1007/s12662-018-0549-5.

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.

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*. https://doi.org/10.1371/journal.pone.0182211.

International Table Tennis Federation (2018, November 20th). Current WR – men. http://results.ittf.link/index.php?option=com_fabrik&view=list&listid=69&Itemid=206. Accessed: 20 Nov 2018

Kondric, M., Furjan-Mandić, G., Kondrič, L., & Gabaglio, A. (2010). Physiological demands and testing in table tennis. *International Journal of Table Tennis Science*, *6*, 165–171.

Kondric, M., Zagatto, A. M., & Sekulic, D. (2013). The physiological demands of table tennis: a review. *Journal of Sports Science & Medicine*, *12*(3), 362–370.

Lopez, A., & Santelices, O. (2012). Personality characteristics of elite table tennis athletes of the Philippines: basis for a proposed recruitment program. *International Journal of Table Tennis Sciences*, *7*, 1–4.

Mähner, T. (2018). Talentdiagnose im Tischtennis. Zur Validität sportmotorischer Tests in der Talentsichtung & Talentselektion. Unpublished manuscript, Lehrstuhl für Trainings- und Bewegungswissenschaft, Universität Bayreuth, Bayreuth.

Meylan, C., Cronin, J., Oliver, J., & Hughes, M. (2010). Reviews: 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.

Müller, L., Müller, E., Kornexl, E., & Raschner, C. (2015). The relationship between physical motor skills, gender and relative age effects in young Austrian alpine ski racers. *International Journal of Sports Science & Coaching*, *10*(1), 69–85. https://doi.org/10.1260/1747-9541.10.1.69.

Munivrana, G., Furjan-Mandic, G., & Kondrič, M. (2015a). Determining the structure and evaluating the role of technical-tactical elements in basic table tennis playing systems. *International Journal of Sports Science & Coaching*, *10*, 111–132.

Munivrana, G., Petrinović, L., & Kondrič, M. (2015b). Structural analysis of technical-tactical elements

in table tennis and their role in different playing zones. *Journal of Human Kinetics*, *47*(1), 197–214. https://doi.org/10.1515/hukin-2015-0076.

Nikolic, I., Furjan-Mandic, G., & Kondric, M. (2014). The relationship of morphology and motor abilities to specific table tennis tasks in youngsters. *Collegium Antropologicum*, *38*(1), 241–245.

Papic, V., Rogulj, N., & Plestina, V. (2009). Identification of sport talents using a web-oriented expert system with a fuzzy module. *Expert Systems with Applications*, *36*, 8830–8838. https://doi.org/10.1016/j.eswa.2008.11.031.

Pion, J. (2015). *The Flemish Sports Compass: from sports orientation to elite performance prediction*. Ghent: University Press.

Pion, J. (2017). *Sustainable investment in sports talent: the path to the podium through the school and the sports club*. Arnhem: HAN University of Applied Sciences Press.

Pion, J., Hohmann, A., Liu, T., Vandorpe, B., Lenoir, M., & Segers, V. (2016). Predictive models reduce talent development costs in female gymnastics. *Journal of Sports Sciences*, *35*(8), 806–811.

Pion, J., Segers, V., Fransen, J., Debuyck, G., Deprez, D., Haerens, L., & Lenoir, M. (2015). Generic anthropometric and performance characteristics among elite adolescent boys in nine different sports. *European Journal of Sport Science*, *15*(5), 357–366. https://doi.org/10.1080/17461391.

Robertson, K., Pion, J., Mostaert, M., Norjali Wazir, M. R. W., Kramer, T., Faber, I. R., Vansteenkiste, P., & Lenoir, M. (2018). A coaches' perspective on the contribution of anthropometry, physical performance, and motor coordination in racquet sports. *Journal of Sports Sciences*, *36*, 2706–2715. https://doi.org/10.1080/02640414.2018.1441941.

Rodrigues, S. T., Vickers, J. N., & Williams, A. M. (2002). Head, eye and arm coordination in table tennis. *Journal of Sports Sciences*, *20*, 187–200.

Roescher, C., Elferink-Gemser, M., Huijgen, B., & Visscher, C. (2010). Soccer endurance development in professionals. *International Journal of Sports Medicine*, *31*, 174–179. https://doi.org/10.1055/s-0029-1243254.

Stemper, T., Bachmann, C., Diehlmann, K., & Kemper, B. (2009). Das Düsseldorfer Modell der Bewegungs-, Sport- und Talentförderung (DüMo). In Bundesinstitut für Sportwissenschaft (Ed.), *Talentdiagnose und Talentprognose*. 2. BISp-Symposium: Theorie trifft Praxis. (pp. 139–142). Köln: Strauss.

Willimczik, K. (1982). Determinanten der sportmotorischen Leistungsfähigkeit im Kindesalter. In H. Howald & E. Hahn (Eds.), *Kinder im Leistungssport* (pp. 141–153). Stuttgart: Birkhäuser.

Zuber, C., Zibung, M., & 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. https://doi.org/10.3389/fpsyg.2016.01088.

Artikel 4

Hohmann*, A., & Siener, M. (2021). Talent Identification in Youth Soccer: Prognosis of U17 Soccer Performance on the Basis of General Athleticism and Talent Promotion Interventions in Second-Grade Children. *Frontiers in Sports and Active Living, 3*, 625645. https://doi.org/10.3389/fspor.2021.625645

frontiers in Sports and Active Living

ORIGINAL RESEARCH
published: 04 June 2021
doi: 10.3389/fspor.2021.625645

Talent Identification in Youth Soccer: Prognosis of U17 Soccer Performance on the Basis of General Athleticism and Talent Promotion Interventions in Second-Grade Children

Andreas Hohmann and Maximilian Siener*

Department of Training and Movement Science, Institute of Sport Science, University of Bayreuth, Bayreuth, Germany

Several talent identification programs in elementary school have implemented motor diagnostics to introduce children to groups of sports, like game sports, or even to particular sports like soccer. However, as in most other sports, in youth soccer, the predictive value of such early testing is still unclear. This prospective study evaluated the midterm prognostic validity of generic motor performance tests. The sample consisted of male second-grade children, which had received a recommendation to participate in soccer. The talent screening campaign was a basic check comprising two anthropometric parameters, five physical fitness, and three motor competence diagnostics of the German Motor Test 6–18. The test data were collected from the participating elementary school classes of the years 2010 to 2014. The soccer competition performance of those children having completed the age of at least 15 years (n = 502) up to the end of the season 2019/2020 (2020, September 30) was recorded. This group of U17 players was then assigned individually to five different competition levels. The prognostic validity of the physical and physiological tests was determined using ANOVAs, odds ratios, and a regression path analysis. All diagnostic methods exhibited medium-to-high prognostic validity over the 8 year time span from the talent screening to the later soccer competitions in the adolescent age groups. For later success in soccer on the province level, the 6-min run (OR = 4.28), dynamic balance (OR = 4.04), and 20-m sprint (OR = 2.46), as well as the participation in the training center of the German Soccer Federation (OR = 5.67) and the diversity of club sport activities (OR = 3.56), were of particular importance.

Keywords: talent identification, talent promotion, youth soccer, child athlete, prospective study, longitudinal design, path analysis, odds ratio

OPEN ACCESS

Edited by:
Bruno Gonçalves,
University of Evora, Portugal

Reviewed by:
Oliver Höner,
University of Tübingen, Germany
Matthieu E. M. Lenoir,
Ghent University, Belgium

****Correspondence:***
Andreas Hohmann
andreas.hohmann@uni-bayreuth.de

Specialty section:
This article was submitted to Elite Sports and Performance Enhancement, a section of the journal Frontiers in Sports and Active Living

Received: *03 November 2020*
Accepted: *11 February 2021*
Published: *04 June 2021*

Citation:
Hohmann A and Siener M (2021) Talent Identification in Youth Soccer: Prognosis of U17 Soccer Performance on the Basis of General Athleticism and Talent Promotion Interventions in Second-Grade Children. Front. Sports Act. Living 3:625645. doi: 10.3389/fspor.2021.625645

INTRODUCTION

Long-term talent development programs combine a diagnostic talent identification process with a sustainable talent promotion strategy (see **Figure 1**). On the diagnostic side of such a talent development model, early *talent orientation* includes early talent testing measures as well as recommendations for specific sports that match best with the individual ability profile.

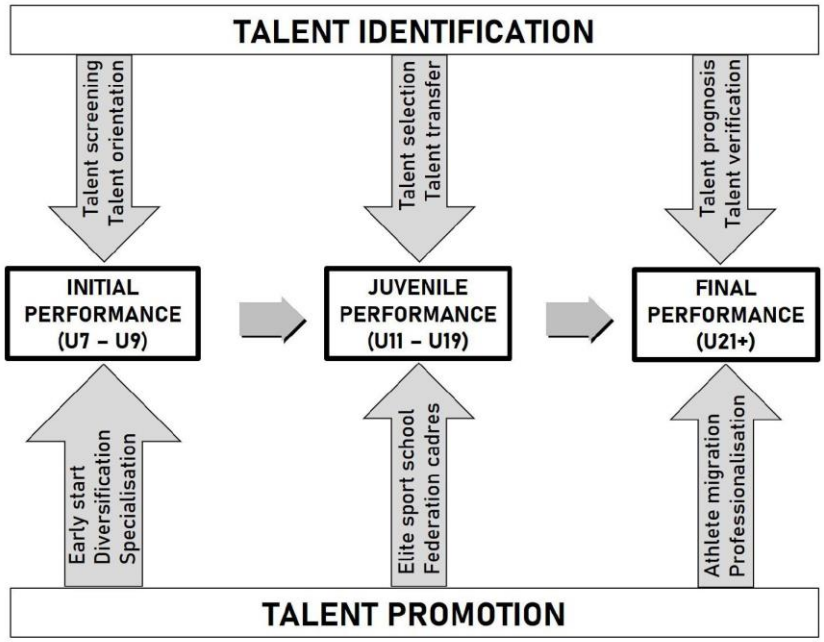

FIGURE 1 | Talent identification and talent promotion as the two intertwined strategies in the long-term talent development model (mod. from Hohmann and Seidel, 2003).

The term talent orientation is related to early talent detection in preferably untrained and thus still heterogeneous samples and "aims at motivating youngsters to choose a sport that matches the individual talent characteristics" (Pion, 2015, p. 22). Following this idea, several talent screening programs in elementary school have implemented motor diagnostics (e.g., Fuchslocher et al., 2011; Golle et al., 2015; Pion, 2015) to orientate the best movers according to the strengths of their individual profile of the general motor giftedness into specific sports where they can transform their physical, physiological, and psychological gifts through a long-term process of diligent learning, deliberate practice, and an extended amount of high-quality training into ultimate achievement (Davids and Baker, 2007; Pion, 2015). The better the individual talent characteristic profile fits to the (future) soccer demands, the higher the chances that the soccer beginners will achieve success and satisfaction in this complex team sport. This assumption is underlined by Suppiah et al. (2015), who state that a wrong choice can never be compensated by training. Engaging in an unsuitable sport might not only be detrimental to fun but also lead to dropout ahead of time. Furthermore, soccer is one of the sports that require early talent orientation (Papic et al., 2009), as on the one hand, a high amount of learning time is needed for the specific technical skills, and on the other hand, the athletes competing on the highest level got younger over the last decade. Thus, besides a systematic long-term athletic development, early talent orientation is of great relevance. In addition to that, an early talent orientation and soccer education starting at the elementary-school age gives the coaches a longer observation period and in this way reduces selection errors during early adolescence. Although most of the soccer federations administer their talent selection campaigns at a later point of time, when youth athletes have already trained for some years and developed into more homogeneous samples (Pion, 2015), early talent orientation at elementary-school age could contribute positively to the ongoing debate in talent research. In this debate, some academics warn against talent selection procedures that are conducted too early (Meylan et al., 2010), whereas others acknowledge that these selections are worthwhile to help the sport federations to focus their resources on the most talented young athletes (Unnithan et al., 2012; Zuber et al., 2015, 2016; Hoener and Votteler, 2016). The latter is especially important in soccer, as this very professional sport attracts the majority of children already at elementary-school age. Especially, if the final step of the scientific verification of the former talent prognosis was included in the talent development program of the federations, an early talent orientation could lead to a more successful talent development model in soccer.

In general, there is a lack of research investigating the prognostic value of different performance prerequisites over the full long-term period from child to adult training (Johnston et al., 2018; Sarmento et al., 2018; Williams et al., 2020). Most of the research concentrated on the middle stages of the juvenile performance development. Although previous studies questioned juvenile success even at the advanced stage of late adolescence as an appropriate indicator for soccer success in adulthood (Guellich, 2014), from recent studies on the early part

of the juvenile training stage between early and late adolescence, there is compelling evidence that early talent orientation is a fruitful endeavor. So in the age group of under-15 (U15) soccer participants, various authors (Le Gall et al., 2010; Carling et al., 2012; Deprez et al., 2015; Hoener and Votteler, 2016; Hohmann et al., 2018; Sieghartsleitner et al., 2018) investigated prognostic periods of at least 3 years. Doing so, Hoener and Votteler (2016; see also Hoener et al., 2017) could show that even on the homogeneous level of the *German soccer competence centers*, a soccer-specific test battery (German Soccer Federation, 2009) provides prognostic valid and also practically worthwhile information about the 7.2 times better odds of the fastest and technically best third of the preselected players to reach the U15 junior national team. Including psychological characteristics, Zuber et al. (2016) showed that 12- to 15- year-old individuals having started early with soccer training had 2.5 times better chances to reach the Swiss U17 national team than players starting later in time with soccer-specific deliberate training.

Although reliable and valid information about the future potential of talented players on the basis of motor abilities and technical skills diagnostics is a valuable tool in talent development programs for sport clubs and federations, several studies question the long-term predictability of future success (Cote et al., 2009; Pankhurst and Collins, 2013; Carling and Collins, 2014). The main reasons for these scientific concerns arise from the often undifferentiated mixture of generic as well as sport-specific tests in talent identification campaigns, poor operationalization of the criterion variables, and the unsystematic timing of cross-sectional diagnostics at single points in time during the long-term athletic development process (Bergkamp et al., 2019). It thus comes as no surprise that the great variety of study design parameters have led to inconsistent research results, providing an ambiguous picture with regard to the prognostic validity of physical and physiological tests addressing generic motor abilities and sport-specific technical skills. Especially in the adolescent U15 age group, some studies verified a medium stability (Deprez et al., 2015) and prognostic validity of physical and physiological tests (Figueiredo et al., 2009; Hoener and Votteler, 2016; Sarmento et al., 2018), whereas others did not find significant associations between test results and later success in youth, junior, or adult soccer (Le Gall et al., 2010; Carling et al., 2012). Recent talent research in soccer by Hoener and Votteler (2016; see also Hoener et al., 2017) and Hohmann et al. (2018) has made clear that a longitudinal investigation of motor predictors' prognostic relevance for long-term success is not only a key topic in talent research (see also Gonaus and Mueller, 2012; Sarmento et al., 2018; Bergkamp et al., 2019) but also an indispensable prerequisite for a sophisticated understanding of the changing prognostic validity of the various diagnostics at the different stages of the non-linear and multifactorial long-term development of elite soccer players (Skorski et al., 2016).

As soccer is based on a complex, multidimensional performance profile (Williams and Franks, 1998; Reilly et al., 2000; Williams and Reilly, 2000; Huijgen et al., 2014; Buekers et al., 2015; Forsman et al., 2016b; Johnston et al., 2018, the early talent screening is by definition focused on a multifaceted variety of general physical, physiological, and psychological performance diagnostics in heterogeneous populations. With this in mind, our study focused on general anthropometric, physical fitness (PF), and motor competence (MC) characteristics. PF can be defined as the overall performance in various strength, speed, and endurance tasks in a specified physical, social, and psychological environment (Demetriou et al., 2019). MC summarizes the degree of proficiency in a wide range of motor tasks as well as the movement quality, coordination, and control leading to a particular motor performance (Bardid et al., 2019). The generality–specificity dilemma says that the testing of soccer-specific skills is not yet suitable for beginners like the second-grade elementary-school children at hand. The reason is that even the best movers do not yet possess the high-fidelity skills of dribbling, passing, or shooting (Hohmann et al., 2018). Notwithstanding this, Pion et al. (2020) assume that even on the basis of a generic test battery, it is possible not only to identify sporting potential in primary school but also, on the one hand, to orient children toward a sport that best suits their personal strengths or, on the other hand, to facilitate the transfer between similar sports. However, the authors concluded that more research is needed on how to assist children in choosing sports that match their own traits and preferences. Thus, the aim of this long-term study is to examine the prognostic validity of a more fundamental and general testing of early athleticism on the second-class age level, although this cohort included some older scholars who had to repeat the second class in the year 2010, with 93.1% of the vast majority of the test participants representing the U9 age group. So for the sake of easier communication in the following, we adhere to the age category U9 to refer to the soccer-specific age group framework.

The present study is a follow-up version of a previous study of Hohmann et al. (2018) to examine the prognostic validity of a broader performance testing at the same age of second grade of elementary school. In contrast to the former study, which investigated prognostic validity of 11 general performance characteristics and two soccer-specific skill tests (agility run and soccer dribbling) over a midterm follow-up period in regard to the U15 age group, in the current study, the follow-up period was extended to a long-term prognosis until the U17 age group. In this study, the test results of the investigated second-class children in two anthropometric, five PF tests, and three MC tests are compared with their successes as juvenile soccer athletes at age U13 and U17. By including the intermediate U13 stage of the participants' soccer performance, this study allows for a more complex picture of the soccer performance development. Athletes who later show better performances should also be able to be distinguished significantly from later weaker athletes already in the U9 age group. By means of this prospective cohort design, this study is aimed on the extension of the current knowledge on the prognostic validity of very early talent characteristics, as the majority of previous studies focused on the later stages of the U12 and U15 age groups.

MATERIALS AND METHODS

Study Design

According to our model of long-term talent development (see **Figure 1**), the relevance of the performance characteristics profile has to be investigated in a stepwise procedure for at least three

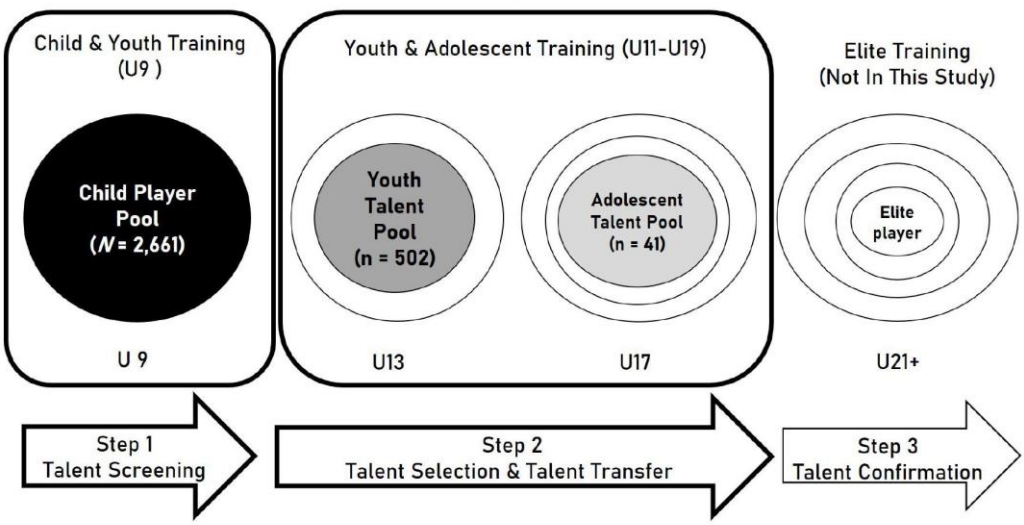

FIGURE 2 | Stepwise talent prediction during the three stages of the talent development process.

prognostic periods from the beginning of a talent development program until the full reach of the professional level as its final destination (**Figure 2**). Each prognosis should allow for midterm soccer performance predictions over a time span of about 5 years (Hoener et al., 2017). As talent development in soccer mostly starts in the U9 age group during elementary-school attendance, the youth training center of the federation starts at age 12 and promotes the talents until age 15. Besides the youth training centers of the soccer federation, the professional soccer clubs run certified soccer academies including players of the same age and also older ones of the age group U17. The final stage of elite adult expertise is reached earliest at about age 21; and in most cases around the age of 23 years, the long-term perspective should cover a development period of about 15 years. This also corresponds to the average time span of 15 years that is needed to reach international soccer excellence (Leite et al., 2009).

The second-grade scholars' physical and physiological test results served as predictors for the participants' later success in soccer at the intermediate point in time of age U13 and finally in late adolescence, which is the U17 age group. The total prognostic time span between the testing and the adolescent competition U17 success was $M = 7.94$ years ($SD = 0.87$).

University staff members and students conducted the testing in all five test campaigns (2010–2014). After the testing, the participation of the children in soccer competitions on any youth soccer division level was recorded from the beginning of the season 2010/2011 (2010, September 30) until the end of the season 2019/2020 (2020, September 30). The data were extracted from the regional print media (tournament and soccer match lists), the official DFB website (www.fussball.de; www.torgranate.de), and the official team lists provided by the soccer clubs.

Before entering the test campaign, all children's parents provided written informed consent for the recording and scientific use of the data collected in the anthropometric and motor tests. The university's ethics department approved the implementation of this study.

Participants

All soccer participants underwent the motor diagnostics at the average age of $M = 8.10$ years ($SD = 0.73$) together with their classmates who did not engage in youth soccer. For the prospective cohort study at hand, we investigated the data from $N = 2{,}661$ male second-grade children with no more than one test data missing. All 2,661 study members took part in the talent screening campaign in the years 2010–2014. Out of this total population, $n = 502$ were found to compete later on in official soccer games independent from the playing division and also reached the age of 15 years (≥180 months) until the deadline of the sport season 2019/2020 (2020, September 30; **Figure 3**). Children who had chosen other competition sports other than soccer were not included in this study to allow for the comparison of the soccer participants with actually untrained children.

Measures

Performance Characteristics

Besides the two physical characteristics body height and body weight, the testing included five physiological fitness, and three MC tests, which were used as performance predictors. The test tasks aimed at the diagnosis of sprint, coordination, balance, flexibility, arm and upper body strength, leg power, ball throw, and endurance performance. The standardization of the test items was secured by protocols, which included a detailed description of test materials and setup, test demonstration,

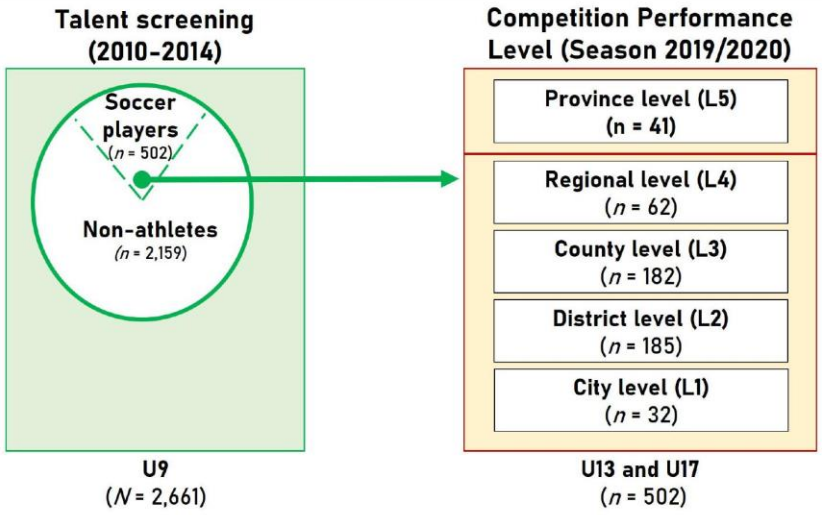

FIGURE 3 | Scheme of the participants' groups of the talent screening campaign (2010–2014) included in this study.

warm-up and training trials, assessment, and registering of the test scores (Boes and Schlenker, 2016). Besides the generic motor tests, each player's body height, body weight, and calendar age (measured by the month of birth within a calendar year) were registered.

20-m Sprint

Time for a 20-m linear running sprint (Boes et al., 2001). The run time was measured by means of light gates (Brower Timing Systems; Draper, USA). The starting position was 0.3 m behind the start line. Between the two possible attempts, a break of at least 2-min was kept. The reliability of this test is $r_{tt} = 0.90$ (Boes and Schlenker, 2016). For easier understanding, in the following, we changed the scale into positive direction for faster run times.

Sideward Jumping

The test contains of 15 s of sideward jumping within two adjacent 50 × 50 cm squares (Kiphardt and Schilling, 1970). The number of two-legged jumps from one square to the other without touching a boundary line was measured. Five trial jumps were allowed before the testing began. Between the two possible attempts, a break of at least 2-min was kept. The objectivity of this test is $r_{obj} = 0.99$, and the reliability is $r_{tt} = 0.89$ (Boes and Schlenker, 2016).

Balancing Backwards

Balance backwards on 6-, 4.5-, and 3-cm wide beams (Kiphardt and Schilling, 1970). For each bar, the number of steps backwards balanced (feet fully raised) until leaving the bar was counted. The maximum number of steps per attempt was limited to eight. For each of the three beams, two possible attempts were made. Thus, the result was the sum of all steps taken (maximum: 48 steps). There was a short practice period before the test was carried out. The objectivity of this test is $r_{obj} = 0.99$, and the reliability $r_{tt} = 0.73$ (Utesch et al., 2015).

Standing Torso Bend Forward

A standing torso bend forward test was performed as a flexibility test (Fetz and Kornexl, 1978). Here, the participants tried to reach as far as possible with their fingertips beyond their feet and to hold this position for at least 3 s. The distance of the fingers in cm to ground level was recorded, whereby a low range aboveground level was recorded as a negative distance. There were two attempts allowed. The objectivity of this test is $r_{obj} = 0.99$, and the reliability $r_{tt} = 0.94$ (Boes and Schlenker, 2016).

Push-Ups

The push-up test was carried out after a short trial period. Within 40 s, the number of fully completed repetitions is counted. Hands have to touch each other when the body is lying down on the floor and also when the arms are extended after the push-up. A complete repetition was only evaluated when the upper body was laid down on the mat and hands touched each other (Haag and Singer, 1981). Only one attempt was carried out here. The objectivity of this test is $r_{obj} = 0.98$, and the reliability $r_{tt} = 0.69$ (Boes and Schlenker, 2016).

Sit-Ups

Similar to the push-up test, the time available for the push-up test was limited to 40 s (Boes et al., 2001). After a short practice phase, only one test was granted, and the number of correctly executed sit-ups was counted. The objectivity of this test is $r_{obj} = 0.92$, and the reliability $r_{tt} = 0.74$ (Klein et al., 2012).

Standing Long Jump

The standing long jump was carried out without a previous tryout. The distance of the two-leg standing jump was measured

in cm (measured from the heel; Fetz and Kornexl, 1978). A break of at least 2-min was observed between the two possible attempts. The objectivity of this test is $r_{obj} = 0.99$, and the reliability $r_{tt} = 0.89$ (Boes and Schlenker, 2016).

6-min Endurance Run

A 6-min endurance run around a volleyball pitch (9 × 18 m) was carried out (Fetz and Kornexl, 1978). There, the number of meters covered was measured. The test was conducted in groups of 15 persons at the same time. The reliability of this test is $r_{tt} = 0.92$ (Boes and Schlenker, 2016).

All players were assessed under similar conditions. The tests were carried out during regular school hours (8–12 pm) by qualified test personnel. The testing always began after a uniform warm-up phase with the 20-m sprint and ended with the 6-min endurance run. In all tests, except for sideward jumping (where the average of the two attempts was taken as test result), the better one of the attempts counted.

All eight tests were examined in a whole series of studies by various authors (Klein et al., 2012; Utesch et al., 2015; Boes and Schlenker, 2016) with regard to the test standards. Boes and Schlenker (2016) analyzed the test battery's psychometric properties for a sample consisting of nearly 50,000 school children and adolescents. The authors found an average test–retest correlation coefficient of $r_{tt} = 0.82$ at elementary-school age (7–11 years), even if there is considerable variation between the tests ranging from $r_{tt} = 0.52$ for balancing backward to $r_{tt} = 0.94$ for the bend forward.

Talent Promotion Interventions

On the side of the talent promotion variables, the club sport diversity of the 502 soccer participants was quantified by the number of different club sport activities with at least one official competition participation attended voluntarily before or in parallel to the talent screening, as well as during the following period of soccer practice during youth age.

Also very early, the sport administration of the government of the Hessen province offers a local talent promotion course, which is conducted by an expert sport teacher. Out of the 502 soccer participants, about 5% ($n = 26$) of the players took part in this talent promotion program. Such an additional 90-min school sport session takes place once a week and is open to about 10–15% of selected pupils with the aim to enhance the basic components of PF and MC. Thus, the generic training program aims primarily at the development of fundamental movement skills (Burrows et al., 2014; Barnett et al., 2016). The talent promotion courses start during the first grade of elementary school (normally at the age of 6–7 years) and end with finishing elementary school after the fourth grade.

As a third talent promotion strategy, two certified schools from the tested region offer sport-specific talent promotion classes, which were attended by about 10% ($n = 51$) of the soccer players. The sport classes consist of about 25 children who are selected and intensively guided by expert school sport officials. The membership in a sport class of one of the two local youth sport schools could comprise a maximum duration 6 years starting at the fifth grade (normally at the age of 11 years) and ending after the 10th grade (at the age of 16 years), which is 2–3 years before finishing high school education. The sport class members take part in two additional training sessions embedded in the official school timetable in the morning. Both training, sessions take place in sport-specific training centers and are guided by qualified expert coaches of the particular sport. The training focuses on general as well as sport-specific performance prerequisites.

As a fourth talent promotion campaign, the German Soccer Federation (DFB) offers between the age of 12 and 15 years 1 weekly soccer training session conducted by expert soccer coaches in one of the 366 soccer training centers distributed all over Germany. Building up on a basic soccer promotion stage (U11 groups and younger), the soccer training (or competence) centers form the second level of the pyramid of the successful talent promotion system of the German soccer system. The soccer training centers prepare about 14,000 promising youth athletes per year over 3 years in the age groups U12, U13, U14, and U15 (in some regions also U11), so the players can participate for 4 years in the program (Schott, 2010). In parallel, more than 50 professional soccer clubs run youth soccer academies, which include players from the same and also older age groups. The youth soccer academies are attached to the professional soccer clubs of the first, second, and third professional soccer divisions (Bundesliga). The youth soccer academies represent the second level of the talent promotion of the DFB and open the door to the third and final step to participate in the different DFB youth national teams. About 8% ($n = 40$) of the investigated soccer players were selected for a membership in one of the two regional talent promotion centers, with four players transferring to different youth soccer academies at age 15 after finishing with the training center.

Competition Performance

The soccer performance level (**Table 1**) reached by the soccer players until the end of the season 2019/2020 was utilized to quantify all athletes' success in early adolescence as criterion variable. All second-grade athletes who participated in the motor diagnostics and participated a minimum of one official soccer competition in a division or city level until the end of the season 2019/2020 or at least held an official club license without taking part in competitions (level 1) were recorded. Based on their success in the youth soccer competitions, the individual soccer performance was ranked from level 1 up to level 5, if the athlete was playing on the province "*Hessen level*." Although the participation in the Hessen soccer division (level 5) was the highest level registered in this study, from a nationwide perspective, such a soccer competition participation represents only a sub-elite soccer performance level. So it has to be stressed that there is still a gap between the soccer performance investigated in this study and the nation's best soccer performances within the U17 age group. To promote especially the elite youth soccer level, the German Soccer Federation not only organizes an U17 national championship by means of a first division competition (Bundesliga) but also promotes an U17 national team for Germany's best players.

TABLE 1 | Five-level scale for the recording of the soccer-specific competition results of the study participants taking part in official soccer competitions of the age groups U13 and U17 ($N = 502$).

5-level scale of U13 and U17 age group soccer performance		
Level 1	$N = 32$	Holder of an official soccer club license or competition participation on the city level (Kreisklasse) at the age of U13 and U17
Level 2	$N = 185$	Competition participation on the district level (Kreisliga) at the age of U13 and U17
Level 3	$N = 182$	Competition participation on the county level (Bezirksliga, Gruppenliga) at the age of U13 and U17
Level 4	$N = 62$	Competition participation on the regional level (Verbandsliga) at the age of U13 and U17
Level 5	$N = 41$	Competition participation on the province level (Hessenliga) at the age of U13 and U17

In contrast to a former study of Hohmann et al. (2018) on younger soccer players, in the U17 age group, the names of almost all study participants were individually reported in the team rosters published by the local media and the website of the Hessen Soccer Federation, thus substantially increasing the study sample of this study. However, the performance levels of game sport athletes are difficult to judge (Roescher et al., 2010; Gonaus and Mueller, 2012; Hoener and Votteler, 2016; Leyhr et al., 2020). To enhance the reliability of the assignments of the soccer players to the different competition performance levels, the records of the soccer performance of the U13 as well as the U17 age group were checked fine-grained, that is, individually player by player, by the head coach of the local competence center of the German Soccer Federation (DFB) who was in charge of the nomination, selection, and education of the youth soccer players between the age period of 12 and 15 years promoted by the DFB during the investigation period. As the scale of the five performance levels of the city, district, county, regional, and Hessen province leagues is not standardized for the different age groups during childhood and adolescence, and thus the age-specific performance levels do not accord perfectly with each other, this might impair the comparability of the individual soccer performances at the U13 and U17 age stages. Thus, in addition to the coach's individual rating, we checked the predictive validity of different age-specific soccer performances in our study sample. The high intercorrelations of the age-specific soccer performance variables (U13 * U15: $r_{tt} = 0.79$, $n = 502$, $p < 0.001$; U13 * U17: $r_{tt} = 0.79$, $n = 502$, $p < 0.001$; U15 * U17: $r_{tt} = 0.89$, $n = 502$, $p < 0.001$) underpinned not only the usability of the criterion variables but also the developmental stability of the individual pathways of the soccer performance development[1].

Statistical Analysis

In order to obtain solid results in regard to the prognostic relevance of the predictors, the data sets of four successive second grade cohorts (2010–2014) were collected, so the samples of the soccer players achieved a sufficient number on the higher competition levels. To gain insight into the prognostic relevance of the 10 predictors (two anthropometric variables and eight motor tests) in the soccer group and to clarify their predictive validity for the single dependent (criterion) variable of the adolescent soccer performance at age U17, univariate ANOVAs were conducted, analyzing mean differences between the five different soccer performance levels. According to Cohen (1988), the magnitude of effects was as classified small ($\eta^2 = 0.01$), medium ($\eta^2 = 0.06$), and large ($\eta^2 = 0.14$).

For the evaluation of motor predictors' prognostic validity, the age influence on the test performances should be considered (Meylan et al., 2010; Carling and Collins, 2014; Hoener and Votteler, 2016; Hoener et al., 2017). So ANOVAs were conducted to check the data set for significant differences of the youth athletes in regard to the yearly birth quarter. To control for confounding effects of the systematic influence of age found in the tests sideward jumping, 20-m sprint, push-ups, sit-ups, and 6-min run, the calendar age (in months) was partitioned out of the results in all 10 predictors by bivariate regression analysis. In the bivariate regression analyses, the test results served as the dependent variable, and the age (in months) as the independent variable (Siener and Hohmann, 2019). Furthermore, to allow for comparisons of the soccer-specific relevance of the different predictors and to estimate the lead of the soccer athletes against the general age group population, all test data were standardized by z-values based on the mean value and standard deviation data of the untrained athletes.

On the side of the talent promotion variables, the (i) club sport diversity was quantified by the number of club sport activities with an official competition participation before or in parallel to the testing and practicing soccer afterwards and thus represented a discrete variable. The participation in a (ii) local talent promotion course offered by the Hessen government sport administration, the attendance of a (iii) sport class of one or both local youth sport schools, and the membership in the (iv) local soccer training center of the German Soccer Federation (DFB) represented dichotomous dummy variables. Whereas, 411, that is, 81.9% of the $N = 502$ study participants, did not take part in any of the three talent promotion measures (ii–iv), 70 children (13.9%) participated in one of the interventions. From the remaining 21 (4.2%) of the investigated youth athletes, 12 of the 51 sport school members exercised also in the local soccer training center, but only seven of the sport school members took part in the governmental talent promotion course. Furthermore, four members of the local soccer training centers visited the governmental talent promotion courses in

[1]The predictive validity of the age-specific soccer performance variables even increased with age, as was shown in the oldest fraction of the investigated sample: U13 * U19: $r_{tt} = 0.87$, $n = 140$, $p < 0.001$; U15 * U19: $r_{tt} = 0.70$, $n = 140$, $p < 0.001$; U17 * U19: $r_{tt} = 0.97$, $n = 140$, $p < 0.001$.

parallel. None of the scholars participated in all three talent promotion programs.

According to former test validation results in youth soccer athletes (Hohmann et al., 2018) and a coach survey including 32 highly qualified soccer coaches ranking the eight test according to their relevance for youth soccer performance, a soccer recommendation score (SRS) was calculated for each participant. The SRS was formed by the mean value of a selection of the five most influential and, respectively, highest ranked general performance prerequisites, thus being a more complex soccer performance predictor than the single tests *per se*. On the basis of the ranking results of the previous studies, each of the five selected tests was hierarchically weighted by a stepwise weight factor (WF) between 1.2 and 2.0 according to their estimated validity for soccer performance. Thus, the 20-m sprint (WF 2.0), sideward jumping (WF 1.8), 6-min run (WF 1.6), standing long jump (WF 1.4), and sit-ups (WF 1.2) were used, because they had turned out to be more soccer-related than the other three tests and the two anthropometric measures. Based on this weighting, the SRS was calculated according to Formula (1).

$$SRS = 2^{*}z_{20m\ sprint} + 1.8^{*}z_{sideward\ jumping} + 1.6^{*}z_{6min\ endurance\ run} + 1.4^{*}z_{standing\ long\ jump} + 1.2^{*}z_{situps} \quad (1)$$

For a clearer view of a player's relative chances depending on several predictors, odds ratios (ORs) for the z-standardized performances were computed. On the basis of the OR, the prognostic validity of the test results was obtained. Furthermore, the questions, how many non-talented athletes would be disqualified (specificity in regard to true negatives), and how many talented youngsters would be successfully identified (sensitivity in regard to true positives) by the test battery were analyzed by calculating the odds stepwise between the 0.5% and up to 99.5% rank thresholds in the z-standardized test performances (Hoener and Votteler, 2016). To determine the optimal relation between drafted talents and excluded non-talents, the Youden index (also *Youden's J*) was calculated. This figure ranges between 0 (random result) and 1 (optimum result) and is calculated by summing up the two parameters sensitivity and specificity. Youden's J allows to quantify the prognostic validity independent of the sample size of the two performance groups (Youden, 1950). In this study, the basic assumption of Youden's J that specificity and sensitivity are of equal importance is fulfilled. Unlike in the soccer federations training centers or in sport school classes, at elementary-school age, there is still sufficient capacity in the soccer clubs to promote all promising youngsters. Thus, there is no need to reduce the number of false-positive soccer recommendation. Also, an increase of the number of true positives has not to be preferred one-sided, as almost all male second-class participants in school get in touch with the sport of soccer anyway and thus could be still detected as soccer talents by teachers and the soccer coaches later on.

A path-analytical model based on stepwise multivariate regression analyses examines the causal pathways of influence of independent variables on a dependent criterion variable measured later in time (Bryman and Carter, 1994). Thus, the influence of the 10 performance characteristics assessed at second grade, as well as the influence of the four different talent promotion interventions (diversity of club sport activities besides soccer, participation in the governmental talent promotion program, attendance of a sport school class, and membership of the German Soccer Federation training center) on the later soccer performance, was determined here. In a first step of the regression path analysis, all 10 performance characteristics (eight test performances and two anthropometric measures) and the four talent promotion interventions, plus the soccer performance level at age U13, were included as predictors, and their influence on the criterion variable performance level at age U17 was evaluated. In a second step, the soccer competition performance at age U17 was eliminated from the next regression analysis, and now the formerly independent early soccer competition performance at age U13 served as the dependent criterion variable.

In all procedures, data were analyzed using SPSS version 25.0 (IBM Corp.), and minimum level of significance was set at $p < 0.05$.

RESULTS

Test Performances of the Soccer Players and the Non-athletes

The descriptive performance characteristics of the test participants are presented in **Table 2**.

The early talent screening results for the age-adjusted and z-standardized performance characteristics in **Figure 4** demonstrate in nine of the 10 diagnostics better performances in the soccer players compared with the non-athletes who served as the reference group ($M_z = 0$). As described above, the five test results of the 20-m sprint, sideward jumping, 6-min run, standing long jump, and sit-ups formed the basis of the SRS. The adequacy of this selection is corroborated by the result that the soccer group exhibited already in the talent screening campaign in the second grade especially in these tests much better soccer-oriented performance prerequisites than did the non-athletes.

Predictive Validity of the General Performance Characteristics in Soccer

ANOVAs and Odds Ratios

Within the group of the U9 soccer players ($n = 502$), the ANOVAs demonstrated significant differences in the performance characteristics, which were in line with the differences in the later performance levels reached by the children as adolescents in the U17 age group. Soccer players who reached higher juvenile performance levels (see **Figure 5**) performed better not only in the SRS [$F_{(4;476)} = 16.36$; $p < 0.001$; $\eta^2 = 0.11$] but also in all eight single motor tests: 20-m sprint [$F_{(4;491)} = 12.06$; $p < 0.001$; $\eta^2 = 0.09$], standing long jump [$F_{(4;489)} = 10.95$; $p < 0.001$; $\eta^2 = 0.08$], sideward jumping [$F_{(4;491)} = 9.15$; $p < 0.001$; $\eta^2 = 0.07$], 6-min run [$F_{(4;479)} = 8.77$; $p < 0.001$; $\eta^2 = 0.07$], push-ups [$F_{(4;491)} = 7.79$; $p < 0.001$; $\eta^2 = 0.06$], balancing backward [$F_{(4;491)} = 6.96$; $p < 0.001$; $\eta^2 = 0.05$], bend forward [$F_{(4;486)} = 3.90$; $p < 0.01$; $\eta^2 = 0.03$], and sit-ups [$F_{(4;491)} = 3.27$; $p < 0.05$; $\eta^2 = 0.02$]. In regard to the

TABLE 2 | Descriptive statistics of age, soccer competition performances, two anthropometric, and eight motor diagnostics, soccer recommendation score, and four different talent promotion interventions in second-grade children taking part in the test campaigns 2010–2014 and competing in youth soccer competitions later on in the age groups U13 and U17 until the end of the season 2019/2020.

Variables	Groups	*N*	M	*SD*	SE	Min	Max
Calendar age							
U9: Calendar age (months)	Soccer players	499	95.23	7.24	0.21	82.00	125.00
	Non-athletes	2,159	95.53	6.48	0.14	75.00	127.00
U13: Calendar age (months)	Soccer players	501	148.74	7.70	0.24	156.00	179.00
U17: Calendar age (months)	Soccer players	502	192.47	7.94	0.36	180.00	203.00
Soccer performance							
U13: Performance level (pts)	Soccer players	501	1.89	0.78	0.03	1.00	5.00
U17: Performance level (pts)	Soccer players	502	2.80	1.02	0.05	1.00	5.00
Performance characteristics							
Body height (cm)	Soccer players	490	129.74	5.63	0.25	110.00	147.00
	Non-athletes	1,997	129.45	6.27	0.14	107.00	154.00
Body weight (kg)†	Soccer players	490	28.08	4.89	0.22	18.30	54.80
	Non-athletes	1,997	28.69	6.11	0.14	16.40	70.004.70
Sideward jumping (reps)†	Soccer players	493	25.55	6.14	0.28	11.00	44.00
	Non-athletes	1,993	23.07	6.40	0.14	0.50	44.50
Balancing backwards (steps)†	Soccer players	493	29.76	8.69	0.39	7.00	48.00
	Non-athletes	1,994	26.42	8.98	0.20	3.00	48.00
Standing long jump (cm)†	Soccer players	491	135.78	17.05	0.77	70.00	181.00
	Non-athletes	1,983	125.36	18.86	0.42	57.00	193.00
20-m sprint (s)†	Soccer players	493	4.50	0.33	0.01	3.54	5.58
	Non-athletes	1,992	4.60	0.39	0.01	3.60	7.15
Push-ups (reps)†	Soccer players	493	14.09	4.09	0.18	1.00	26.00
	Non-athletes	1,992	13.24	3.84	0.09	0.00	26.00
Sit-ups (reps)†	Soccer players	493	20.06	5.22	0.23	0.00	35.00
	Non-athletes	1,992	17.51	5.69	0.13	0.00	35.00
Bend forward (cm)†	Soccer players	488	0.87	5.25	0.24	−19.50	15.00
	Non-athletes	1,981	−0.54	6.06	0.14	−23.00	20.00
6-min run (m)†	Soccer players	481	1,019.68	120.40	5.49	604.00	1,359.00
	Non-athletes	1,953	907.31	133.51	3.02	108.00	1,306.00
Soccer recommendation score							
Soccer recommendation score (*z*-value)†	Soccer players	481	−0.29	1.14	0.05	−4.62	3.00
Talent promotion interventions (U17 soccer performance)							
German Soccer Federation training center (pts)*	Members	40	3.98	0.80	0.13	3.00	5.00
	Non-members	462	2.70	0.98	0.05	1.00	5.00
Sport school class (pts)*	Members	51	3.40	1.05	0.15	1.00	5.00
	Non-members	451	2.73	1.00	0.05	1.00	5.00
Governmental talent promotion program (pts)*	Members	27	3.27	1.04	0.20	1.00	5.00
	Non-members	475	2.77	1.02	0.05	1.00	5.00
Club sports diversity (pts)^{ns}	High diversity (≥3)	14	3.21	1.05	0.28	2.00	5.00
	Low diversity (≤2)	488	2.79	1.02	0.05	1.00	5.00

†*Significant between soccer players and non-athletes;* **significant between members and non-members;* ^{ns} *not significant between high and low club sports diversity.*

two anthropometric measures, the soccer players' body weight corresponded negatively to the individual performance level [$F_{(4;488)} = 4.43$; $p < 0.01$; $\eta^2 = 0.03$], whereas the body height [$F_{(4;488)} = 1.08$; $p = 0.366$; $\eta^2 = 0.006$] did not go systematically hand in hand with the later-on soccer performance level in the U17 age categories.

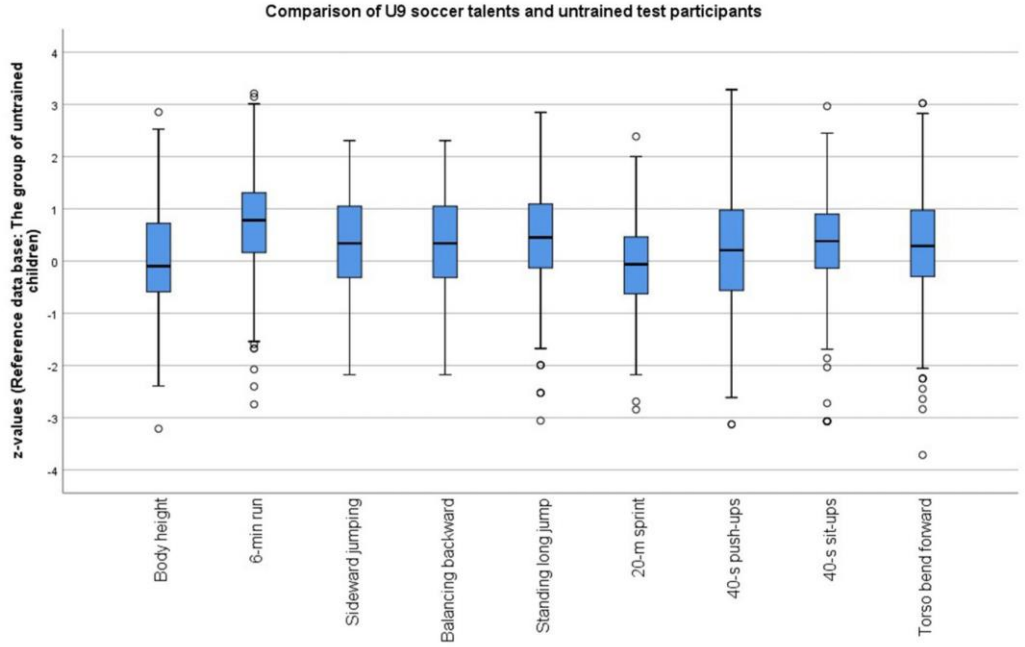

FIGURE 4 | Age-adjusted and *z*-standardized test performances of the soccer players and non-athletes of the second grade of elementary school in the test campaigns 2010–2014.

ORs for each single test as well as for the SRS represent the prognostic validity of the investigated predictors and make the sport-specific relevance of the different tests comparable. In the context of this study, the ORs quantify the relative chances of an U9 participant to reach the performance level 5 until the adolescent age group U17. In **Figure 6**, the ORs were calculated for participants who had achieved an individual test score among the best 16% of the total group ($z \geq 1.0$, corresponding to $PR \geq 84$) in the SRS, and in any of the eight single tests and the two body dimensions. With the use of this cutoff limit, the ORs of the performance characteristics 6-min run ($OR = 4.27$; $X^2 = 19.74$; $p < 0.001$), balancing ($OR = 4.03$; $X^2 = 18.08$; $p < 0.001$), and standing long jump ($OR = 2.18$; $X^2 = 4.47$; $p < 0.05$) show a significant difference between the two talent groups. On the other hand, the ORs for the body weight points at a significant negative perspective ($OR = 0.11$; $X^2 = 6.69$; $p < 0.05$) as the group of the heavier athletes have worse odds as the lighter counterparts.

The ORs for the soccer federation-, school-, and government-based talent promotion interventions were based on the binary code (Yes or No) for the participation in these campaigns. The OR for the diversity of competition participation in other sports before and besides soccer was based on the limit of at least three other sport activities. The chances for the youngsters to reach actually the maximum of soccer performance found in this study and thus to play particularly on the province level at the U17 age were substantially higher for those players who attended the local DFB soccer training center ($X^2 = 22.64$; $p < 0.001$; $\eta = 0.21$). In contrast, the odds for the participants of the afternoon training course of the governmental talent promotion program ($X^2 = 2.15$; $p = 0.142$) and for the members of the sport school classes ($X^2 = 2.52$; $p = 0.112$) were not systematically better than those for the abstainers. In regard to the extent of diversity of the individual training activities in different competitive club sports, those children focusing solely on soccer or taking part in only one or two additional sport were assigned to group 0, whereas the children participating in three or even more competitive sport club activities before or besides their soccer engagement were assigned to group 2. The better odds for the multi-sport children ($X^2 = 3.96$; $p < 0.05$; $\eta = 0.09$) are in line with the quota that three out of 13 children (23%) competing in at least three different club sports besides soccer reached the highest province level in soccer performance at the age of U17.

In regard to the utility of the talent identification campaign, the questions of how many non-talented athletes will be disqualified (specificity of the testing) and how many talented youngsters will be successfully identified (sensitivity of the testing) by the test battery are of great interest (Hoener and Votteler, 2016; Bergkamp et al., 2019). **Figure 7** documents that at a given cutoff limit of $z = 1.0$ ($PR = 84$) in the SRS, according to the sensitivity curve 24.4% of the participants, were correctly identified as future successful soccer players (true positives). The specificity of the testing amounts to 86.3%, which means that this fraction of low talented players (true negatives) would be excluded from talent promotion if the campaign aimed

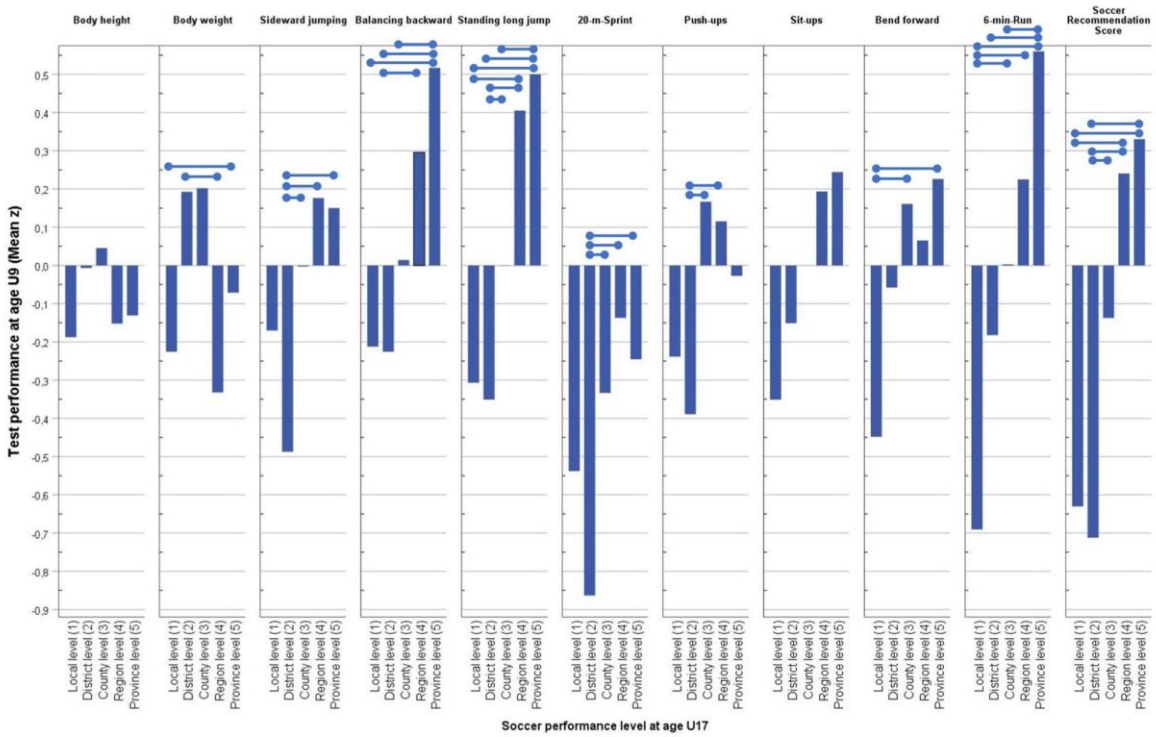

FIGURE 5 | Age-adjusted and *z*-standardized test performances of the U9 participants in regard to the later competition performance levels at adolescence (U17). Data of the 20-m sprint were transformed into positive direction. Variances of subgroups were homogeneous, but not in body weight [Levene (4;488) = 5.39; $p < 0.001$]; significant differences (Bonferroni *post-hoc* test; $p < 0.05$) between subgroups are marked by connection lines).

at securing at least a provincial soccer performance level. To optimize the balance between drafted talents and excluded non-talents, *Youden's J* of the SRS in **Figure 7** documents that the best compromise is reached by a cutoff limit at PR 70 ($z \geq 0.5$), where 51.2% of the future group of successful soccer players could be correctly identified, and 78.3% of the later-on less successful participants would be sorted out. In case of such a cutoff limit, the success rate in the group of the 120 drafted U9 talents would be 19.16%, as 23 of the draftees finally had reached the highest province level at age U17. Comparatively, the base rate in the tested 502 U9 soccer players was 7.97%, as a total of 41 talents reached the highest level in the future.

Regression Path Analysis

Besides the prognostic validity of the early performance characteristics, the long-term influence of different talent promotion interventions is of major importance. In parallel to the testing campaign and besides the 3 weekly school sport lessons of 45-min duration in elementary and grammar schools, the youth athletes are subject to a variety of sport performance-enhancing measures. Thus, in a first step of the path analysis, we investigated not only the prognostic validity of the 10 physical, physiological, and coordinative performance characteristics but also the influence of the (i) diversity of the voluntary club sport participation and of the three official talent promotion interventions, (ii) governmental talent promotion program for the best movers, (iii) sport classes of the two local sport schools, and the (iv) training center of the German Soccer Federation. In the first step, also the U13 soccer performance was added to the list of regressors to predict the U17 soccer performance. In a second step, we left out the U17 soccer performance and substituted it by the U13 soccer performance. Now, the 10 performance characteristics and the four talent promotion measures served to predict the U13 soccer performance. Both regression analyses were tested for multicollinearity. Neither in the U17 regression model ($0.492 < TOL < 0.976$; $1.025 < VIF < 2.033$) nor in the U13 model the tolerance (TOL) coefficients and variance inflation factors (VIFs), respectively ($0.492 < TOL < 0.970$, $1.024 < VIF < 2.033$) confirmed such an intercorrelation between the independent variables that none of these predictors had to be excluded from one of the two combined models (Eid et al., 2017). **Figure 8** shows the complete result of the regression-based path model documenting the statistical relevant pathways of influence of the predictors on the soccer performance development. The configuration of the model components in **Figure 8** is based on the process structure of our abstract talent development model shown in the beginning (see **Figure 1**), thus adding the social talent promotion

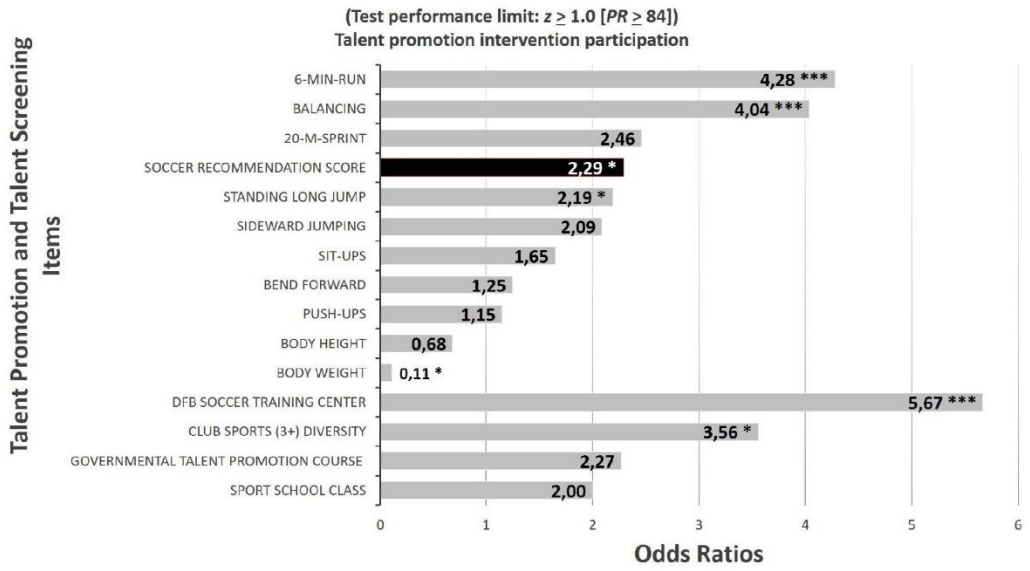

FIGURE 6 | Odds ratios (at the $z = 1.0$ and $PR = 84$ threshold) of the two body dimensions and eight generic motor tests and the soccer recommendation score, as well as of four talent promotion interventions to predict the relative chances of the U9 participants to reach the highest (province) level of soccer competition performance at adolescent age (U17). Significant results are marked by asterisks: $^{*}p < 0.05$; $^{***}p < 0.001$.

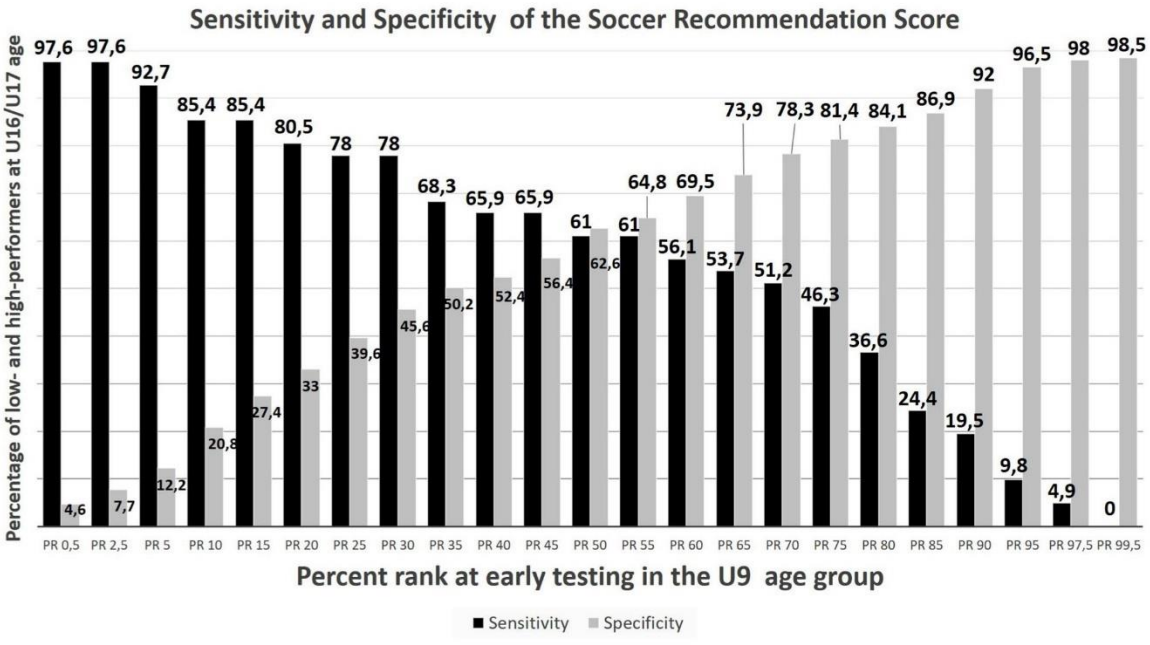

FIGURE 7 | Specificity and sensitivity of the soccer recommendation score (SRS) of the U9 participants to predict a future soccer performance on the highest provincial level (level 5) at adolescence (U17).

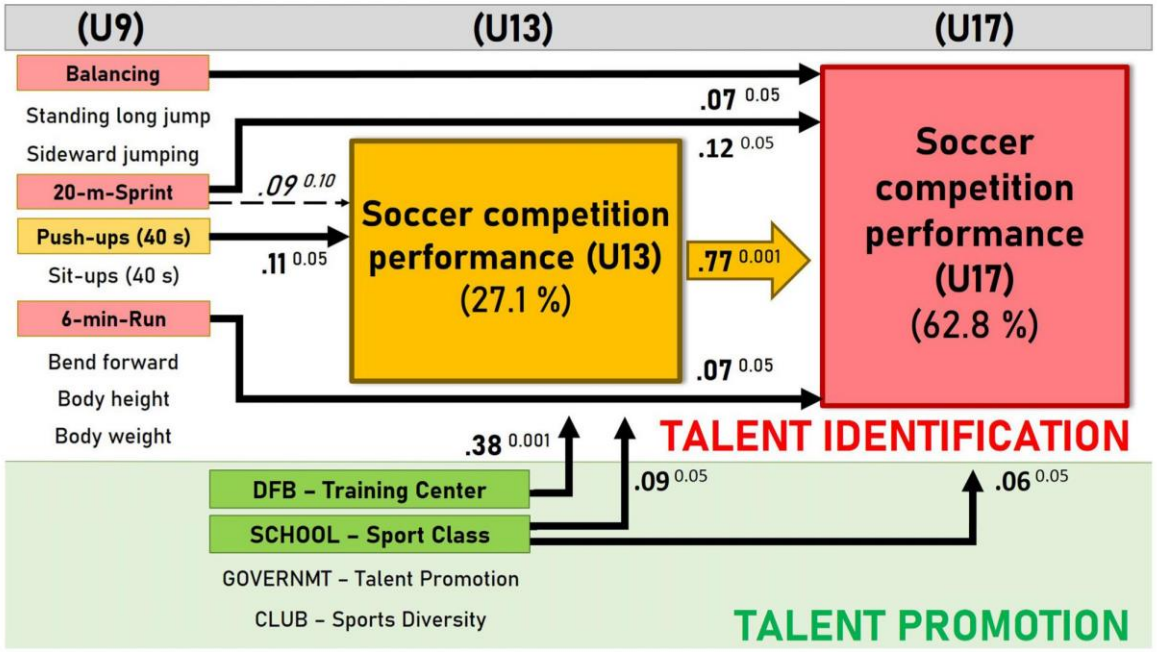

FIGURE 8 | Regression path analysis of early performance characteristics at U9 age and their relevance for midterm performance development in youth and adolescent soccer players ($N = 502$).

interventions from the sport-specific environment of the youth soccer players to the portfolio of the physical, physiological, and coordinative performance characteristics of the soccer athletes.

As can be seen in the path-analytical model in **Figure 8**, the adolescent soccer performance at age U17 [$F_{(15;463)} = 54.83$; $p < 0.001$] is predominantly determined by the intermediate soccer performance at age U13. Besides the intermediate soccer expertise at age U13, the adolescent soccer performance is also significantly influenced by early PF (20-m sprint and 6-min run; $p < 0.05$) as well as MC (dynamic balancing backward; $p < 0.05$). During the early stage of youth development, the intermediate U13 soccer performance [$F_{(14;464)} = 13.71$; $p < 0.001$] is primarily determined by the early MC at age U9, which is in particular represented by the difficult coordination of the upper limbs in the challenging push-up task ($p < 0.05$) and the sprint run performance (20-m sprint; $p < 0.10$), although the latter effect is smaller and only shows a tendency toward significance. The upper body coordination in the push-up test in the second graders shows a direct effect (beta = 0.11) on early soccer competition performance at age U13 and an indirect effect via this mediator on soccer competition performance at age U17. The indirect effect (beta = 0.08) reflects a full mediation because after controlling for the mediator, body coordination has no significant direct effect on soccer performance at age U17 (Preacher and Hayes, 2004). Finally, there was no systematic effect of the standing long jump, sideward jumping, abdominal strength endurance, torso flexibility, body height, or body weight on the soccer performance at early (U13) and late adolescent age (U17).

On the side of the talent promotion measures, a significant relation could be found between the membership in a sport class at school and the soccer competition performance of the children at age U13 (beta = 0.09; $p < 0.05$), and also at the later age U17 (beta = 0.06; $p < 0.05$). The highest effect is seen in the sport school support on the soccer success at age U13 (beta = 0.38; $p < 0.001$). On the other hand, no significant effect could be found between the engagement in the governmental afternoon sport course and the subsequent soccer performance at adolescent age. Also, in the total group, the number of additional club sport activities besides soccer was not related to the soccer competition performance at the ages U13 and U17. As only the players of the lowest and highest performance levels participated to a greater extent in more than one other sport besides soccer training, we obtained a significant curvilinear relation between the early diversity of club sport engagement and the future success in adolescent soccer competition [quadr $R^2 = 0.02$; $F_{(2;497)} = 5.90$; $p < 0.01$].

DISCUSSION

It is imperative that early talent screening and sports orientation campaigns based on the assessment of the performance predictors and competition performance have to be related to a long-term talent development process and properly allocated

to defined stages of a complex talent development model. Compared with recent studies (Huijgen et al., 2013; Hoener and Votteler, 2016; Hoener et al., 2017; Leyhr et al., 2018) investigating the prognostic relevance of mainly soccer-specific skills as talent predictors in male youth soccer, this study differs in that it assessed mainly the more general athleticism and MC of a very young male sample over a long-term period from the early talent screening in second grade (U9) until the gateway to expert performance at adolescent age (U17).

The present results are in line with Gil et al. (2007), who assumed only a short-term advantage of larger body dimensions but on the other hand contradict the later importance of the anthropometric variables body height and body weight found by Hoener et al. (2017) in early adolescence for more or less successful adult players. These findings could have resulted from two different factors. On the one hand, the elimination of the calendar age from our test data by partializing out the age (in months) from all test performance data could have led to a more homogeneous "age group." On the other hand, there could exist a considerably lower relative age effect in the early talent development stage investigated in this study compared with middle or late adolescence (Musch and Grondin, 2001; Carling et al., 2009). Due to the early stage of the talent screening campaign, the U9 participants are still far from the pre-pubertal acceleration of the development of body dimensions, which might reduce the impact of the anthropometric predictors on the soccer performance of the young people. Nevertheless, to fully understand the relationship between body size and biological maturity (Meylan et al., 2010; Lago-Peñas et al., 2011; Lovell et al., 2015; Patel et al., 2019) and soccer performance development from childhood to adolescence, more differentiated analyses would be worthwhile.

According to the ANOVA results, and especially the ORs of the PF and MC tests, this longitudinal study verified the prognostic validity of all eight generic motor tests in the sport of soccer. In this context, the tendency of a significant long-term prognostic validity of the MC test balancing backward, as well as the explosive PF test standing long jump already at the U9 age, comes as no surprise (Reilly et al., 2000; Wong et al., 2009; Deprez et al., 2015; Sarmento et al., 2018) and might be an indicator of early athleticism that is relevant for soccer talents, be it innate or a specific outcome of early soccer training experience (Franks et al., 1999). All in all, at least the design of our study is comparable with the longitudinal approach of Hoener and Votteler (2016) and Leyhr et al. (2018), although their studies analyzed the prognostic relevance of soccer-specific motor diagnostics on the talent development stage from early adolescence (U12–U15) until junior age (U16–U19). In contrast, their study focused on elite youth athletes, having already excelled during early, mid, and late adolescence. Although our study was based on more general soccer performance prerequisites, it might help to complete the picture of the soccer performance development reaching from the search for an early soccer-related profile of general athleticism in a very heterogeneous sample at elementary-school age until the soccer-specific talent selection measures at junior age, which are normally based on very homogeneous soccer populations and include high-fidelity diagnostics (Bergkamp et al., 2019). In line with our results, Deprez et al. (2015), Figueiredo et al. (2009), Gonaus and Mueller (2012), Le Gall et al. (2010), and Zuber et al. (2015, 2016) underlined the midterm (≥2 years) relevance of running endurance and sprint speed in youth soccer players. Interestingly, in our study, the 20-m sprint and the 6-min run performances were systematically correlated (r_{tt} = 0.39; n = 484; $p < 0.001$), which might refer to an age-specific, general running ability. In regard to the direct pathway between balancing and later soccer performance, Mirkov et al. (2010) likewise stressed the high relevance of coordination for future success in soccer. Another reason could be that children who already played soccer before the testing have improved their motor coordination and, thus, achieved better test results. Last but not least, also the participation of 70 soccer athletes in the different talent promotion programs, which stress especially general coordination ability and fundamental movement skills, might have contributed to the long-term impact of dynamic balance on the soccer performance. As a consequence of the particular results in the OR and the path-analytical model, the composition of the SRS should be revisited. It seems that a supplement of the highly valid dynamical balance test could contribute to a better soccer talent forecast at least at the elementary-school age, which would be in line not only with the findings of Mirkov et al. (2010) but also with the test ranking of the coaches obtained before.

The structure of the path-analytical model outlined in **Figure 8** was based on our process-oriented talent development model described in the beginning (see **Figure 1**) and is consistent with the general model of talent development outlined by Abbot and Collins (2002), Fisher and Bailey (2008), and Heller and Hany (1986), where the performance-enhancing effect of the talent promotion environment is taken for granted. It is well-known and in line with our findings that elite sport schools (Granacher and Borde, 2017), as well as the talent promotion training centers of the German Soccer Federation (Hoener et al., 2015), have contributed largely to performance development and sporting success. On the other hand, we could not confirm a positive effect of the province-wide talent promotion campaign "Talent Search and Talent Promotion" of the regional government on the performance of youth soccer players. The lack of evidence might be due to the early-onset, low 90-min per week volume and short duration of the campaign between ages 6 and 10, as well as the generic training program focusing predominantly on fundamental movement skills. So our result is not surprising, as the objectives of this program are not to foster soccer-specific skills but more general PF and MC independent from the children's decision on the talent development pathway. As coach-led, soccer-specific practice necessary for later success (Haugaasen et al., 2014; Hornig et al., 2014) does not happen in this course, it might exert only little effect on future soccer expertise at adolescent age.

Our findings were not fully consistent in regard to the controversial debate of whether a high diversity of early club sport activities enhanced talent development by a greater portfolio of protective coordination patterns and transferable technical skills (Cote and Hay, 2002; Baker, 2003; Baker et al., 2003; Soberlak and Cote, 2003; Cote et al., 2009; Myer et al.,

2015) contributes to better learning pre-requisites to optimize soccer-specific technical skills (Huijgen et al., 2010, 2013) or has more negative than positive consequences for long-term athletic development by impairing sport-specific technical and tactical education due to reduced time, energy, and logistic resources (Helsen et al., 1998; Ward et al., 2007; Forsman et al., 2016a; Zuber et al., 2016). In our path-analytical model, the contribution of the diversity of club sport participation did not play a significant role, as in the total group of all soccer players there was no significant linear correlation with the performance levels reached at adolescence. In general, our path-analytical model confirmed the findings of Ford et al. (2009), Ford and Williams (2012), and Ward et al. (2007), as it did not underline a positive effect of an early diversification of club sport participation. On the other hand, at least the ORs of the diversified club sport participation were in line with Guellich et al. (2017) that such a multi-sport engagement could serve as a promising strategy to reach the highest province soccer performance level later on. The explanation for the missing relation might lie in the fact that only in the lowest and highest performance groups of the soccer players a major fraction of children engaged in three or even more competitive sports. In the medium soccer performance groups of the district or county performance levels (2–3), most of the youth athletes concentrated solely on soccer or participated in one or two additional sports only. Our interpretation for this situation is that at the low soccer performance level, a substantial number of athletes with good general performance characteristics preferred other sports as main disciplines and thus participated in soccer training and competition games just as side events to their main sport. This interpretation is in line with the significant quadratic bivariate regression function found between the training diversity that is the number of sports performed during youth development and the adolescent soccer performance level reached at adolescence. Nevertheless, we assume that the relation between the diversity of early club sport engagement and future soccer success again disappears when the best youth soccer athletes performing on national or international youth level were included, as the findings of Ford et al. (2012) and Hornig et al. (2014) suggest. In our study, all of the province level soccer athletes practiced in the three sports—track and field, tennis, and table tennis only—which is in line with the specialized sampling approach of Sieghartsleitner et al. (2018) postulating that additional side-sport activities should show at least complementary relations and thus provide positive contributions to the main sport soccer to enhance the specific PF or MC profile needed there.

There are some limitations in this study. The operationalization of the individual soccer performance by means of the highest soccer performance level reached by the club team is still debatable (Baker et al., 2015; Bergkamp et al., 2019; Leyhr et al., 2020). To compensate for the lack of individual in-match player performance assessment, we introduced an additional control step to have the individual soccer performance criterion finally checked fine-grained player-per-player by the regional talent coordinator of the German Soccer Federation (DFB). Besides that, we can also assume that due to natural selection and club policies on the higher U17 county and province levels, the individual performances correspond pretty close to the teams' overall soccer performance. Nevertheless, there might be some internal validity problems at least on the lower levels 1–2, which are on a local level. Here, the player participation in a youth club team might also depend to a certain degree on the preference of early peer relations or shorter travel distances to the training and competition sites and thus underestimate the individual soccer performance level. Thus, on the two lowest levels, team success might not fully represent the individual athletes' performance capacity. Furthermore, due to less frequent reporting on individual player participation in the media and internet match reports on the lowest levels 1–2, the range in the criterion soccer performance could be somewhat restricted, so that the pathways between the predictors and the later-on soccer performances found in our study might be underestimated (Bergkamp et al., 2019). Especially, in our path-analytical model, more precisely classified low-level participants could improve the significance and prognostic validity of the different performance characteristics of the U9 players. Besides the low fidelity of the criterion variable team performance level, a limitation of our study is also that the investigated performance levels end up at the sub-elite Hessen province level. From a nationwide perspective on talent identification, only the highest youth soccer performance levels of the U17 Regionalliga (level 6) and Bundesliga (level 7) truly represent youth elite soccer performance. So it has to be stressed that there is still a gap between the soccer performance investigated in this study and the nation's best soccer performances within the U17 age group. In the U9 age group, the generic test battery used here generally allows for a first comprehensive performance diagnosis to identify good movers with a suitable profile of anthropometry, PF, and MC for a deliberate soccer training. Besides the initial quality of the performance characteristics other factors like sport-specific training initiation and soccer experience, learning conditions and facilities, quality of training and training volume, and field position might influence the performance development at youth age. So the lack of information about these potentially confounding variables is a limitation of the present study.

Also, there is still room for improvement in the game sports where test items focusing on technical skills and agility without or in combination with decision-making could be added to improve the prognostics (O'Connor et al., 2016; Ulbricht et al., 2016; Schorer et al., 2017). On the one hand, such predictors of higher fidelity may lead to a better explanation of the variance between the soccer performance groups (Bergkamp et al., 2019). On the other hand, in soccer, Hohmann et al. (2018) observed at elementary-school age a lower prognostic validity of sport-specific skill tests compared with generic physical or physiological tests, which the authors attributed to the higher difficulty of technical tests and the short time span of soccer-specific education. In order to do justice to the multidimensional character of football performance (Huijgen et al., 2010), in future studies at primary school age, the general tests should be supplemented, if not by football-specific tests, then at least by semi-specific tests to test fundamental ball-handling skills (Rylander et al., 2019). Despite such concerns, it is obvious that research has to be extended to the early age group of

elementary-school children, as Ford et al. (2020) have reported recently that a substantial fraction of worldwide youth soccer academies have installed talent identification systems at the age group level of 8 to 11 year-old players already. In addition to sport-specific test extensions, psychological test items might also provide further indications of potential future top performers (Morris, 2000; Epuran et al., 2008; Forsman et al., 2016a; Johnston et al., 2018; Murr et al., 2018a,b), although it is still unclear whether such attempts could be a fruitful endeavor at the early stage of talent development during elementary-school age already. At least in the U12 age group, Zuber et al. (2015), as well as Hoener et al. (2015), have made promising attempts to integrate psychological testing in a talent identification campaign of the Swiss Soccer Federation (SFV) and the German Soccer Federation (DFB), respectively. All in all, further research is needed not only to contribute to the extension of the current knowledge on the prognostic validity of motor assessment on this early talent development stage but also to clarify whether such general talent diagnostic campaigns are suitable to support youth sport practitioners in the task of recommending and assigning the right sport to each of the promising individuals (Pion, 2015).

Nevertheless, despite multifaceted and optimized test tasks as well as sophisticated multivariate statistical tools, not every talented athlete can always be surely identified, be it at elementary-school age or during puberty and adolescence. In contrast to linear talent prediction methods, some unexpected players develop into professional athletes, as even during late adolescence very few career paths are straightforward (Elferink-Gemser et al., 2011; Gulbin et al., 2013; Li et al., 2018). Thus, very often unsteady development of performance characteristics (Vaeyens et al., 2008; Fransen et al., 2017) and non-linear pathways to the top of the scene are determined by many interacting dynamic parameters to which each athlete reacts in a unique way (Phillips et al., 2010), and which should be reflected in the research methodology as well (Den Hartigh et al., 2016; Fransen et al., 2017; Leyhr et al., 2020).

CONCLUSION

Our study shows that early talent screening and sports orientation can be used to make valid statements in regard to future success of young soccer players. It also provides reliable empirical knowledge on the prognostic relevance of eight generic PF and MC tests in a regional talent screening and sports orientation campaign. The results show motor predictors' prognostic validity over a long-term period (on average about 8 years) after controlling all test data for calendar age. The specificity of the generic testing in the second grade is very high, and the majority of non-talents is advised to pursue sports other than soccer for which they were better suited. However, due to the medium sensitivity of the testing, recommending elementary-school children to focus solemnly on the sport soccer still remains a very complex practical and theoretical problem. The practical application of the test battery assessing speed, endurance, and coordinative abilities turned out to be a useful tool for talent orientation as a combination of multifaceted testing and a subsequent sports recommendation on the basis of the best match of the individual profile with the concrete demands of specific sports like soccer.

Further studies with alternate methodological approaches, like person-oriented pattern analyses (Zuber et al., 2015, 2016; Zibung et al., 2016), and a promising combination of linear and non-linear prognostic tools (Pion et al., 2016) should be further examined and compared with each other to identify the corresponding strengths and weaknesses. In doing so, the research on talent orientation may provide coaches with more scientifically valid tools for supporting their talent development strategies as well as offering a deeper understanding of the effectiveness of talent promotion interventions in the process of the long-term development of talented youth soccer players.

DATA AVAILABILITY STATEMENT

The data associated with the paper are not publicly available but are available from the corresponding author on reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Municipality of Fulda, University of Bayreuth. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. The Declaration of Helsinki was checked and adhered to in all parts of this study.

AUTHOR CONTRIBUTIONS

AH contributed on to the design and implementation of the research. MS performed the test data collection. All authors contributed to the final version of the manuscript.

FUNDING

This publication was funded by the German Research Foundation (DFG) and the University of Bayreuth in the funding program Open Access Publishing.

ACKNOWLEDGMENTS

This study is part of the research project Evaluation of a model for talent identification at elementary school age, which was supported by the German Institute of Sport Science (Bonn, Germany; ZMVI4-070504/15-16). Also, we would like to express our gratitude to the municipality of Fulda, the Fulda district authorities, and the participating elementary schools for supporting the test campaigns 2010–2014. Furthermore, we would like to thank Matthias Wilde, the regional talent coordinator of the German Soccer Federation (DFB), for contributing in the final check of the individual soccer performance level of all players included in the study.

REFERENCES

Abbot, A., and Collins, D. (2002). A theoretical and empirical analysis of a 'State of the Art' talent identification model. *High Abil. Stud.* 13, 157–178. doi: 10.1080/1359813022000048798

Baker, J. (2003). Early specialization in youth sport: a requirement for adult expertise? *High Abil. Stud.* 14, 85–94. doi: 10.1080/13598130304091

Baker, J., Cote, J., and Abernethy, B. (2003). Sport specific training, deliberate practice and the development of expertise in team ball sports. *J. Appl. Sport Psychol.* 15, 12–25. doi: 10.1080/10413200305400

Baker, J., Wattie, N., and Schorer, J. (2015). "Defining expetise: a taxonomy for researchers in skill acquisition and expertise," in *Routlegde Handbook of Sport Expertise*, eds J. Baker and D. Farrow (London: Routledge), 183–195.

Bardid, F., Utesch, T., and Lenoir, M. (2019). Investigating the construct of motor competence in middle childhood using the BOT-2 Short Form: an item response theory perspective. *Scand. J. Med. Sci. Sports* 29, 1980–1987. doi: 10.1111/sms.13527

Barnett, L. M., Lai, S. K., Veldman, S. L. C., Hardy, L. L., Cliff, D. P., Morgan, P. J., et al. (2016). Correlates of gross motor competence in children and adolescents: a systematic review and meta-analysis. *Sports Med.* 46, 1663–1688. doi: 10.1007/s40279-016-0495-z

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

Boes, K., Opper, E., Woll, A., Liebisch, R., Breithecker, D., and Kremer, B. (2001). The Karlsruhe test system for children (KATS-K). *Haltung Bewegung* 21, 4–66.

Boes, K., and Schlenker, L. (2016). *Der Deutsche Motorik-Test 6-18 (German Motor Test 6-18)*. Hamburg: Feldhaus.

Bryman, A., and Carter, D. (1994). *Quantitative Data Analysis for Social Scientists*. New York, NY: Routledge.

Buekers, M., Borry, P., and Rowe, P. (2015). Talent in sports. Some reflections about the search for future champions. *Mov. Sport Sci.* 88, 3–12. doi: 10.3917/sm.088.0003

Burrows, E. J., Keats, M. R., and Kolen, A. M. (2014). Contributions of after school programs to the development of fundamental movement skills in children. *Int. J. Exerc. Sci.* 7, 236–249.

Carling, C., and Collins, D. (2014). Comment on "Football-specific fitness testing: adding value or confirming the evidence?" *J. Sports Sci.* 32, 1206–1208. doi: 10.1080/02640414.2014.898858

Carling, C., Le Gall, F., and Malina, R. M. (2012). Body size. skeletal maturity. and functional characteristics of elite academy soccer players on entry between 1992 and 2003. *J. Sports Sci.* 30, 1683–1693. doi: 10.1080/02640414.2011.637950

Carling, C., Le Gall, F., Reilly, T., and Williams, F. (2009). Do anthropometric and fitness characteristics vary according to birth date distribution in elite youth academy soccer players? *Scand. J. Med. Sci. Sports* 19, 3–9. doi: 10.1111/j.1600-0838.2008.00867.x

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: L. Erlbaum Associates.

Cote, J., and Hay, J. (2002). "Children's involvement in sport: a developmental perspective," in *Psychological Foundations of Sport*, eds J. M. Silva and D. Stevens (Boston, MA: Allyn and Bacon), 484–502.

Cote, J., Lidor, R., and Hackfort, D. (2009). ISSP position stand: to sample or to specialize? Seven postulates about youth sport activities that lead to continued participation and elite performance. *Int. J. Sport Exerc. Psychol.* 7, 7–17. doi: 10.1080/1612197X.2009.9671889

Davids, K., and Baker, J. (2007). Genes, environment and sport performance. *Sports Med.* 37, 961–980. doi: 10.2165/00007256-200737110-00004

Demetriou, Y., Bucksch, J., Hebestreit, A., Schlund, A., Niessner, C., Schmidt, S. C. E., et al. (2019). Germany's 2018 report card on physical activity for children and youth. *German J. Exerc. Sport Res.* 49, 113–126. doi: 10.1007/s12662-019-00578-1

Den Hartigh, R. J. R., Van Dijk, M. W. G., Steenbeek, H. W., and Van Geert, P. L. C. (2016). A dynamical network model to explain the development of excellent human performance. *Front. Psychol.* 7:532. doi: 10.3389/fpsyg.2016.00532

Deprez, D., Buchheit, M., Fransen, J., Boone, J., Lenoir, M., Philippaerts, R., et al. (2015). Characteristics of high-level youth soccer players: variation by playing position. *J. Sports Sci.* 33, 243–254. doi: 10.1080/02640414.2014.934707

Eid, M., Gollwitzer, M., and Schmitt, M. (2017). *Statistik und Forschungsmethoden (Statistics and Research Methods) (7. Auflage)*. Weinheim: Beltz.

Elferink-Gemser, M. T., Jordet, G., Coelho-E-Silva, M. J., and Visscher, C. (2011). The marvels of elite sports: how to get there? *Br. J. Sports Med.* 45, 683–684. doi: 10.1136/bjsports-2011-090254

Epuran, M., Holdevici, I., and Tonita, F. (2008). *Performance Sport Psychology. Theory and Practice*. Bucharest: Fest.

Fetz, F., and Kornexl, E. (1978). *Sport Motor Tests*. Berlin: Bartels and Wernitz.

Figueiredo, A. J., Gonçalves, C. E., Coelho, E., Silva, M. J., and Malina, R. M. (2009). Youth soccer players. 11–14 years: maturity, size, function, skill and goal orientation. *Ann. Hum. Biol.* 36, 60–73. doi: 10.1080/03014460802570584

Fisher, R., and Bailey, R. (Eds.). (2008). *Talent Identification and Development. The Search for Sporting Excellence*. Berlin: ICSSPE.

Ford, P. R., Carling, C., and Garces, M., et al. (2012). The developmental activities of elite soccer players aged under-16 years from Brazil, England, France, Ghana, Mexico, Portugal and Sweden. *J. Sports Sci.* 30, 1653–1663. doi: 10.1080/02640414.2012.701762

Ford, P. R., Delgado Bordonau, J. L., Bonanno, D., Tavares, J., Groenendijk, C., and Fink, C. (2020). A survey of talent identification and development processes in the youth academies of professional soccer clubs from around the world. *J. Sports Sci.* 38, 1269–1278. doi: 10.1080/02640414.2020.1752440

Ford, P. R., Ward, P., and Hodges, N. J., et al. (2009). The role of deliberate practice and play in career progression in sport: the early engagement hypothesis. *High Abil. Stud.* 20, 65–75. doi: 10.1080/13598130902860721

Ford, P. R., and Williams, A. M. (2012). The developmental activities engaged by elite youth soccer players who progressed to professional status compared to those who did not. *Psychol. Sport Exerc.* 13, 349–352. doi: 10.1016/j.psychsport.2011.09.004

Forsman, H., Blomqvist, M., Davids, K., Konttinen, N., and Liukkonen, J. (2016a). The role of sport-specific play and practice during childhood in the development of adolescent Finnish team sport athletes. *Int. J. Sports Sci. Coach.* 11, 69–77. doi: 10.1177/1747954115624816

Forsman, H., Grasten, A., Blomqvist, M., Davids, K., Liukkonen, J., and Konttinen, N. (2016b). Development of perceived competence, tactical skills, motivation, technical skills, and speed and agility in young soccer players. *J. Sports Sci.* 34, 1311–1318. doi: 10.1080/02640414.2015.1127401

Franks, A. M., Williams, A. M., Reilly, T., and Nevill, A. (1999). Talent identification in elite youth soccer players: physical and physiological characteristics. Communication to the 4th World Congress on Science and Football, Sydney. *J. Sports Sci.* 17:812.

Fransen, J., Bennett, K., Woods, C., French-Collier, N., Deprez, D., Vaeyens, R., et al. (2017). Modelling age-related changes in motor competence and physical fitness in high-level youth soccer players: implications for talent identification and development. *Sci. Med. Football* 1, 203–208. doi: 10.1080/24733938.2017.1366039

Fuchslocher, J., Romann, M., Ruedisueli, L. R., Birrer, D., and Hollenstein, C. (2011). Das Talentselektionsinstrument PISTE: wie die Schweiz Nachwuchsathleten auswählt. *Leistungssport* 41, 22–27.

German Soccer Federation (2009). *Talent Development Program. Training Guideline (Deutscher Fußball-Bund (DFB). Talentförderprogramm. Leitfaden für die Ausbildung)*. Münster: Philippka.

Gil, S., Ruiz, F., Irazusta, A., Gil, J., and Irazusta, J. (2007). Selection of young soccer players in terms of anthropometric and physiological factors. *J. Sports Med. Phys. Fitness* 47, 25–32.

Golle, K., Muehlbauer, T., Wick, D., and Granacher, U. (2015). Physical fitness percentiles of German children aged 9–12 years: findings from a longitudinal study. *PLoS ONE* 10:e0142393. doi: 10.1371/journal.pone.0142393

Gonaus, C., and Mueller, E. (2012). Using physiological data to predict future career progression in 14- to 17-year-old Austrian soccer academy players. *J. Sports Sci.* 30, 1673–1682. doi: 10.1080/02640414.2012.713980

Granacher, U., and Borde, R. (2017). Effects of sport-specific training during the early stages of long-term athlete development on physical fitness, body composition, cognitive, and academic performances. *Front. Physiol.* 8:810. doi: 10.3389/fphys.2017.00810

Guellich, A. (2014). Selection. de-selection and progression in German football talent promotion. *Eur. J. Sport Sci.* 14, 530–537. doi: 10.1080/17461391.2013.858371

Guellich, A., Kovar, P., Zart, S., and Reimann, A. (2017). Sport activities differentiating match-play improvement in elite youth footballers – a 2-year longitudinal study. *J. Sports Sci.* 35, 207–215. doi: 10.1080/02640414.2016.1161206

Gulbin,. J., Weissensteiner, J., Oldenziel, K., and Gagné, F. (2013). Patterns of performance development in elite athletes. *Eur. J. Sport Sci.* 13, 605–614. doi: 10.1080/17461391.2012.756542

Haag, H., and Singer, R. (1981). "Development of a test to measure motor fitness," in *Sport Education and Evaluation*, eds H. Haag et al. (Schorndorf: Hofmann).

Haugaasen, M., Toering, T., and Jordet, G. (2014). From childhood to senior professional football: a multi-level approach to elite youth football players'engagement in football-specific activities. *Psychol. Sport Exerc.* 15, 336–344. doi: 10.1016/j.psychsport.2014.02.007

Heller, K. A., and Hany, E. A. (1986). "Identification, development and achievement analysis of talented and gifted children in West Germany," in *Identifying and Nurturing the Gifted. An International Perspective*, eds K. A. Heller and J. F. Feldhusen (Toronto, ON: Huber), 67–82.

Helsen, W. F., Starkes, J. L., and Hodges, N. J. (1998). Team sports and the theory of deliberate practice. *J. Sports Sci.* 20, 12–34. doi: 10.1123/jsep.20.1.12

Hoener, 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:e0182211. doi: 10.1371/journal.pone.0182211

Hoener, 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. *J. Sports Sci.* 34, 2269–2278. doi: 10.1080/02640414.2016.1177658

Hoener, 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. *J. Sports Sci.* 33, 145–159. doi: 10.1080/02640414.2014.928416

Hohmann, A., and Seidel, I. (2003). Scientific aspects of talent development. *Int. J. Phys. Educ.* 40, 9–20.

Hohmann, A., Siener, M., and He, R. (2018). Prognostic validity of talent orientation in soccer. *German J. Exerc. Sport Res.* 48, 478–488. doi: 10.1007/s12662-018-0549-5

Hornig, M., Aust, F., and Guellich, A. (2014). Practice and play in the development of German top-level professional football players. *Eur. J. Sport Sci.* 14, 1–10. doi: 10.1080/17461391.2014.982204

Huijgen, B. C. H., Elferink-Gemser, M. T., Ali, A., and Visscher, C. (2013). Soccer skill development in talented players. *Int. J. Sports Med.* 34, 720–726. doi: 10.1055/s-0032-1323781

Huijgen, B. C. H., Elferink-Gemser, M. T., Lemmink, K. A. P. M., and Visscher, C. (2014). Multidimensional performance characteristics in selected and deselected talented soccer players. *Eur. J. Sports Sci.* 14, 2–10. doi: 10.1080/17461391.2012.725102

Huijgen, B. C. H., Elferink-Gemser, M. T., Post, W., et al. (2010). Development of dribbling in talented youth soccer players aged 12-19 years: a longitudinal study. *J. Sports Sci.* 28, 689–698. doi: 10.1080/02640411003645679

Johnston, K., Wattie, N., Schorer, J., and Baker, J. (2018). Talent identification in sport: a systematic review. *Sports Med.* 48, 97–109. doi: 10.1007/s40279-017-0803-2

Kiphardt, E. J., and Schilling, F. (1970). *Body Coordination Test for Children (KTK). Manual.* Weinheim: Beltz.

Klein, M., Froehlich, M., and Emrich, E. (2012). On the precision of selected items of the German Motor Test 6-18 (Zur testgenauigkeit ausgewählter items des Deutschen Motorik-Tests DMT 6-18). *Leipziger Sportwissenschaftliche Beiträge* 53, 23–45.

Lago-Peñas, C., Casais, L., Dellal, A., Rey, E., and Domínguez, E. (2011). Anthropometric and physiological characteristics of young soccer players according to their playing positions: relevance for competition success. *J. Strength Condit. Res.* 25, 3358–3367. doi: 10.1519/JSC.0b013e318216305d

Le Gall, F., Carling, C., Williams, M., and Reilly, T. (2010). Anthropometric and fitness characteristics of international. professional and amateur male graduate soccer players from an elite youth academy. *J. Sci. Med. Sport* 13, 90–95. doi: 10.1016/j.jsams.2008.07.004

Leite, N., Baker, J., and Sampaio, J. (2009). Paths to expertise in Portuguese National team athletes. *J. Sports Sci. Med.* 8, 560–566.

Leyhr, D., Kelava,A., Raabe, J., and Hoener, 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:e0196324. doi: 10.1371/journal.pone.0196324

Leyhr, D., Raabe, J., Schultz, F., Kelava, A., and Hoener, O. (2020). The adolescent motor performance development of elite female soccer players: a study of prognostic relevance for future success in adulthood using multilevel modelling. *J. Sports Sci.* 38, 1342–1351. doi: 10.1080/02640414.2019.1686940

Li, P., de Bosscher, V., and Weissensteiner, J. (2018). The journey to elite success: a thirty-year longitudinal study of the career trajectories of top professional tennis players. *Int. J. Perform. Anal. Sport* 18, 1–12. doi: 10.1080/24748668.2018.1534197

Lovell, R., Towlson, C., Parkin, G., Portas, M., Vaeyens, R., and Cobley, S. (2015). Soccer player characteristics in English lower-league development programmes: the relationships between relative age, maturation, anthropometry and physical fitness. *PLoS ONE* 10:e0137238. doi: 10.1371/journal.pone.0137238

Meylan, C., Cronin, J., Oliver, J., and Hughes, M. (2010). Reviews: talent identification in soccer: the role of maturity status on physical. physiological and technical characteristics. *Int. J. Sports Sci. Coach.* 5, 571–592. doi: 10.1260/1747-9541.5.4.571

Mirkov, D. M., Kukolj, M., Ugarkovic, D., Koprivica, V. J., and Jaric, S. (2010). Development of anthropometric and physical performance profiles of young elite male soccer players: a longitudinal study. *J. Strength Condit. Res.* 24, 2677–2682. doi: 10.1519/JSC.0b013e3181e27245

Morris, T. (2000). Psychological characteristics and talent identification in soccer. *J. Sports Sci.* 18, 715–726. doi: 10.1080/02640410050120096

Murr, D., Feichtinger, P., Larkin, P., O'Connor, D., and Hoener, O. (2018a). Psychological talent predictors in youth soccer: a systematic review of the prognostic relevance of psychomotor, perceptual-cognitive and personality-related factors. *PLoS ONE* 13:e0205337. doi: 10.1371/journal.pone.0205337

Murr, D., Raabe, J., and Hoener, O. (2018b). The prognostic value of physiological and physical characteristics in youth soccer: a systematic review. *Eur. J. Sport Sci.* 18, 62–74. doi: 10.1080/17461391.2017.1386719

Musch, J., and Grondin, S. (2001). Unequal competition as an impediment to personal development: a review of the relative age effect in sport. *Dev. Rev.* 221, 147–167. doi: 10.1006/drev.2000.0516

Myer, G. D., Jayanthi, N., Difiori, J. P., Faigenbaum, A. D., Kiefer, A. W., Logerstedt, D., et al. (2015). Sport specialization, part I: does early sports specialization increase negative outcomes and reduce the opportunity for success in young athletes? *Sports Health* 7, 437–442. doi: 10.1177/1941738115598747

O'Connor, D., Larkin, P., and Williams, A. M. (2016). Talent identification and selection in elite youth football: an Australian context. *Eur. J. Sport Sci.* 16, 837–844. doi: 10.1080/17461391.2016.1151945

Pankhurst, A., and Collins, D. (2013). Talent identification and development: the need for coherence between research system and process. *Quest* 65, 83–97. doi: 10.1080/00336297.2012.727374

Papic, V., Rogulj, N., and Plestina, V. (2009). Identification of sport talents using a web-oriented expert system with a fuzzy module. *Expert Syst. Appl.* 36, 8830–8838. doi: 10.1016/j.eswa.2008.11.031

Patel, R., Nevill, A., Cloak, R., Smith, T., and Wyon, M. (2019). Relative age, maturation, anthropometry and physical performance characteristics of players within an Elite Youth Football Academy. *Int. J. Sports Sci. Coach.* 14, 714–125. doi: 10.1177/1747954119879348

Phillips, E., Davids, K., Renshaw, I., and Portus, M. (2010). Expert performance in sport and the dynamics of talent development. *Sports Med.* 40, 271–283. doi: 10.2165/11319430-000000000-00000

Pion, J. (2015). *The Flemish Sports Compass. From Sports Orientation to Elite Performance Prediction.* Ghent: University Press Zelzate.

Pion, J., Hohmann, A., Liu, T., Vandorpe, B., Lenoir, M., and Segers, V. (2016). Predictive models reduce talent development costs in female gymnastics. *J. Sports Sci.* 2016, 1–6. doi: 10.1080/02640414.2016.1192669

Pion, J., Teunissen, J. W., Ter Welle, S., Spruijtenburg, G., Faber, I. R., and Lenoir, M. (2020). "Chapter 13: how similarities and differences between sports lead to talent transfer: a process approach," in *Talent Identification and Development in Sport: International Perspectives*, 2nd Edn, eds J. Baker, S. Cobley, and J. Schorer (London: Routledge), 184–196.

Preacher, K. J., and Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behav. Res. Methods Instrum. Comput.* 36, 717–731. doi: 10.3758/BF03206553

Reilly, T., Williams, A. M., Nevill, A., and Franks, A. (2000). A multidisciplinary approach to talent identification in soccer. *J. Sports Sci.* 18, 695–702. doi: 10.1080/02640410050120078

Roescher, C., Elferink-Gemser, M., Huijgen, B., and Visscher, C. (2010). Soccer endurance development in professionals. *Int. J. Sports Med.* 31, 174–179. doi: 10.1055/s-0029-1243254

Rylander, P., Karlsteen, M., Kougioumtzis, K., and Gustafsson, J. E. (2019). The specificity versus generality of ball-handling skills - empirical evidence for a general ball-handling ability. *Hum. Mov. Sci.* 66, 477–486. doi: 10.1016/j.humov.2019.05.024

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

Schorer, J., Rienhoff, R., Fischer, L., and Baker, J. (2017). Long-term prognostic validity of talent selections: comparing national and regional coaches, laypersons and novices. *Front. Psychol.* 8:1146. doi: 10.3389/fpsyg.2017.01146

Schott, U. (2010). Von der Talentförderung profitieren das A-team und die Bundesliga (Talent promotion is beneficial for National Team and Bundesliga). *Fußballtraining* 28, 6–13.

Sieghartsleitner, R., Zuber, C., Zibung, M., and Conzelmann, A. (2018). "The early specialised bird catches the worm!" - a specialised sampling model in the development of football talents. *Front. Psychol.* 9:188. doi: 10.3389/fpsyg.2018.00188

Siener, M., and Hohmann, A. (2019). Talent orientation: the impact of motor abilities on future success in table tennis. *German J. Exerc. Sport Res.* 49, 232–243. doi: 10.1007/s12662-019-00594-1

Skorski, S., Skorski, S., and Faude, O. (2016). The relative age effect in elite German youth soccer: implications for a successful career. *Int. J. Sports Physiol. Perform.* 11, 370–376. doi: 10.1123/ijspp.2015-0071

Soberlak, P., and Cote, J. (2003). The developmental activities of professional ice hockey players. *J. Appl. Sport Psychol.* 15, 41–49. doi: 10.1080/10413200305401

Suppiah, H. T., Low, C. Y., and Chia, M. (2015). Detecting and developing youth athlete potential: different strokes for different folks are warranted. *Br. J. Sports Med.* 49, 878–882. doi: 10.1136/bjsports-2015-094648

Ulbricht, A., Fernandez-Fernández, J., Mendez-Villanueva, A., and Ferrauti, A. (2016). Impact of fitness characteristics on tennis performance in elite junior tennis players. *J. Strength Condit. Res.* 30, 989–998. doi: 10.1519/JSC.0000000000001267

Unnithan, V., White, J., Georgiou, A., Iga, J., and Drust, B. (2012). Talent identification in youth soccer. *J. Sports Sci.* 30, 1719–1726. doi: 10.1080/02640414.2012.731515

Utesch, T., Strau,ß, B., Tietjens, M., Büsch, D., Ghanbari, M.-C., and Seidel, I. (2015). Die Überprüfung der Konstruktvalidität des Deutschen Motorik-Tests 6-18 für 9- bis 10-Jährige. *Zeitschr. Sportpsychol.* 22, 77–90. doi: 10.1026/1612-5010/a000143

Vaeyens, R., Lenoir, M., Williams, A. M., and Philippaerts, R. M. (2008). Talent identification and development programmes in sport: current models and future directions. *Sports Med.* 38, 703–714. doi: 10.2165/00007256-200838090-00001

Ward, P., Hodges, N. J., Starkes, J. L., and Williams, M. A. (2007). The road to excellence in soccer: deliberate practice and the development of expertise. *High Abil. Stud.* 18, 119–153. doi: 10.1080/13598130701709715

Williams, A. M., Ford, P. R., and Drust, B. (2020). Talent identification and development in soccer since the millennium. *J. Sports Sci.* 38, 1199–1210. doi: 10.1080/02640414.2020.1766647

Williams, A. M., and Franks, A. (1998). Talent identification in soccer. *Sports Exerc. Injury* 4, 159–165.

Williams, A. M., and Reilly, T. (2000). Talent identification and development in soccer. *J. Sports Sci,*. 18, 657–667. doi: 10.1080/02640410050120041

Wong, P.-L., Chamari, K., Dellal, A., and Wisløff, U. (2009). Relationship between anthropometric and physiological characteristics in youth soccer players. *J. Strength Condit. Res.* 23, 1204–1210. doi: 10.1519/JSC.0b013e31819f1e52

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer* 3, 32–35.

Zibung, M., Zuber, C., and Conzelmann, A. (2016). The motor subsystem as a predictor of success in young football talents: a person-oriented study. *PLoS ONE* 11:e0161049. doi: 10.1371/journal.pone.0161049

Zuber, C., Zibung, M., and Conzelmann, A. (2015). Motivational patterns as an instrument for predicting success in promising young football players. *J. Sports Sci.* 33, 160–168. doi: 10.1080/02640414.2014.928827

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. *Front. Psychol.* 7:1088. doi: 10.3389/fpsyg.2016.01088

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Hohmann and Siener. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Artikel 5

Siener*, M., Ferrauti, A., & Hohmann, A. (2021). Early talent identification in tennis: A retrospective study. *International Journal of Racket Sports Science*, 3(2).

Int. j. racket sports sci. vol. 3(2), 2021, 26-38. eISSN: 2695-4508
DOI: 10.30827/ Original Investigation

Early talent identification in tennis: A retrospective study

Identificación temprana de talento en tenis: un estudio retrospectivo

Maximilian Siener [1*]; Alexander Ferrauti [2]; Andreas Hohmann [1]

1 Institute of Sport Science, Department of Training and Movement Science, University of Bayreuth, Bayreuth, Germany.
2 Faculty of Sport Science, Department of Training and Exercise Science, Ruhr University Bochum, Bochum, Germany.

Received: 25-06-2021
Accepted: 29-01-2022

Abstract

Talent identification often begins at the age of entry into a sport: even under the age of 9 years old (U9). However, the success of such early talent identification is questionable. Therefore, the aim of this long-term retrospective study is to examine whether today's more successful junior tennis players already differed from today's less successful junior tennis players in terms of physical fitness and motor competence when both groups were U9. If significant differences in performance characteristics between successful and less successful tennis players were already apparent at this young age, such characteristics could be used to forecast talent at an early stage. Based on their current tennis success, a total of 174 junior tennis players were divided into national *ranked players* (n = 16: players who achieved a place in the official national junior tennis ranking list of the German Tennis Federation) and *non-ranked players* (n = 158). All of these players had already participated in two anthropometric and nine physical fitness and motor competence tests at U9 (e.g., sprint, endurance run, ball throw). Using a MANCOVA and a correlation analysis, we retrospectively examined whether the two current performance groups had differed significantly in their U9 test scores and whether these athletes' U9 test performance scores correlated with their current playing success. No significant ($p < 0.05$) differences were found between ranked and non ranked junior players in terms of U9 body weight and height. However, with the exception of flexibility, all physical fitness tests and motor competence tests showed significant results. The ball throw was the most relevant test parameter, as it showed the highest prognostic validity (effect size $\eta^2 = .157$ and $r = .360$). This test was followed by the two test tasks standing long jump (effect size $\eta^2 = .081$ and $r = .287$) and endurance run (effect size $\eta^2 = .065$ and $r = .296$). Overall, the U9 findings are in line with the results from other studies of U12–U18 tennis players. Therefore, it can be assumed that talent specific characteristics remain stable over a certain period of time and that U9 test performances may provide an early indication of later playing success.

Keywords: *talent, tennis, performance testing, prognosis, success.*

Resumen

La identificación de talentos a menudo suele comenzar desde la edad de entrada al deporte, es decir, incluso por debajo de los 9 años (U9). Sin embargo, el éxito de dicha identificación de talento temprana es cuestionable. Por lo tanto, el objetivo de este estudio retrospectivo a largo plazo es analizar si los jugadores juveniles de tenis más exitosos hoy en día ya eran diferentes de los jugadores juveniles de tenis menos exitosos hoy en día en términos de aptitud física y competencia motora cuando los dos grupos eran U9. Si a esta temprana edad ya eran notables diferencias significativas en las características de desempeño entre los jugadores de tenis exitosos y menos exitosos, dichas características podrían usarse para predecir el talento desde una etapa temprana. Basados en su actual éxito en el tenis, un total de 174 jugadores juveniles de tenis fueron divididos en jugadores de clasificación nacional (n= 16: jugadores que obtuvieron un lugar en la lista oficial de clasificación nacional juvenil de tenis de la Federación Alemana de Tenis) y jugadores por fuera de la clasificación (n=158). Todos estos jugadores ya habían participado en dos pruebas antropométricas y nueve de aptitud física y competencia motora en U9 (ej. sprint, carrera de resistencia, lanzamiento de balón). Usando un análisis de correlación y el MANCOVA, analizamos retrospectivamente si los dos grupos de desempeño actual diferían

Correspondence author: Maximilian Siener, maximilian.siener@uni-bayreuth.de

Cite this article as:
Siener, M., Ferrauti, A., & Hohmann, A. (2021). Early talent identification in tennis: A retrospective study. *International Journal of Racket Sports Science*, *3*(2), 26-38.

This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

significativamente de los puntajes en sus pruebas en U9, y si esos puntajes de las pruebas de rendimiento en U9 de estos atletas se correlacionaban con su éxito de juego actual. No se encontraron diferencias significativas ($p < 0,05$) entre los jugadores juveniles dentro la clasificación y por fuera de ella en términos de peso y altura en U9. Sin embargo, exceptuando la flexibilidad, todas las pruebas de aptitud física y competencia motora presentaron resultados significativos. El lanzamiento de balón fue la prueba más relevante ya que demostró la mayor validez pronóstica (tamaño del efecto $\eta 2 = .157$ y $r = .360$). A esta prueba le siguieron dos actividades de prueba, salto de longitud de pie (tamaño del efecto $\eta 2 = .081$ y $r = .287$) y carrera de resistencia (tamaño del efecto $\eta 2 = .065$ y $r = .296$). En general, los resultados en U9 están en línea con los resultados de otros estudios de jugadores de tenis en U12-U18. Por lo tanto, puede asumirse que las características de talento específicas permanecen estables durante cierto periodo de tiempo y que el desempeño en las pruebas en U9 puede ser un indicador temprano de futuro éxito en el juego.

Palabras clave: *talento, tenis, prueba de desempeño, pronóstico, éxito.*

INTRODUCTION

With over 1.2 billion fans and more than 80 million players, tennis is one of the most popular sports in the world (International Tennis Federation, 2021). Tennis is also a highly complex sport that requires not only general endurance but also speed, agility, upper body power, and coordination (Filipcic & Filipcic, 2005; Kramer et al., 2017; Robertson et al., 2018). Apart from professional tennis players needing these performance prerequisites to defend their top positions in the world (Reid & Schneiker, 2008), it is also necessary to understand which performance characteristics are already relevant for tennis players at an early age. This may not only contribute to a successful future tennis performance (Till & Baker, 2020); it may also orientate a *best mover* to this particular sport according to the particular strengths of his or her individual performance profile. Importantly, the better the individual talent characteristics match the (future tennis) demands, the higher the chances that the beginners will achieve success and satisfaction in this complex sport. This assumption is underlined by Suppiah et al. (2015), who state that a wrong choice can never be compensated for by training. Engaging in an unsuitable sport might not only be detrimental to fun but also lead to drop outs ahead of time. Conversely, if children like a recommended sport, the talented athletes could transform their physical, physiological, and psychological gifts through a long term process of diligent learning, deliberate practice, and an extended amount of high quality training into optimal achievements (Davids & Baker, 2007; Pion, 2015).

With the help of physical fitness and motor competence tests, young athletes are analyzed and their future potential is assessed (Kramer, Huijgen, Lyons, et al., 2016; Ulbricht et al., 2015). These so called sports orientation and talent identification (TID) campaigns have become increasingly common in recent years (Johnston et al., 2018). This is also due to national sports organizations investing more and more effort into the systematic identification of talented young players. In a professionalized competitive environment, a relaxed approach no longer appears acceptable (De Bosscher et al., 2008), and talent identification could become the key to national elite sport performance. Therefore, TID is designed to identify promising young athletes at an early stage (Hohmann & Seidel, 2003; Pion, 2015). Thus, the testing and scouting of athletes begin as soon as they enter a sport. On average, the general sport entry age of professional athletes is about 9 years (8.5 ± 2.5 years: Güllich & Emrich, 2014; 9.1 ± 3.7 years: Vaeyens et al., 2009). In this context, tennis and other racket sports are known to have an early starting age (5–8 years) (Faber et al., 2016). Studies by Li et al. (2020) have shown that 75% of all Top 300 tennis players began playing tennis between 3 and 7 years of age, and only 4% of the Top 300 tennis players started after the age of 10. This might be related to the fact that, firstly, the former group of athletes had more time to freely gain competitive experience, and secondly – particularly before puberty – there are sensitive learning phases that promote motor learning and thus offer the opportunity for the acquisition of technical skills (Knudsen, 2004). In addition, a tennis education starting at a young age gives the coaches a longer observation period; this reduces *talent selection* errors during early adolescence and enhances practitioners' talent identification decisions. Accordingly, it is not surprising that the evaluation of young athletes begins when they are under nine years (U9) of age (Potočnik et al., 2020; Tomkinson et al., 2017). However, because in many cases these young athletes have only little technical experience so far, these TIDs often consist of several generic test items rather than specific, more technically demanding skill tests (Hohmann et al., 2018; Niessner et al., 2020). This approach is supported by studies by Faber et al. (2020), who found a small but significant correlation between more sport specific, coordinative technical skills (e.g., speed while dribbling or aiming at a target) and previous training hours in racket players aged between 8 to 10 years old. In contrast, generic tests such as sprinting or standing long jump had no significant relationship with training volume.

While there are already some studies on the success of TID in other sports, such as soccer (see Sarmento et al., 2018), in tennis it is still questionable whether TID campaigns in the U9 age group (Hohmann et al., 2018; Pion, 2015) can provide any information

whatsoever about later tennis performance. Although there are many studies on TID in junior tennis (U12–U18) and in professional tennis (Baiget et al., 2016; Van Den Berg et al., 2006), there are either no studies or hardly any on younger age groups (e.g., U9). This might be related to the fact that for this young age group, which lacks a developed tournament and ranking system, no proper differentiation of better and weaker performance groups can be established (Siener & Hohmann, 2019). The assessment of different tennis performance groups usually arises with the entrance into the junior ranking system of the U12 and is then often determined by the ranking position (Ulbricht et al., 2016). For this reason, it has been common practice to use as a template for TID in youth professional adult tennis players' profiles, collected from cross sectional studies (Hohmann & Seidel, 2003). This assumes that the same invariant skills are crucial for tennis success in both age groups. However, this has not yet been adequately demonstrated by long term studies (Baker et al., 2020). Therefore, for a prognostically valid evaluation of U9 test performances, either a prospective study design must be employed or these test performances must be considered retrospectively (see also Mostaert et al., 2020; Till et al., 2015). Either way, both approaches are highly time consuming and only possible in a longitudinal study.

The present study starts at the earliest possible point in time at which a distinction between different performance groups in tennis can be made. From U12, tennis players are able to qualify for the national junior tennis rankings of the German Tennis Federation and thus stand out from other athletes. These junior tennis rankings cover the age range up to U18 (Deutscher Tennis Bund [DTB], 2020). Within the group of junior tennis players, two performance groups – junior ranked players (RPs) and non ranked players (NPs) – can be distinguished. Based on these two performance groups, it is possible to retrospectively analyze the performance shown in the physical fitness and motor competence tests of the U9. Thus, the aim of this long-term retrospective study was to compare the initial U9 physical fitness and motor competence test performances of today's junior ranked and non–ranked tennis players from the national tennis ranking list of the German Tennis Federation (U12–U18) and, based on this, to address whether TID in U9 seems at all possible. Overall, this study makes the first attempt to close the gap in research on TID at U9 and to investigate the validity of talent prediction by physical fitness testing at this young age. If it transpires that at this young age, there are already significant differences in performance characteristics between later successful and less successful tennis players, these characteristics can be used to make a talent forecast for players when they are only 8 years old.

MATERIALS AND METHODOLOGY

General study design

In this retrospective (long-term) study, we examined whether there is a relationship between the playing success achieved by junior tennis players (U12–U18) and this cohort's general childhood performance (U9). For this purpose, 174 junior tennis players were divided into two groups based on their current playing success: (A) 16 players who have achieved a position in the official national tennis ranking lists of the German Tennis Federation (RPs); and (B) 158 players who have not achieved a position on the rankings (NPs).

All these players (A+B) had already been tested in the U9 category in two anthropometric tests, six physical fitness tests, and three motor competence tests (Fig. 1). This now allows us to analyze whether the two performance groups had already showed differences in their performance characteristics at the age of 8 and, if so, in which generic tests the differences were particularly evident. The results can then indicate whether certain characteristics/predictors detectable as early as U9 may point to later success at junior age and thus make TID worthwhile at this earlier stage.

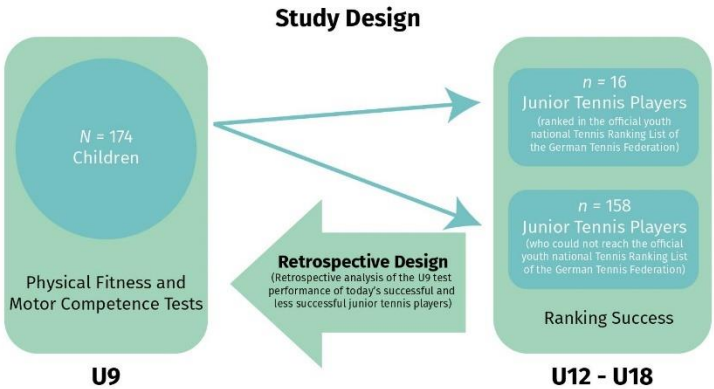

Figure 1. Study Overview

Participants

A total of 174 junior tennis players (U12–U18; *M* = 156.3 months, *Min* = 132 months, *Max* = 206 months) were included in this study. The sample consists of 62 girls (♀) and 112 boys (♂). Among these junior tennis players, 16 RPs (n♂ = 11, n♀ = 5) achieved a place in the official national junior tennis ranking list of the German Tennis Federation (DTB, 2020). To be included in this junior ranking, a player must have at least 10 wins in junior ranking matches or have already earned a position in the adult rankings. All other junior tennis players (NPs; n = 158, n♂ = 101, n♀ = 57) were registered in clubs and actively participated in regional and local club competitions and tournaments but did not achieve the necessary number of wins in official ranking matches to earn a place on the official national junior ranking list of the German Tennis Federation.

As part of the study, all participants and their parents were fully informed about the content of the testing and the resulting studies, and their consent was obtained before the study began. The study and research design were in line with the Declaration of Helsinki and was approved by the ethics committee of the Municipality of Fulda (Germany), as well as the State Office of School Education of Fulda (Germany).

Measurements

Physical Fitness and Motor Competence Tests of U9 Testing

All junior tennis players were already tested at U9 with a test battery of two anthropometric (body weight and height) tests; six physical fitness tests (sprint, flexibility, arm and upper body strength, leg power, and endurance performance); and three mixed motor competence tests (coordination, balance, and ball throw performance). Each of the standardized tests was performed according to existing test protocols, which include a detailed description of the test items, the exact test set up, the demonstration of the test item, the execution of the test phases, and the measurements (Bös et al., 2009; Siener et al., 2021):

20-m Sprint

The sprint performance was recorded using a 20-m sprint (~21.9 yards). In each of the two possible attempts, the test persons started 0.3 m before the starting line. The time was stopped by means of light gates (Brower Timing Systems; Draper, USA). The test met an objectivity value of .86 and a reliability value of .96 (Bös et al., 2009).

Sideward Jumping

The number of two-legged jumps that a participant could perform between two 50-cm x 50-cm (1 cm ≈ 0.4 inches) squares within 15 s was measured. Only jumps where none of the boundary lines was touched counted. In total, two attempts were made, with a break of at least 2 min. between each attempt. The mean value of both tests was used for further calculations. The objectivity of this test is .99 and the reliability is .89 (Bös et al., 2009).

Balancing Backwards

The participants balanced backwards on a 6-cm, 4.5-cm, and 3-cm wide wooden beam. Two attempts were made on each beam, and the number of steps (feet fully raised) before leaving the beam was counted. A maximum of eight steps/points could be achieved per beam, so that the total maximum number of points for this test task is limited to 48. The test achieved an objectivity of .99 and a reliability of .73 (Bös et al., 2009).

Standing Forward Bend

In this flexibility test, the participants have to try to reach as low as possible with their hands. The ground height is evaluated as 0 cm, and everything that goes beyond that (below ground level) is entered as a positive value. The achieved value had to be held for at least 3 sec. in each of the two attempts. The test achieved an objectivity of .99 and a reliability of .94 (Bös et al., 2009).

Push-Ups

In this test, the participants had to perform as many push-ups as possible over 40 sec. An execution was only evaluated if the technique was correct and the test person subsequently lay back down in the starting position. Only one attempt was performed. Bös et al. (2009) rated the objectivity with .98 and the reliability with .69.

Sit-Ups

The sit-up test also evaluated the number of correctly performed sit-ups in 40 sec. Only one attempt was performed. The objectivity of this test was .92 and the reliability was .74 (Klein et al., 2012).

Standing Long Jump

In the standing long jump test, the jumping distance was measured in centimeters. Each test person had two attempts, of which the better attempt was recorded. Between both attempts, a complete break was ensured. The test achieved an objectivity of .99 and a reliability of .89 (Bös et al., 2009).

Ball Throw

In the ball throw test, the throw distance was measured in centimeters orthogonal to the line of

release. The test persons threw with an 80-gm ball from a standing position three times in a row. As in the previous tests, the best value was evaluated. The test could be evaluated in our own studies (n = 1800) with a reliability of .77.

6 min Endurance Run

In the endurance test, the participants tried to run as many laps as possible around a 9-m x18-m volleyball field in 6 min. The achieved distance was noted in meters. The test was carried out by a total of 15 people at the same time. The objectivity of this test is .87 and the reliability is .92 (Bös et al., 2009).

All tests were conducted by qualified personnel during regular school hours (8–12 am), and a uniform warm-up was held before starting. The 6-min endurance run was always the last test in the test series.

STATISTICAL ANALYSES

For all analyses, the software SPSS (version 26; SPSS Inc, Chicago, IL, USA) was used.

A univariate ANOVA found that age affects the data and that the performance of U9 players increased significantly with age. In order to avoid this age bias, bivariate regressions to age in months were used to z standardize the test value residuals separately for both genders (Siener et al., 2021). Since this procedure poses the risk that in homogeneous groups certain test values are distorted by the group composition, the results of about 4000 children predominantly not from sporting clubs (see Hohmann et al., 2018; Tomkinson et al., 2017) were additionally used for z standardization. Thus, all data are available as age and gender independent z values. To ensure a better comparison of the sprint data, the z values were additionally multiplied by "-1", thus turning the better sprint results into positive z values.

T-Tests and a MANCOVA were used to check whether the test results differed significantly between the gender and tennis success groups. A covariate weight was chosen. Effect sizes for *partial eta squared* (η^2) smaller than 0.01 are interpreted as trivial, effect sizes between 0.01 and 0.059 are small, between 0.06 and 0.139 are moderate, and values higher than 0.14 are large.

To gain a better insight into the influence of the individual test items on later tennis success, *bivariate correlations* were also calculated. The Spearman correlations were classified according to the following pattern: trivial (0 – 0.1), small (0.1 – 0.3), moderate (0.3 – 0.5), and large (0.5 – 0.7).

RESULTS

The initial U9 test results of the junior tennis players show that in almost all test tasks, RPs perform significantly better than NPs (Table 1). Particularly noteworthy here are the standing long jump, the sideward jumping, the balancing, the endurance run, and the ball throw, which all have values of p < 0.001 (t-Tests). In sideward jumping, the maximum value of the RPs was five jumps (12.5%) higher than the maximum value of the NPs. In the ball throw and the standing long jump, the later better athletes had a clear advantage in their youth. In both tests, approximately 84% of the NPs did not reach the average test result of RPs. In the sideward jumping, the balancing, and the endurance run, 84% of the RPs were also above the average result of the weaker performance group. For the forward bend test task (p = 0.158) and body weight (p = 0.910), no significant differences in the two performance groups could be found. In addition, the later ranked players are on average 4 months older than their weaker tennis colleagues (p < 0.01).

In examining the raw scores of the two performance groups separately by gender, it is notable that no significant group difference can be found for boys (n♂ = 112) in the test scores for body weight (p = 0.149), height (p = 0.8), sprint (p = 0.106), push-ups (p = 0.065), and forward bends (p = 0.111). All other test results showed significant group differences (p < 0.01). *Girls* (n♀ = 62) showed comparable results. Only the test results for body height (p = 0.044) and sprint (p = 0.013) were also significant in contrast to the results for boys.

Also, after eliminating the age effect, RPs achieved better z-values than the NPs in all test items (Fig. 2). Except for the 20-m sprint, however, the ascending order of the single tests was almost identical in both groups. Also, in regard to the order of the RPs' tests, the ball throw (M_z = 2.03) was in first place, followed by the standing long jump (M_z = 1.53), the endurance run (M_z = 1.42), and sideward jumping (M_z = 1.41). For the ball throw, female tennis players (M_z = 2.95, n = 5) performed significantly better (p < 0.05) than their male counterparts (M_z = 1.61, n = 11). The same holds true for body height (M_z♀ = 0.92, M_z♂ = 0.07) and the standing long jump (M_z♀ =1.88, z♂ = 1.37). The anthropometric measures of the RPs were M_z = 0.34 (body height) and M_z = −0.30 (body mass). RPs at U9 were, therefore, on average slightly taller and lighter than the NPs. However, while in a t-test comparison, significant differences between the two performance groups can be found in almost all generic test items (with the exception of $p_{\text{Forward bends}}$ = 0.086), there are no significant differences in the anthropometric test values of body mass (p = 0.078) and height (p = 0.234).

Table 1.
– Descriptive statistics for the former U9 test results of the junior tennis players.

	Groups	N	M	SD	95% CL		Min	Max	*p*
					LL	UL			
Calendar age (months)	NPs	158	93.8	5.0	93.0	94.6	83	110	0.007
	RPs	16	97.5	6.4	94.1	100.9	88	112	
Test results									
Body height (cm)	NPs	158	129.1	5.7	128.2	130.0	117	145	0.034
	RPs	16	132.3	4.5	129.8	134.7	127	143	
Body mass (kg)	NPs	158	27.2	4.1	26.6	27.9	20.0	39.3	0.910
	RPs	16	27.3	1.9	26.3	28.3	23.4	30.6	
Sideward jumping (repeats)	NPs	157	27.3	6.1	26.3	28.2	6.5	40.5	0.001
	RPs	16	33.0	5.1	30.3	35.7	27.0	45.0	
Balance backward (steps)	NPs	158	30.3	8.5	28.9	31.6	8	48	0.001
	RPs	16	38.3	5.7	35.2	41.3	28	48	
Standing long jump (cm)	NPs	157	135.1	16.2	132.5	137.6	82	190	0.001
	RPs	16	153.0	17.2	143.8	162.2	125	178	
20 m sprint (s)	NPs	158	4.45	0.36	4.39	4.51	3.10	5.32	0.011
	RPs	16	4.21	0.34	4.03	4.39	3.50	4.72	
Push ups (repeats)	NPs	158	14.6	3.6	14.0	15.1	4	24	0.041
	RPs	16	17.1	5.2	14.3	19.9	9	25	
Sit ups (repeats)	NPs	158	19.2	5.1	18.4	20.0	2	30	0.002
	RPs	16	23.4	4.1	21.3	25.6	15	29	
Forward bends (cm)	NPs	158	1.98	5.94	1.05	2.92	-11	18	0.158
	RPs	16	4.16	4.90	1.55	6.77	-10	12	
6 min run (m)	NPs	154	959	130.8	938	979	545	1259	0.001
	RPs	16	1078	80.7	1035	1121	891	1200	
Ball throw (m)	NPs	155	13.5	4.03	12.9	14.2	3.8	27.6	0.002
	RPs	16	18.8	5.35	15.9	21.6	9.2	28.3	

M = mean; SD = standard deviation; SE = standard error; CL = confidence limit; LL = lower limit; UL = upper limit; Min = minimum; Max = maximum; NPs = non-ranked players; RPs = ranked players; p = *p*-value of the *t*-test

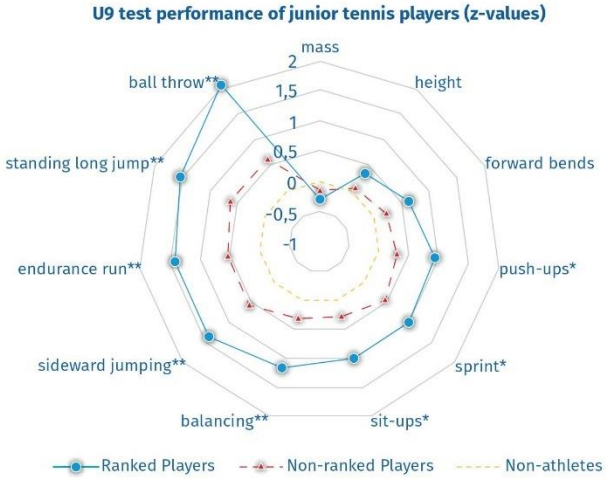

Figure 2. Initial U9 test performances (z-values) of junior ranked tennis players, non- ranked tennis players, and non-athletes (* p ≤ 0.05; ** p ≤ 0.01).

The MANCOVA (Table 2) shows that body mass only had a significant ($p < 0.05$) influence on the test values of balancing, sit ups, and the endurance run. Gender only had a significant influence on the ball throw in the analyses ($p_{Gender} = .012$; $p_{Performance*Gender} = .056$). All other hypotheses concerning a difference between the genders had to be rejected. As already suspected in *Fig. 2*, a *partial Eta squared* of $\eta^2 = .157$ shows the strongest effect size of group differences in the ball throw. The standing long jump, the endurance run, and sideways jumping show moderate effect sizes, while all other effects are small. It is notable that the standard deviation (SD) for the RPs fluctuates strongly. For the NPs, on the other hand, the SD of almost 1, which is usual for z values, is achieved.

Looking at the box plots of RPs and NPs for the different test tasks (*Fig. 3*), it is notable that the RPs have the biggest advantage, especially in the ball throw. Nevertheless, not all of the 16 RPs can show very high values. There is also one athlete with a *z*-value of $z = -1.2$ and three athletes with z values below $z = 1.0$ (range of 6.06). All other RPs show above-average throwing performances. Such large fluctuations in performance cannot be seen in test items the endurance run (range of 2.3) and balancing (range of 2.16). Here the results are comparatively close to each other, and none of the RP values is below a *z*-value of $z = 0$. Also, in body mass, no major fluctuations can be detected in the RPs (range of 1.15). The results show that in the ranked player group 75% (12/16) of the tennis players at the age of 8 yrs achieved a very good test result ($z \geq 1.0$) in the standing long jump, the endurance run, and the ball throw. In addition to the three tests mentioned, more than 50% of the ranked junior tennis players were initially (when U9) able to achieve a very good test score in the sideward jumping (68.8%) and sit ups (56.3%).

Table 2.
– Results of the MANCOVA for the different former U9 test disciplines

	NPs (Mean ± SD)	RPs (Mean ± SD)	F	Sig.	Partial Eta Squared (η2)
Ball Throw	0.639 ± 0.999	2.033 ± 1.553	29.943	.0001	.157
Standing Long Jump	0.634 ± 0.874	1.532 ± 0.906	14.175	.0001	.081
Sideward Jumping	0.558 ± 0.949	1.407 ± 0.773	11.401	.001	.066
Endurance Run	0.537 ± 0.975	1.423 ± 0.564	11.239	.001	.065
Sit-Ups	0.281 ± 0.908	1.010 ± 0.710	9.697	.002	.057
Balancing	0.300 ± 0.938	1.157 ± 0.605	9.542	.002	.056
Sprint	0.440 ± 0.912	0.987 ± 0.851	5.468	.021	.033
Push-Ups	0.288 ± 0.927	0.924 ± 1.360	5.090	.025	.031
Forward Bends	0.198 ± 0.952	0.634 ± 0.564	2.548	.112	.016

SD = standard deviation; NPs = junior non ranked tennis players; RPs = junior ranked tennis players; Effect sizes (η2): 0.01 ≤ small, 0.06 ≤ medium, 0.14 ≤ strong

Apart from the forward bends, no significant results can be seen for the correlations of tennis ranking success with body height or body mass (Table 3). All other test values achieved significant correlations, mostly in the moderate range.

The ball throw reaches the highest correlation value with r = .360, followed by the standing long jump ($r = .287$), and the endurance run ($r = .296$). However, these two values cannot be considered a moderate result. Also, all other test values have small correlation effects.

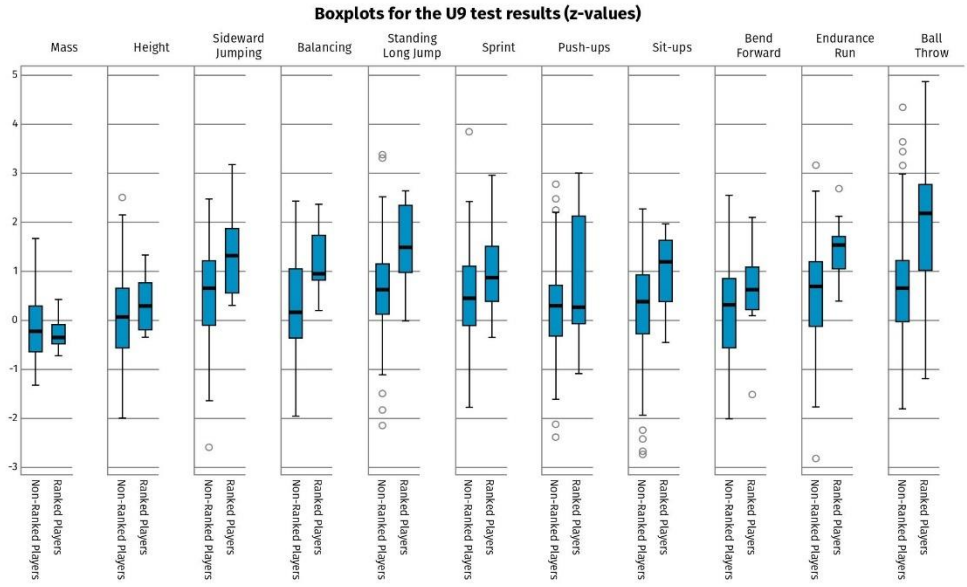

Figure 3. Boxplots of the former U9 test results (z-values) of junior non-ranked tennis players and ranked tennis players.

Table 3.
– Bivariate correlations of the different U9 test results and the chance to achieve a ranking position in the official youth national ranking lists of the German Tennis Federation (p ≤ 0.05; ** p ≤ 0.01).*

	Mass	Height	Sideward Jumping	Balancing	Standing Long Jump	Sprint	Push-ups	Sit-ups	Forward Bends	Endurance run	Ball Throw
Tennis Ranking Success	-.077	.091	.256**	.263**	.287**	.226**	.187*	.231**	.130	.296**	.360**

DISCUSSION

The aim of this study was to investigate whether TID in tennis is already possible in the U9 age group and whether physical fitness tests and motor competence tests provide an early indication of later junior success. For this purpose, a long-term study retrospectively investigated the initial U9 test performances (based on physical fitness and motor competence tests) of today's junior tennis players in the official junior ranking list of the German Tennis Federation, as well as the weaker tennis players without a successful placement in the ranking list. If measurable differences in certain physical fitness and motor competence characteristics/predictors between these two current junior performance groups can already be identified at U9, then conversely, a prediction of later junior success can also be prospectively made at U9 on the basis of these characteristics, and the validity of early TID can be evaluated.

It was shown that the group of junior ranked tennis players were already significantly superior to the group of non ranked players in almost all test items at the age of 8 (U9). Looking at the results of the U9 tests more closely, it is noticeable that the differences are particularly high in test tasks the ball long throw, the endurance run, and the standing long jump, while U9 body height and body weight did not show any significant difference between the two groups of junior tennis players.

In *Table 1*, it can be assumed that the RPs as children (U9) had a clear advantage over their peers. The mean values of all test items are higher for the later group of more successful athletes than for the later group of NPs. With the exception of body weight and flexibility, the raw values already show significant differences between the two performance groups. However, the values still cannot be interpreted well, as age can also have an effect on performance. Considering the age in months, it is also notable that the later RPs are, on average, about 4 months older than the NPs. This is consistent with various studies on the *relative age effect*, according to which older athletes of a given age group have an advantage over their younger peers (Musch & Grondin, 2001). According to the data, this advantage does not seem to vanish even after about 5 years (Smith et al., 2018). Also, comparing the results of similar tests by Tomkinson et al. (2017) and Bös et al. (2009), it is of note that the participants tested in this study performed better in certain test items than the comparison groups of the other studies. Therefore, it can be assumed that regional differences influence performance. For this reason, too, it makes sense to standardize the **raw scores**. For the *z*-value standardization, in addition to the gender and age in months of the tennis players, the data of about 2000 non athletes of the same region were also used. This prevents regional test bias from blurring the results (Hohmann et al., 2018). In addition, **male and female** athletes can be evaluated together, as is usual in talent identification campaigns (Pion, 2015). While the raw scores of female and male tennis players still differ significantly in the four test tasks the sprint, the forward bends, the endurance run, and the ball throw, the MANCOVA of *z*-values only shows significant differences between the genders in the ball throw, with girls scoring comparatively better than boys. Overall, however, it must be noted that with this exception, the sample is still relatively homogeneous at this age (Siener & Hohmann, 2019).

However, the *z*-standardization to the non athletes also has the disadvantage that the *z*-values are comparatively higher than usual. This is especially noticeable in the **ball throw**, where individual tennis players can show *z*-values of $z = 4$ and more. Therefore, the results must always be seen in relation to the regional non athletes. Nevertheless, the differences in the individual test items now can be better seen. In this study, the ball throw is also the most important U9 test value. Both the NPs and RPs achieved the highest average test values of $z = 0.64$ and $z = 2.03$. It is not surprising that the ball throw, as a combined exercise of technique and strength, has such a high influence on subsequent success in tennis, since the two movements, the tennis serve and throwing abilities, have similar characteristics (Fett et al., 2020). The serve is of particular importance in tennis, as it lays the foundation for the rally and can put pressure on the opponent right from the beginning to achieve a point win. As early as 1992, Roetert et al. could prove a significant correlation between national tennis success and striking speed (correlations: $r_{forehand} = 0.68$, $r_{backhand} = 0.59$, $r_{serve} = 0.57$). Similar results have been shown in recent studies by Ulbricht et al. (2016) on the influence of fitness characteristics on tennis

performance. They showed that in 902 tennis players aged from 11 to 16 years, serve velocity (correlations: r♀ = −0.43 to −0.64, r♂ = −0.33 to −0.49) and upper body power (medicine ball throw; r♀ = −0.26 to −0.49; r♂ = −0.20 to −0.49) have the greatest impact on tennis performance. We were able to demonstrate a significant correlation between the ball throw and the RP of r = 0.360 (p < 0.01), which is comparable to the studies described above. *Table 2* also shows the highest effect size for the ball throw, with a *partial eta squared* of η^2 = 0.157. This is the only one of the various test items that can demonstrate a strong effect. A high influence on tennis performance could be proven not only for the serve but also for the upper body strength in general. In cross-sectional studies by Kramer et al. (2017) on 86 Dutch junior elite tennis players, a significant correlation between the combined upper body strength abilities of the ball throw and two medicine ball throws with the ranking of the male U13 players (Pearson correlation: r = −0.5*) was demonstrated. This is supported in studies by Fett et al. (2017) on U16 *Davis Cup* players and regional squad tennis players, who demonstrated a large effect size (*Cohen's d* = 1.04) for the throwing performance (forehand medicine ball throw) in *t*-tests.

In college tennis competitions, Kovacs (2007) measured an average playing time of about 1.5 hr. With such a long game duration, it is not surprising that due to exhaustion, hitting accuracy can drop as low as 81% (Davey et al., 2002). Thus, **endurance** has an important impact on tennis success and in our study reaches a moderate effect strength of η^2 = 0.065. Box plots also show that the range of results is relatively small, which is also shown in *Table 2* in SD = 0.564. A certain level of endurance seems to be indispensable for achieving rankings. For example, in a study with 40 male RPs (ranked 1 to 40) at the age of 15 years, a 20-m shuttle run (number of laps) correlated significantly with the achieved ranking (Meckel et al., 2015). However, the achieved correlation of r = −0.581* is much higher than the correlation value of r = 0.296** proven here. In addition to Meckel and colleagues (2015), other authors demonstrated the importance of endurance performance for tennis success in the U15 category (Filipcic et al., 2010). While in older athletes good endurance, due to the shifting of the fiber distribution, can have a rather negative effect on sprint ability, at the age of 8 years a significant correlation between the two abilities is still evident (r = 0.331**). Nevertheless, the influence of **sprinting** is weaker compared to endurance and has a correlation value of only r = 0.226** (Sprint–Tennis Ranking Success). The *partial eta squared* also turns out to be only a *small* effect, with η^2 = 0.03. This seems surprising at first sight, since studies by Girard and Millet (2009) on male U15 tennis players showed correlation values between the 20-m sprint performance and the ranking of r = 0.74**. However, their study was based on a small sample, with only 12 participants. Nevertheless, Filipcic et al. (2010) also demonstrated similar results in a larger sample (N = 159, U16–U19). In investigations by Ulbricht et al. (2016), correlations of r♂ = 0.31** and r♀ = 0.19* were found in the U14 category (N = 431), but in contrast to the serve velocity (r = 0.33 to 0.64 of U12–U16), these values were rather low. Also, the effect size between national and regional athletes was mostly small (d_{Cohen} = 0.00 to 0.21 for U14–U16; exception: d_{Cohen} = 0.63** for ♂ U12). Ulbricht explains the low significance of the classic linear sprint tests by the fact that typical movements in tennis are limited to a radius of 3 to 4 meters. The maximum speed would therefore never be reached. This fact is even more relevant for the U9 players, where the 20-m sprint is more dependent on speed endurance than in older research groups. Therefore, at a second view, the weak performance of the 20-m sprint is understandable.

The two test tasks **sideward jumping** and **standing long jump** have moderate effect size of a *partial eta squared* of 0.066 and 0.081. The standing long jump is also in second place in the overall U9 motor skills profile, and both RPs and NPs were able to achieve very good values at a young age. The explosive power of the leg muscles is also of great importance for short fast starts and quick first steps. Both exercises were investigated in a series of agility tests (see Girard & Millet, 2009; Ulbricht et al., 2016). Roetert et al. (1992) was one of the first to point out the significant correlation between the hexagonal test and the tennis ranking position. Although a classical agility test is missing in our study, the high value in the standing long jump may be an indication of corresponding abilities in tennis.

None of the RPs was extremely light or heavy at the age of 8. Accordingly, no significant results could be found in **body mass** comparing the two performance groups. Also, the **body height** did not lead to a significant result (*t*-test: p = 0.24). This is surprising, since a larger person has an advantage over smaller persons in the serve and also in the range of covered space on the court. Due to the higher hitting position of the ball, the relative field size to be hit in the opponent's field is larger than from a lower hitting position of the ball (Vaverka & Cernosek, 2013). A higher hitting position also allows the ball to be played at greater speed, so it should not be surprising that coaches give larger athletes a higher chance of success (Robertson et al., 2018). Nevertheless, the height of the junior tennis players tested in the U9 had no direct influence on their tennis ranking success.

Overall, U9 results are comparable to results from other performance studies of U11–U16 tennis players (Fett et al., 2017; Ulbricht et al., 2016). Therefore, it can be assumed that talent specific characteristics remain stable over time (Hohmann et al., 2018) and that athletes who tend to perform better later on already emerge in U9. However, the results cannot

consistently confirm the prognostic validity of TID in U9, as a high association with later ranking success could be found in only a few test items. Future TID predictions could therefore benefit from more test items or holistic talent assessments (e.g., motivation; Zuber et al., 2016). Nevertheless, for coaches, the results demonstrated can serve for a first cautious assessment of their training groups.

LIMITATIONS

The **test battery** used here covers a wide range of generic motor tests and almost all basic abilities (speed, strength, flexibility etc.). These tests intentionally correspond to the physical fitness and motor competence tests often used in TID campaigns in Europe (Niessner et al., 2020; Pion, 2015; Potočnik et al., 2020). Nevertheless, it should be noted that the results of the study shown here also depend to a large extent on the tests used. According to the retrospective design, the U9 tests here were already given up to 9 years ago; accordingly, it was not possible to draw on the latest test developments of recent years (Faber et al., 2017; Faber et al., 2018; Fernandez-Fernández et al., 2014). However, although Koopmann et al. (2020) were able to highlight the use of sport specific tests in the context of TID for the junior level, it is questionable whether similar successes in TID can be already predicted in U9 through sport specific tests in tennis. This is because the use of sport specific tests can quickly overburden children without many years of sport specific technical experience and, accordingly, would presumably over recommend early specialized athletes. However, specialization as early as 8 years of age is at the same time considered critical in the eyes of many scientists (LaPrade et al., 2016). Therefore, for future testing of the U9 players, sport-specific tests should only be used to a small extent. Nevertheless, the study shown here could have benefited from an additional *change of direction agility test, a stroke velocity test, or* a *handgrip test* (Ulbricht et al., 2016). However, tennis success does not only depend on physical fitness. Zuber et al. (2016) have therefore made initial attempts to integrate *psychological tests* into TID campaigns. In the future, it will be interesting to see how TID develops further.

The small **sample size** is another limitation of the study. With only 174 participants and also only 16 ranked players, the sample is very small. The results shown here are therefore not generalizable and must always be interpreted in relation to group size. Following Gagné (2010), only about every 10th athlete can be described as talented. Accordingly, if a group of 50 ranked players were to be studied, a total sample of approximately 500 participants would be needed. However, investigating such large sample sizes is highly time-consuming and cost-intensive, especially for long-term studies, which is why only a few studies have made this effort (Bergkamp et al., 2019; Gonaus & Müller, 2012; Höner et al., 2017; Höner & Votteler, 2016).

Another problem could be the joint consideration of genders. In Germany, it is common to train girls and boys together at a young age and thus to screen them together in the initial TID. However, various studies (Fernandez-Fernández et al., 2014; Kramer, Huijgen, Elferink-Gemser, & Visscher, 2016; Sannicandro et al., 2012; Ulbricht et al., 2016) suggest that boys' physical fitness and motor competence (in relation to tennis) may differ from that of girls. Usually, these differences do not emerge until puberty (Grosser et al., 2008). In the U9 sample used here, no significant differences in age-adjusted z-scores between the genders were found except for ball throwing (where ranked girls were slightly better than ranked boys). Nevertheless, even the same test values could have a different effect for the different genders later. However, our calculations showed that the results for test items with the strongest effect values were almost identical for both genders. Unfortunately, due to the even smaller sample size (especially for the 62 girls), these results are difficult to evaluate. Therefore, in future studies, the differentiation of the genders should be considered again more intensively.

CONCLUSION

This long-term retrospective study shows that TID appears possible even at U9 and that early test performance can be used to predict later junior tennis success. Today's junior tennis players in the national tennis ranking list of the German Tennis Federation were already superior in U9 to current weaker junior players in almost all early test items. This is especially evident in throwing performance, endurance running, and the standing long jump. Throwing performance is the most important factor. In all analyses (MANCOVA and correlation), it was shown to have the greatest influence. All other test items could only show medium to small effect sizes. Retrospective body height and weight at the time of U9 testing did not show any difference between the groups. Nevertheless, the study demonstrates that a valid talent prognosis seems to be possible even at the U9 level.

PRACTICAL APPLICATION

The results show that TID can be successful as early as U9. Coaches should particularly focus on throwing power in their search for talent, as this has the greatest effect on reaching the national youth rankings later on. Namely, every second person with a throwing performance of $z \geq 2$ reached the rankings. In addition to throwing power, endurance and jumping power were also good indicators of later success. In both parameters, with one exception, later RPs reached above-average values ($z \geq 0$). However, the results also

show that even weaker athletes ($z_{ball\ throw}$ = −1.3) can still achieve success. Therefore, physical fitness and motor competence should not be used alone for TID but should be considered in addition to the coach's opinion.

CONFLICT OF INTERESTS

The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript. Also, the authors did not conduct any studies with animal participants for this article. As for other studies citied in this article, information on ethical guidelines may be found in the respective sources. All experiments comply with the current laws of the country in which they were performed.

REFERENCES

Baiget, E., Iglesias, X., & Rodríguez, F. A. (2016). Aerobic fitness and technical efficiency at high intensity discriminate between elite and subelite tennis players. *International Journal of Sports Medicine, 37*(11), 848-854. https://doi.org/10.1055/s-0042-104201

Baker, J., Wilson, S., Johnston, K., Dehghansai, N., Koenigsberg, A., Vegt, S. de, & Wattie, N. (2020). Talent Research in Sport 1990-2018: A Scoping Review. *Frontiers in Psychology, 11*, 607710. https://doi.org/10.3389/fpsyg.2020.607710

Bergkamp, T. L., Niessen, A. S. M., den Hartigh, R., Frencken, W. G., & 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

Bös, K., Schlenker, L., Albrecht, C., Büsüch, D., Lämmle, L., Müller, H., Oberger, J., Seidel, I., Tittlbach, S., & Woll, A. (2009). *Deutscher Motorik-Test 6-18 (DMT 6-18)*. Feldhaus.

Davey, P. R., Thorpe, R. D., & Williams, C. (2002). Fatigue decreases skilled tennis performance. *Journal of Sports Sciences, 20*(4), 311-318. https://doi.org/10.1080/026404102753576080

Davids, K., & Baker, J. (2007). Genes, environment and sport performance: why the nature-nurture dualism is no longer relevant. *Sports Medicine (Auckland, N.Z.), 37*(11), 961-980. https://doi.org/10.2165/00007256-200737110-00004

De Bosscher, V., Bingham, J., Shibli, S., van Bottenburg, M., & De Knop, P. (2008). *The global Sporting Arms Race. An international comparative study on sports policy factors leading to international sporting success*. Aachen: Meyer & Meyer.

Deutscher Tennis Bund. (2020). *Jugendrangliste: Deutsche Ranglisten der Juniorinnen und Junioren*. Deutscher Tennis Bund. www.dtb-tennis.de/Tennis-National/Ranglisten/Jugend

Faber, I. R., Zamoscinska, M., Teunissen, J. W., & Pion, J. (2020). The relationship between sport-specific training and a perceptuo-motor skills assessment as part of talent identification in young table tennis players (8-10 years). *International Journal of Racket Sports Science, 2*(1), 42-54. https://doi.org/10.30827/Digibug.63720

Faber, I. R., Bustin, P. M. J., Oosterveld, F. G. J., Elferink-Gemser, M. T., & Nijhuis-Van der Sanden, Maria W G (2016). Assessing personal talent determinants in young racquet sport players: A systematic review. *Journal of Sports Sciences, 34*(5), 395–410. https://doi.org/10.1080/02640414.2015.1061201

Faber, I. R., Elferink-Gemser, M. T., Oosterveld, F. G. J., Twisk, J. W. R., & Nijhuis-Van der Sanden, M. W. G. (2017). Can an early perceptuo-motor skills assessment predict future performance in youth table tennis players? An observational study (1998-2013). *Journal of Sports Sciences, 35*(6), 593-601. https://doi.org/10.1080/02640414.2016.1180421

Faber, I. R., Pion, J., Munivrana, G., Faber, N. R., & Nijhuis-Van der Sanden, M. W. G. (2018). Does a perceptuomotor skills assessment have added value to detect talent for table tennis in primary school children? *Journal of Sports Sciences, 36*(23 - Table Tennis Special Issue), 593-601. https://doi.org/10.1080/02640414.2017.1316865

Fernandez-Fernández, J., Ulbricht, A., & Ferrauti, A. (2014). Fitness testing of tennis players: How valuable is it? *British Journal of Sports Medicine, 48*(1), 22-31. https://doi.org/10.1136/bjsports-2013-093152

Fett, J., Ulbricht, A., & Ferrauti, A. (2020). Impact of Physical Performance and Anthropometric Characteristics on Serve Velocity in Elite Junior Tennis Players. *Journal of Strength and Conditioning Research, 34*(1), 192-202. https://doi.org/10.1519/JSC.0000000000002641

Fett, J., Ulbricht, A., Wiewelhove, T., & Ferrauti, A. (2017). Athletic performance, training characteristics, and orthopedic indications in junior tennis Davis Cup players. *International Journal of Sports Science & Coaching, 12*(1), 119-129. https://doi.org/10.1177/1747954116684393

Filipcic, A., & Filipcic, T. (2005). The influence of tennis motor abilities and anthropometric measures on the competition successfulness of 11 and 12 year-old female tennis players. *Acta Universitatis Palackianae Olomucensis, Gymnica, 35*(2), 35-41. http://oldwww.upol.cz/fileadmin/user_upload/Veda/AUPO/AUPO_Gymnica_35-2.pdf#page=35

Filipcic, A., Pisk, L., & Filipcic, T. (2010). Relationship between the results of selected motor tests and competitive successfulness in tennis for different age categories. *Kinesiology, 42*(2), 175-183.

Gagné, F. (2010). Motivation within the DMGT 2.0 framework. *High Ability Studies, 21*(2), 81-99. https://doi.org/10.1080/13598139.2010.525341

Girard, O., & Millet, G. P. (2009). Physical determinants of tennis performance in competitive teenage players. *Journal of Strength and Conditioning Research, 23*(6), 1867-1872. https://doi.org/10.1519/jsc.0b013e3181b3df89

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. https://doi.org/10.1080/02640414.2012.713980

Grosser, M., Starischka, S., & Zimmermann, E. (2008). *Das neue Konditionstraining: Sportwissenschaftliche Grundlagen, Leistungssteuerung und Trainingsmethoden, Übungen und Trainingsprogramme* (10th ed.). BLV-Sportwissen. BLV.

Güllich, A., & Emrich, E. (2014). Considering long-term sustainability in the development of world class success. *European Journal of Sport Science, 14*(Suppl 1), S383-S397. https://doi.org/10.1080/17461391.2012.706320

Hohmann, A., & Seidel, I. (2003). Scientific aspects of talent development. *International Journal of Physical Education, 40*(1), 9-20.

Hohmann, A., Siener, M., & He, R. (2018). Prognostic validity of talent orientation in soccer. *German Journal of Exercise and Sport Research, 48*(4), 478-488. https://doi.org/10.1007/s12662-018-0549-5

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. https://doi.org/10.1080/02640414.2016.1177658

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*(8). https://doi.org/10.1371/journal.pone.0182211

International Tennis Federation. (2021). *What We Do to Promote Tennis Around the World | ITF*. International Tennis Federation. https://www.itftennis.com/en/about-us/organisation/what-we-do/

Johnston, K., Wattie, N., Schorer, J., & Baker, J. (2018). Talent Identification in Sport: A Systematic Review. *Sports Medicine (Auckland, N.Z.), 48*(1), 97-109. https://doi.org/10.1007/s40279-017-0803-2

Klein, M., Fröhlich, M., & Emrich, E. (2012). Zur Testgenauigkeit ausgewählter Items des Deutschen Motorik-Tests DMT 6-18. *Leipziger Sportwissenschaftliche Beiträge, 53*(1), 23-45.

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience, 16*(8), 1412-1425. https://doi.org/10.1162/0898929042304796

Koopmann, T., Faber, I. R., Baker, J., & Schorer, J. (2020). Assessing Technical Skills in Talented Youth Athletes: A Systematic Review. *Sports Medicine (Auckland, N.Z.), 50*(9), 1593-1611. https://doi.org/10.1007/s40279-020-01299-4

Kovacs, M. (2007). Tennis physiology: Training the competitive athlete. *Sports Medicine, 37*, 189-198. https://doi.org/10.2165/00007256-200737030-00001

Kramer, T., Huijgen, B. C., Elferink-Gemser, M. T., & Visscher, C. (2016). A longitudinal study of physical fitness in elite junior tennis players. *Pediatric Exercise Science, 28*(4), 553-564. https://doi.org/10.1123/pes.2016-0022

Kramer, T., Huijgen, B. C., Elferink-Gemser, M. T., & Visscher, C. (2017). Prediction of tennis performance in junior elite tennis players. *Journal of Sports Science and Medicine, 16*(1), 14-21.

Kramer, T., Huijgen, B. C., Lyons, J., Elferink-Gemser, M. T., & Visscher, C. (2016). Physical development of young talented tennis players. In M. Coelho e Silva, A. J. Figueiredo, R. M. Malina, & M. T. Elferink-Gemser (Eds.), *Youth Sports: growth, maturation and talent* (pp. 93-108). Portugal: Coimbra University Press.

LaPrade, R. F., Agel, J., Baker, J., Brenner, J. S., Cordasco, F. A., Côté, J., Engebretsen, L., Feeley, B. T., Gould, D., Hainline, B., Hewett, T., Jayanthi, N., Kocher, M. S., Myer, G. D., Nissen, C. W., Philippon, M. J., & Provencher, M. T. (2016). AOSSM Early Sport Specialization Consensus Statement. *Orthopaedic Journal of Sports Medicine, 4*(4), 2325967116644241. https://doi.org/10.1177/2325967116644241

Li, P., Weissensteiner, J. R., Pion, J., & Bosscher, V. de (2020). Predicting elite success: Evidence comparing the career pathways of top 10 to 300 professional tennis players. *International Journal of Sports Science & Coaching, 15*(5-6), 793-802. https://doi.org/10.1177/1747954120935828

Meckel, Y., Hophy, A., Dunsky, A., & Eliakim, A. (2015). Relationships between physical characteristics and ranking of young tennis players. *Central European Journal of Sport Sciences and Medicine, 10*(2), 5-12.

Mostaert, M., Pion, J., Lenoir, M., & Vansteenkiste, P. (2020). A Retrospective Analysis of the National Youth Teams in Volleyball: Were They Always Faster, Taller, and Stronger? *Journal of Strength and Conditioning Research*. Advance online publication. https://doi.org/10.1519/JSC.0000000000003847

Musch, J., & Grondin, S. (2001). Unequal Competition as an Impediment to Personal Development: A Review of the Relative Age Effect in Sport. *Developmental Review, 21*(2), 147-167. https://doi.org/10.1006/drev.2000.0516

Niessner, C., Utesch, T., Oriwol, D., Hanssen-Doose, A., Schmidt, S. C. E., Woll, A., Bös, K., & Worth, A. (2020). Representative Percentile Curves of Physical Fitness from Early Childhood to Early Adulthood: The MoMo Study. *Frontiers in Public Health, 8*, 458. https://doi.org/10.3389/fpubh.2020.00458

Pion, J. (2015). *The Flemish Sports Compass: from sports orientation to elite performance prediction.* [Doctoral dissertation, Universiteit Gent]. ResearchGate. https://doi.org/10.13140/RG.2.1.4426.8887

Potočnik, Ž. L., Jurak, G., & Starc, G. (2020). Secular Trends of Physical Fitness in Twenty-Five Birth Cohorts of Slovenian Children: A Population-Based Study. *Frontiers in Public Health, 8*, 561273. https://doi.org/10.3389/fpubh.2020.561273

Reid, M., & Schneiker, K. (2008). Strength and conditioning in tennis: Current research and practice. *Journal of Science and Medicine in Sport, 11*(3), 248-256. https://doi.org/10.1016/j.jsams.2007.05.002

Robertson, K., Pion, J., Mostaert, M., Norjali Wazir, Mohd Rozilee Wazir, Kramer, T., Faber, I. R., Vansteenkiste, P., & Lenoir, M. (2018). A coaches' perspective on the contribution of anthropometry, physical performance, and motor coordination in racquet sports. *Journal of Sports Sciences, 36*(23 - Table Tennis Special Issue), 2706-2715. https://doi.org/10.1080/02640414.2018.1441941

Roetert, E. P., Garrett, G. E., Brown, S. W., & Camaione, D. N. (1992). Performance Profiles of Nationally Ranked Junior Tennis Players. *Journal of Applied Sport Science Research, 6*(4), 225-231. https://doi.org/10.1519/1533-4287

Sannicandro, I., Piccinno, A., Cofano, G., Pascalis, S. de, & Rosa, A. R. (2012). Analysis of some variable performances of under 12 tennis players. *Medicina Dello Sport; Rivista di Fisiopatologia Dello Sport, 65*(4), 473-484.

Sarmento, H., Anguera, M. T., Pereira, A., & Araújo, D. (2018). Talent Identification and Development in Male Football: A Systematic Review. *Sports Medicine (Auckland, N.Z.), 48*(4), 907-931. https://doi.org/10.1007/s40279-017-0851-7

Siener, M., & Hohmann, A. (2019). Talent orientation: the impact of motor abilities on future success in table tennis. *German Journal of Exercise and Sport Research, 49*(3), 232-243. https://doi.org/10.1007/s12662-019-00594-1

Siener, M., Faber, I. R., & 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

Smith, K. L., Weir, P. L., Till, K., Romann, M., & Cobley, S. (2018). Relative Age Effects Across and Within Female Sport Contexts: A Systematic Review and Meta-Analysis. *Sports Medicine (Auckland, N.Z.), 48*(6), 1451-1478. https://doi.org/10.1007/s40279-018-0890-8

Suppiah, H. T., Low, C. Y., & Chia, M. (2015). Detecting and developing youth athlete potential: Different strokes for different folks are warranted. *British Journal of Sports Medicine, 49*(13), 878-882. https://doi.org/10.1136/bjsports-2015-094648

Till, K., & Baker, J. (2020). Challenges and [Possible] Solutions to Optimizing Talent Identification and Development in Sport. *Frontiers in Psychology, 11*, 664. https://doi.org/10.3389/fpsyg.2020.00664

Till, K., Cobley, S., O'Hara, J., Morley, D., Chapman, C., & Cooke, C. (2015). Retrospective analysis of anthropometric and fitness characteristics associated with long-term career progression in Rugby League. *Journal of Science and Medicine in Sport, 18*(3), 310-314. https://doi.org/10.1016/j.jsams.2014.05.003

Tomkinson, G., Carver, K., Atkinson, F., Daniell, N., Lewis, L., Fitzgerald, J., Lang, J., & Ortega, F. (2017). European normative values for physical fitness in children and adolescents aged 9-17 years: Results from 2 779 165 Eurofit performances representing 30 countries. *British Journal of Sports Medicine, 52*(22), 1445-1456. https://doi.org/10.1136/bjsports-2017-098253

Ulbricht, A., Fernandez-Fernández, J., & Ferrauti, A. (2015). The German physical condition tennis testing program: Measurements and implications for training. *Journal of Medicine and Science in Tennis, 20*(1), 6-16.

Ulbricht, A., Fernandez-Fernández, J., Mendez-Villanueva, A., & Ferrauti, A. (2016). Impact of fitness characteristics on tennis performance in elite junior tennis players. *Journal of Strength and Conditioning Research, 30*(4), 989-998. https://doi.org/10.1519/jsc.0000000000001267

Vaeyens, R., Güllich, A., Warr, C. R., & Philippaerts, R. (2009). Talent identification and promotion programmes of Olympic athletes. *Journal of Sports Sciences, 27*(13), 1367-1380. https://doi.org/10.1080/02640410903110974

Van Den Berg, L., Coetzee, B., & Pienaar, A. E. (2006). The influence of biological maturation on physical and motor performance talent identification determinants of U-14 provincial girl tennis players. *Journal of Human Movement Studies, 50*, 273-290.

Vaverka, F., & Cernosek, M. (2013). Association between body height and serve speed in elite tennis players. *Sports Biomechanics, 12*(1), 30-37. https://doi.org/10.1080/14763141.2012.670664

Zuber, C., Zibung, M., & 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*. https://doi.org/10.3389/fpsyg.2016.01088

Artikel 6

Siener*, M., Faber, I. R., & 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

applied sciences

MDPI

Article

Prognostic Validity of Statistical Prediction Methods Used for Talent Identification in Youth Tennis Players Based on Motor Abilities

Maximilian Siener [1,*], Irene Faber [2] and Andreas Hohmann [1]

[1] Institute of Sport Science, University of Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany; Andreas.Hohmann@uni-bayreuth.de
[2] Institute of Sport Science, University of Oldenburg, Ammerländer Heerstraße 114-118, 26129 Oldenburg, Germany; irene.faber@uni-oldenburg.de
* Correspondence: Maximilian.Siener@uni-bayreuth.de

Abstract: (1) Background: The search for talented young athletes is an important element of top-class sport. While performance profiles and suitable test tasks for talent identification have already been extensively investigated, there are few studies on statistical prediction methods for talent identification. Therefore, this long-term study examined the prognostic validity of four talent prediction methods. (2) Methods: Tennis players (N = 174; $n_{♀}$ = 62 and $n_{♂}$ = 112) at the age of eight years (U9) were examined using five physical fitness tests and four motor competence tests. Based on the test results, four predictions regarding the individual future performance were made for each participant using a linear recommendation score, a logistic regression, a discriminant analysis, and a neural network. These forecasts were then compared with the athletes' achieved performance success at least four years later (U13-U18). (3) Results: All four prediction methods showed a medium-to-high prognostic validity with respect to their forecasts. Their values of relative improvement over chance ranged from 0.447 (logistic regression) to 0.654 (tennis recommendation score). (4) Conclusions: However, the best results are only obtained by combining the non-linear method (neural network) with one of the linear methods. Nevertheless, 18.75% of later high-performance tennis players could not be predicted using any of the methods.

Keywords: prognostic validity; RIOC; talent; talent identification; sport; neural network

check for updates

Citation: Siener, M.; Faber, I.; Hohmann, A. Prognostic Validity of Statistical Prediction Methods Used for Talent Identification in Youth Tennis Players Based on Motor Abilities. *Appl. Sci.* **2021**, *11*, 7051. https://doi.org/10.3390/app11157051

Academic Editor: Mark King

Received: 3 June 2021
Accepted: 26 July 2021
Published: 30 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In professional sport, talent identification in the junior sector is of great importance [1]. After all, nations and clubs that manage to identify and train talented young athletes early on have an advantage in later sporting competition [2–4]. The search for future competitive athletes, therefore, begins with young athletes. In order to find suitable players, in recent years, more and more attention has been paid to tests that are designed to map the performance of children before they begin participating in a certain sport (i.e., talent detection), and when they begin training and competition (i.e., talent identification), these are believed to reflect the potential of a child for a certain sport [5]. Compared with scouting, testing has an advantage in that it can be carried out on children who have not yet achieved any success in playing. In addition, such a method can support decision-making for selection procedures in talent development programs, which are construed to help those young athletes who aim for and have the potential for elite sports.

Tennis, like other racquet sports, is considered an early-starting sport [6,7] because players require proper perceptuo-motor skills and a sound sports-specific technique on the elite level, which often must be developed at an early age [8]. Here, the great importance of early talent identification campaigns is evident. The common feature of most campaigns is that they begin at primary school age and, based on physical fitness and motor competence

tests, predict future success based on a child's performance prerequisites profile [9–12]. This early testing and desired early start in tennis ensures that, in addition to a large number of training hours—athletes complete 20-30 h of technical training per week, even at a young age [13]—the athletes can gain competitive experience early and over a longer period of time [14]. Moreover, this offers a good opportunity for the acquisition of technical skills because there are sensitive learning phases that promote motor learning, particularly before puberty [15–18]. Perhaps, for this reason too, national sports organizations are investing more and more effort into the systematic identification of talented young players. In this professionalized, competitive environment, a "relaxed approach" is no longer sustainable [19], and talent identification is becoming the key to national elite sport performance [20,21]. In this context, statistical analysis has advantages as compared with the usual expert ratings of trainers and scientists, and they can be used to identify talented athletes and reduce the costs of talent development [22].

However, while the previous focus of scientific talent research has been very much on the prerequisite profile of later professional tennis players [23–25] or the improvement of prognostic test batteries [8], there have been very few studies on actual prediction calculations in sport [26]. The resulting common practice is characterized by assessing children with specific tests and comparing their performance profile with that of professional players (top down). However, performance profiles at such a young age may still differ considerably from the later performance profile in adulthood [27]. Additionally, some factors initially considered unimportant for the sport, such as balance in soccer, may be crucial for later development and success. Baker et al. [27] describe the typical procedure as a kind of "performance identification", as compared with the required "talent identification" (future potential). For this reason, it is necessary to follow young athletes who are tested over the long term over their careers [28,29] and objectively determine the prognosis of future performance based on all test parameters (bottom up). In addition to classical linear methods [30], non-linear methods [31,32] have become established as common forecasting methods. However, these prediction methods have not yet been tested to determine their prognostic validity.

The aim of this study was, therefore, to compare the prognostic validity of common statistical prediction methods regarding the future performance success of young tennis players based on their juvenile performance profiles. Thus, based on physical fitness and motor competence tests, the performance levels of young tennis players were examined using various statistical prediction methods, and forecasts of future success were made. The prognostic validity of these common methods can be assessed on the basis of these predictions and the later tennis performance achieved by the participants, and the analytical methods can be evaluated with regard to their practical relevance (i.e., sensitivity and specificity). In this context, it is not only about how well the respective forecasts turn out to be but also about which method can most identify talented young tennis players.

2. Materials and Methods

2.1. General Study Design

For this purpose, in a long-term study, club tennis players, at the age of 8 years (U9), were tested using two anthropometric, five physical fitness, and four motor competence tests. Using a tennis-specific recommendation score composed of weighted test performances, a discriminant analysis, a binary logistic regression, and a neural network (multilayer perceptron), four separate forecasts were made for each participant as either a future ranking player in the German tennis rankings (top performer) or as a weaker club player without a ranking (low performer). The predictions were then compared with the tennis performance achieved about 5 years later (U13-U18). For each of the four methods, the achieved sensitivity and specificity of the prognostic classification were evaluated, and the prognostic validity levels of the methods were compared based on their values of relative improvement over chance (RIOC values; [33,34]).

2.2. Participants

The participants in this study included N = 174 junior tennis players (U13-U18), with $n_{♀}$ = 62 female and $n_{♂}$ = 112 male athletes. The mean age of the participants was 156 ± 16 months (min = 132, max = 206). All tennis players were club players who actively participated in club matches and tournaments.

2.3. Tennis Success

In terms of performance success, junior tennis players (U13-U18) were classified into two categories: top performers (TPs) and low performers (LPs). Players who achieved enough wins and points in ranked tournaments and were therefore listed in the current national tennis ranking lists [35,36] were classified as top performers (N = 16; $n_{♂}$ = 11, $n_{♀}$ = 5). In the group of low performers, 158 tennis players with only average performance were identified. These players were unable to perform beyond local and regional successes and at no time fulfilled the necessary requirements to obtain a national ranking. Finally, this group of low performers comprised n = 57 female tennis players and n = 101 male tennis players.

2.4. Anthropometric Characteristics and Motor Abilities at U9

All junior tennis players were already tested at U9. The U9 testing included two anthropometric, five physical fitness, and four motor competence tests. The standardization of the test items is captured in protocols, which include a detailed description of the materials, set-up, assignment, demonstration, training phase, testing phase, and test scores registrations [37]. The test tasks aimed at the diagnosis of sprint, coordination, balance, flexibility, arm and upper body strength, leg power, ball throw, and endurance performance.

2.4.1. 20 m Sprint (SP)

The time for a 20 m linear running sprint was measured by means of light gates (Brower Timing Systems; Draper, USA). The starting position was 0.3 m behind the start line. Between the two potential attempts, a break of at least 2 min was allowed. The objectivity of this test is 0.86, and its reliability is 0.96 [38].

2.4.2. Sideward Jumping (SJ)

The test involves 15 s of sideward jumping within two adjacent 50 cm × 50 cm squares. The number of two-legged jumps from one square to the other without touching a boundary line was measured. Five trial jumps were allowed before the testing began. Between the two potential attempts, a break of at least 2 min was allowed. The objectivity of this test is 0.99, and its reliability is 0.89 [37].

2.4.3. Balancing Backwards (BB)

Players were asked to balance backwards on 6 cm, 4.5 cm, and 3 cm wide beams. For each beam, the number of steps backwards taken while balanced (feet fully raised) before leaving the bar was counted. The maximum number of steps per attempt was limited to eight. For each of the three beams, two attempts were made. Thus, the result was the sum of all steps taken (maximum: 48 steps). There was a short practice period before the test was carried out. The objectivity of this test is 0.99, and the reliability is 0.73 [38].

2.4.4. Standing Bend Forward (SBF)

A standing bend forward test was performed as a flexibility test. Here, the participants attempted to reach as far as possible with their fingertips beyond their feet and hold this position for at least three seconds. The distance between the fingers and ground level was recorded in cm, and a range of very low values measured from just above ground level were recorded as negative distances. Two attempts were allowed. The objectivity of this test is 0.99, and its reliability is 0.94 [37].

2.4.5. Push-Ups (PU)

The push-up test was carried out after a short trial period. Within 40 s, the number of fully completed repetitions was counted. A complete repetition was only evaluated when the upper body was laid down. Only one attempt was allowed. The objectivity of this test is 0.98, and its reliability is 0.69 [37].

2.4.6. Sit-Ups (SU)

Similar to the push-up test, the time available for the sit-up test was limited to 40 s. After a short practice phase, only one test was granted, and the number of correctly executed sit-ups was counted. The objectivity of this test is 0.92, and its reliability is 0.74 [39].

2.4.7. Standing Long Jump (SLJ)

The standing long jump was carried out without a previous practice session. The distance of the standing jump was measured in cm (measured from the heel). A break of at least 2 min was observed between two attempts. The objectivity of this test is 0.99, and its reliability is 0.89 [37].

2.4.8. Ball Throw (BT)

The ball throw was performed from a standing position with an 80 gr ball. The distance to the impact point of the ball was measured along a line orthogonal to the point of release. The result was rounded to the nearest 10 cm. After an initial trial, three scoring attempts were made, of which only the farthest test value was used for subsequent calculations. The reliability of the ball throw test has been demonstrated in a series of our own studies, and $r = 0.77$ ($n = 1800$).

2.4.9. Six min Endurance Run (ER)

A 6 min endurance run around a volleyball pitch (9 × 18 m) was carried out. There, the number of meters covered was measured. The test was conducted in groups of 15 persons at the same time. The objectivity of this test is 0.87, and the reliability is 0.92 [37].

All players were assessed under similar conditions. The tests were carried out during regular school hours (8-12 a.m.) by qualified test personnel. The testing always began after a uniform warm-up phase with the 20 m sprint and ended with the 6 min endurance run. In all tests, except for sideward jumping (where the average of the two attempts was taken as test result), the better of the attempts counted. Table 1 shows a typical dataset for two participants tested in U9 (each row represents one participant). In addition to the nine general motor test items and the anthropometric parameters, gender and test age (in months) were also recorded. Additionally, the ranking success (TP) achieved at junior age was recorded later.

Table 1. An example of test data acquisition.

Sex	Age	TP	Height	Weight	SP	SJ	BB	SBF	PU	SU	SLJ	BT	ER
1	97	1	132	27.3	4.21	33	38	4	17	23	153	19	1078
0	94	0	129	27.3	4.45	27	30	–2	14	19	135	14	959

Legend: Sex (1 = female, 0 = male); Age (months); TP = top performer (1 = yes, 0 = no); Height (cm); Weight (kg); SP = sprint (seconds); SJ = sideward jumping (repeats); BB = balancing backwards (steps); SBF = standing bend forward (cm); PU = push-ups (repeats); SU = sit-ups (repeats); SLJ = standing long jump (cm); BT = ball throw (m); ER = endurance run (m).

All tests (with the exception of the ball throw) have been examined in a series of studies by various authors [37–39] with regard to the test standards. The results show high objectivity and reliability coefficients, even if these vary considerably between tests. The average total retest reliability of the eight test items (without the ball throw) is $r_{general} = 0.85$. The objectivity of the test battery is $r_{obj} = 0.95$ (range: 0.87–0.99). However, the validity

of the test procedures has not yet been sufficiently verified. While Bös et al. [37] have focused primarily on content-logical validity and mainly used expert ratings (expert rating: $M = 1.83$; with grades from 1 to 5), other authors have used correlations to check the criterion-related validity or confirmatory factor analyses to determine construct validity [40]. For most of the individual tests, sufficient to very good test validity can be attested to for the latter two validity categories ($r_{validity} = 0.69$). Thus far, there are no major studies on prognostic validity.

2.5. Statistical Analyses

All evaluations and analyses were performed using SPSS (Version 26.0; SPSS Inc., Chicago, IL, USA), and the (bilateral) significance level was set to $p < 0.05$. Unless otherwise indicated, significant findings in figures and tables are marked with *. Significance values of $p < 0.01$ are marked with **.

Some studies [29,41,42] show that age can have a direct influence on test results. To eliminate such an age bias, a univariate analysis of variance (ANOVA) was used for all test variables to examine these data for significant differences in the U9 age groups of boys and girls. It was found that performance increases with age. In order to avoid these relative age effects, all data were first separated according to gender and then further calculated for both datasets independently. In each of the two datasets, bivariate linear regressions were then calculated for each test item separately, and the resulting residuals were saved. In these regressions, age (in months) served as the independent variable, and the respective single test served as the dependent variable. The respective residuals were finally z-standardized, and these z-standardized residual values for both gender groups were merged once again [43–45]. This resulted in an age-independent and gender-independent overall dataset, which was used for all further calculations. For the sake of simplicity, these z-standardized residuals are now called z-values for subsequent analyses. In addition, the prefix of the z-values for the 20 m sprint was inverted to allow for an easier comparison with the other test values. The z-values were approximately normally distributed, as shown by the Shapiro–Wilk test ($p > 0.05$) and a visual inspection of the histograms.

2.5.1. Prediction Methods

Because talent-detection campaigns differ in their prediction processes, four methods of recommendation calculation were analyzed and compared. In addition to a classical linear method using a tennis-specific recommendation score (weighting), a (binary) logistic regression, a linear discriminant analysis, and a neural network (multilayer perceptron) were used. All calculations were performed using the selection rate of the forecasts provided by SPSS. In each of the three classification analyses, that is, (binary) logistic regression, discriminant analysis, and neural network, junior tennis players' performance (top performer or low performer) was used as the outcome variable (dependent variable), and the z-standardized test scores (z-values) of the U9 tests were used as input variables (independent variables). The linearly calculated tennis recommendation score was used as an external classification criterion [26].

For the three analysis methods, linear regression, discriminant analysis, and neural network (multilayer perceptron), the data were divided into five equally sized random subsets prior to the calculations. Four subsets (80% of the data) were used to create or train the respective analysis method. Using the remaining 20% (test set or hold-out), the classification method was finally tested. This calculation method was performed a total of five times, so that each of the five subsets was used once as a test set (hold-out). The test set (hold-out) results were then averaged. This procedure is known as k-fold cross-validation (CV; with $k = 5$). It is intended to prevent an analysis method from having already seen a case to be classified in training. The results correspond therefore rather to a situation from practice, in which usually an unknown person is to be classified. However, since the random selection of the five subsets can also have an influence on the classification results, according to Kolias et al. [46], the 5-fold cross-validation was performed five times using

different random subsets, and the resulting test set classifications were averaged over all 25 predictions (5-fold 5-sample CV). To ensure a comparison of the methods, the same random partitions were used for all three methods.

2.5.2. Tennis-Specific Recommendation Score

The *Tennis-Specific Recommendation Score* (TRS) is based on a tennis-specific weighting of individual tests for talent identification. The TRS was established on the basis of expert ratings and derived from empirical tests of professional, adult ranked tennis players [26]. It can therefore be considered an external talent criterion. It was calculated separately for each participant and indicates the suitability of the child's talent make-up for the particular demands of tennis. The test values of body height, body mass index (BMI), standing long jump, 20 m sprint, ball throw, sideward jumping, 6 min endurance run, balancing backward, and bend forward were included according to Formula (1).

$$\begin{aligned} z_{TRS} = [2.0 \times z_{standing\ long\ jump} &+ 1.8 \times z_{sideward\ jumping} + 1.7 \times z_{height} \\ &+ 1.3 \times \left(z_{sprint} + z_{balancing} - z_{bmi}\right) \\ &+ 1.2 \times \left(z_{ball\ throw} + z_{bend\ forward}\right) + 1.1 \times z_{endurance\ run}] \\ &\div 12.9 \end{aligned} \quad (1)$$

Due to the specific calculation involved, z_{TRS} is comparable to a normal z-value. Thus, also here, a value of $z_{TRS} = 0$ indicates average suitability. In order not only to give a general tennis recommendation ($z_{TRS} > 0$) but also to predict future top performers, it is necessary to define a suitable threshold value, above which a participant is assigned to the group of potential top performers. Sensitivity and specificity can thus be determined using the selected threshold value, e.g., $z_{TRS} = 1.3$. In this example, the chosen z-value of $z_{TRS} = 1.3$ corresponds to a selection rate of 10%, which is close to the observed percentage of top performers in this sample (16/174 = 10.875%). Thus, using this method, participants with $z_{TRS} \geq 1.3$ were automatically predicted to be top performers, and participants with $z_{TRS} < 1.3$ were predicted to be low performers.

2.5.3. Binary Logistic Regression Analysis

A binary logistic regression examines the influence of independent variables on a binary-coded dependent variable. Thus, the influence of the test values on later tennis performance was determined in this way. The group of youth top performers was marked with a "1", and the group of low performers was marked with a "0". Because the "Enter" method was chosen, all variables of a block were recorded in common in one step, and their influence on the regression was evaluated simultaneously. In trial calculations, it was found that the results of the "backward" and "forward" method in the classification correspond to those of the "Enter" method. Based on the multiple regression, the output is also presented as a categorical variable, and therefore, it is possible to predict affiliation with a performance group. Sensitivity and specificity can be easily calculated. The separation value of this categorical output variable was $u_{cut} = 0.5$ by default for these analyses. This means that, with an individual output value of $u_{cut} \geq 0.5$, the participant was recommended as a top performer. In order to represent a valid and also practical result, the logistic regression was performed as a k-fold cross-validation ($k = 5$), and this procedure was repeated five times. The results of the 25 individual analyses were ultimately averaged.

2.5.4. Discriminant Analysis

Discriminant analysis can be used to investigate differences in feature profiles of different groups and then to assign an unknown profile to a suitable group. Similar to binary logistic regression, group memberships are calculated, and thus the sensitivity and specificity can be analyzed. Since the grouping variable "tennis performance" had only two values (top performer/low performer), only one discriminant function was determined. Using the discriminant coefficients of this function, a performance prediction could be

calculated for each single hold-out case. Since the group variable was dichotomous, the recommendation calculation was performed using a linear method similar to binary logistic regression. The a priori probability of the groups was assumed to be 50:50 (settings: groups equal). For a practical investigation, a 5-fold cross-validated procedure was also considered This procedure was repeated five times with different partition groups, and the results were averaged (see Binary Logistic Regression Analysis and Neural Network Analysis). The individual talent characteristics profile (hold-out) of the required athlete was therefore not already in the training group, making the identification of the case more difficult. This form of analysis therefore offers lesser sensitivity and specificity but corresponds more closely to the "natural" application of talent prognosis in sports practice.

2.5.5. Neural Network Analysis

In contrast to the three more traditional linear or multilinear prediction methods, a neural network (multilayer perceptron) was applied as the fourth type of analysis. Thus, with the *multilayer perceptron* (MLP) tool from SPSS 26 (IBM), a non-linear "feed-forward" classification procedure was used to analyze the prognostic validity of the group assignment [47,48].

The classification calculation of the neural network (SPSS 26) is composed of three basic parts *training*, *validation*, and *test* (hold-out). The network weights are mainly determined via *training*, with *validation* measuring the model errors. Only the *test* set is independent of the creation of the network, which provides an honest estimate of the predictive power of the model. The distribution of the *training*, *validation*, and *test* (hold-out) components was carried out for smaller groups, as in the present case, according to a 60-20-20 scheme [49,50]. Each part (*training*, *validation*, and *test*) is randomly selected from the total sample. The calculations for the neural network therefore corresponded to the 5-fold CV already used in the logistic regression and discriminant analysis. Again, the 5-fold CV was repeated using five different partition divisions [46] and the corresponding 25 *test* classifications were averaged (5-fold 5-sample CV). Additionally, this method is intended to compensate for the small dataset and avoid misinterpretation due to volatile parameter estimates [51].

The architecture of the network contained the eleven test variables (*z*-values) as input neurons (covariates without additional scaling or normalization) and the two performance classes (top performer/low performer) as output neurons (dependent variable). The neurons of one hidden layer in-between were generated by the program independently and depended on a randomly chosen training start vector. Due to the small amount of data, the architecture was limited to a maximum of one hidden layer with one to ten neurons [45]. However, in various test calculations with the data set used here, it was shown that even with different start vectors, the calculated neural network solutions never exceeded more than five neurons (plus one bias neuron) in the hidden layer. Figure 1 shows a typical example of the (fully connected) neural network with nine input neurons (+1 bias neuron), four neurons (+1 bias neuron) in the hidden layer, and the two output neurons of the binary target variable (Figure 1). Experimental calculations with several hidden layers showed no significant improvements in the network. The learning of the network is iteration-based and terminated when no further reduction of the error quotient is apparent. The type of training used was "batch training" [52], which is the recommended method for smaller datasets. All other settings were selected according to the default settings of SPSS.

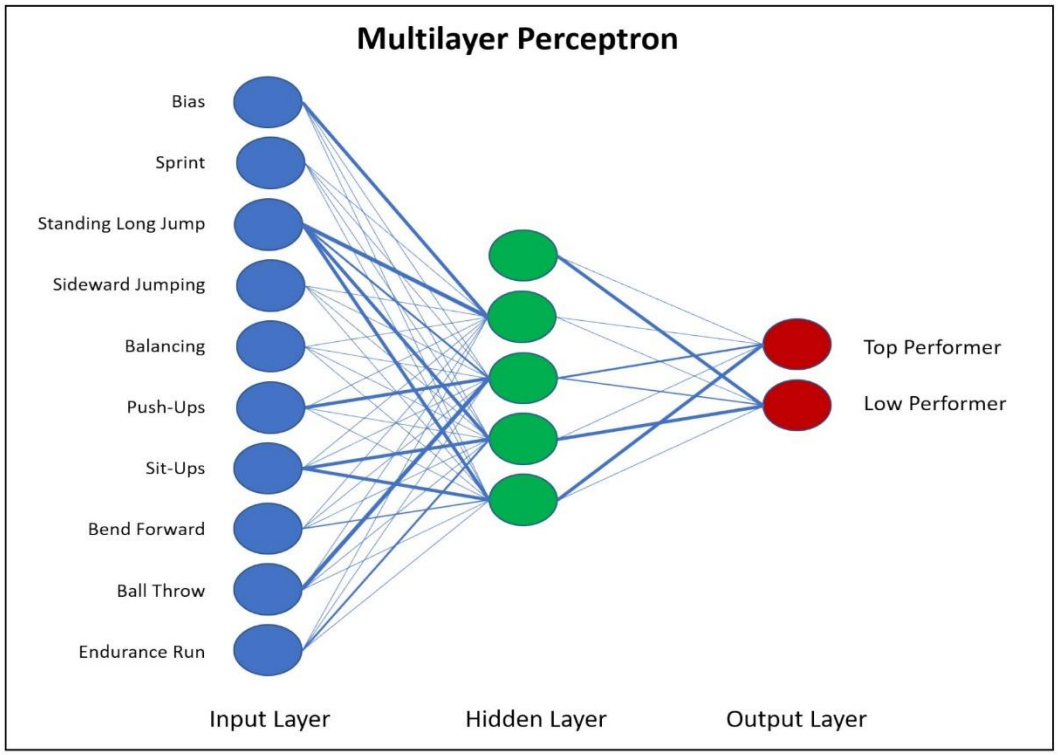

Figure 1. Example of a Neural Network (Multilayer Perceptron).

2.5.6. Prognostic Validity of the Analyses

Prognostic validity can be determined by the theoretical accuracy (sensitivity and specificity) of the performance forecasts of the various prediction methods. For all calculations of the parameters of the analysis methods, only the classification results of the corresponding hold-outs were used. Table 2 represents a typical example of a 2 × 2 classification table. Variable A corresponds to true-positive predictions, B to false-positive predictions, C to false-negative predictions, and D to true-negative predictions.

Table 2. Example for sensitivity and specificity calculation.

		Predicted/Recommended		
		Low Performer	**Top Performer**	Σ
Observed/Existing	Top Performer	C	A	A + C
	Low Performer	D	B	B + D
	Σ	C + D	A + B	A + B + C + D

Sensitivity (recall; Formula (2)) indicates how many top performers (A = true positives) from the observed group (A + C = 16) were correctly predicted. For example, if twelve top performers were correctly identified/classified as top performers, the sensitivity is 75%. The specificity (Formula (3)) represents the percentage of correctly predicted low performers (D = true negatives) in the group of low performers (B + D).

$$\text{Sensitivity} = \frac{A}{A + C} \quad (2)$$

$$\text{Specificity} = \frac{D}{B+D} \tag{3}$$

Furthermore, the positive and negative predictive values (Formulas (4) and (5)) of the predictions were determined (positive predictive value = precision). These two parameters indicate the probability of actually living out a corresponding forecast. It therefore corresponds to the percentage of predictions that actually occurred.

$$\text{Positive Predictive Value} = \frac{A}{A+B} \tag{4}$$

$$\text{Negativ Predictive Value} = \frac{D}{C+D} \tag{5}$$

In general, the hit rate is the percentage of correctly predicted recommendations. Accordingly, the true-positive predictions (A) and the true-negative predictions (B) are added together and divided by the total sample size (Formula (6)). To calculate the random hit rate (Formula (8)), the selection rate (S) is needed. The selection rate is the percentage of top performer recommendations among all recommendations (Formula (7)). It is determined in advance or via the analysis types. Using the selection rate, the random hit rate (Formula (8)) can now be determined. For this purpose, the selection rate (S) is multiplied by the number of existing top performers (A + C). This calculates the number of correct top performer recommendations in a random draw from the group of top performers. In addition, the number of correct low performer recommendations in a random draw from the group of low performers (B + D) is also determined. This is done using the counter-probability of the selection probability (1 − S). Both values are then summed and divided by the total sample size to obtain the random hit rate. The maximum hit rate can also be easily calculated using the counter-probability (Formula (9)). To do so, the over-recommendation of a group caused by the selection rate is subtracted from the total probability (=1). For example, if 20 top-performer recommendations (A + B) have been made by an analysis (selection rate = 11.5%) and only 16 top performers (A + C) exist (~9.2%), then four recommendations ([A + B] − [A + C]) are automatically wrong (~2.3%). Therefore, these false over-recommendations are eliminated, and the maximum hit rate would be 97.7%. The number of over-recommended cases is obtained by subtracting false-positive (B) cases from false-negative (C) cases ([A + B] − [A + C] = B − C). If there is an under-recommendation and fewer than 16 top performer recommendations are made, the subtraction is reversed (C − B). Both can be expressed in a simplified way via | B - C | (see Formula (9)).

$$\text{Hit Rate} = \frac{A+D}{A+B+C+D} \tag{6}$$

$$\text{Selection Rate} = S = \frac{A+B}{A+B+C+D} \tag{7}$$

$$\text{Random Hit Rate} = \frac{S \times (A+C) + (1-S) \times (B+D)}{A+B+C+D} \tag{8}$$

$$\text{Maximum Hit Rate} = 1 - \frac{|B-C|}{A+B+C+D} \tag{9}$$

Because the numbers of top performer forecasts (selection rate) can vary between methods, an RIOC value (Relative Improvement Over Chance; [33,34]) was ultimately calculated for each method (x_{RIOC}). This value determines the *relative hit accuracy*. In most analyses, a maximum hit rate of 100% cannot be achieved. This can lead to a misinterpretation of the validity parameters (e.g., phi or kappa; [34]). The RIOC index avoids this problem by calculating the actual hit rate in relation to the potential maximum hit rate (see Formula (10)).

$$x_{RIOC} = \frac{\text{hit rate} - \text{random hit rate}}{\text{max. hit rate} - \text{random hit rate}} \tag{10}$$

The calculated value varies, usually in the range $0 < x_{RIOC} \leq 1$, but it can also have negative values. With a value of $x_{RIOC} = 0.33$, the classification is considered good. Above 0.66 is considered very good [53]. Using this calculation scheme, the four methods can be directly compared and evaluated independent of their actual selection rates. The real selection rate for talent diagnostics and, thus, also the maximum hit rate, always depends on the talent campaigns and support programs of the participating countries, and therefore, this selection rate can vary considerably in practice.

The Youden Index J (also *Youden's* J) was determined as a further comparative value (Formula (11)). It is calculated by summing sensitivity and specificity, and therefore, it shows prognostic validity independent of the sizes of the two performance groups [54]. Usually, J reaches values between 0 (random result) and 1 (optimum result).

$$\mathrm{J} = \text{Sensitivity} + \text{Specificity} - 1 \tag{11}$$

In addition to *Youden's* J, the Area Under ROC Curve [55] and F_1 score were also ways to evaluate the predictive power of an analytical method (Formula (12)). The F_1 score is based on the harmonic mean, and its calculation is equally divided between sensitivity (recall) and positive predictive value (precision). It can take values between 0 and 1, where 1 stands for the maximum predictive strength. In contrast to *Youden's* J, which includes sensitivity and specificity in equal measure, the focus here is on the predictive power of true-positive cases.

$$\mathrm{F_1} = 2 \times \frac{\text{Sensitivity} \times \text{Positive Predictive Value}}{\text{Sensitivity} + \text{Positive Predictive Value}} \tag{12}$$

2.5.7. Classification of Individual Tennis Players

In order to gain a more precise insight into the calculated prognosis, the participants predicted to be future top performers were also recorded for each type of analysis. Thus, we could determine which individual top performers were correctly identified via which method and which low performers were erroneously judged to be top performers. Using a *classification map*, it was therefore easy to determine how the predictions of the various analysis methods were distributed. In addition, the probability that an athlete would later become a top performer, as calculated via the various methods, was recorded.

3. Results

3.1. Test Performance

In considering the test results of the five physical fitness, four motor competence, and two anthropometric test items (U9), significant differences between later top performers and low performers (U13-U18) can be observed. The mean values of the later top tennis players were significantly higher than those of the later low performers for nearly all test variables. While no differences could be found for body weight ($p = 0.94$) and bend forward ($p = 0.16$), all other test items showed significant differences ($p \leq 0.05$). The mean values of the test items sideward jumping, balancing backwards, standing long jump, 6 min endurance run, and ball throw showed highly significant differences ($p \leq 0.01$). The minimum of the sideward jumping test was $MIN = 27.0$ repetitions (rps) for top performers, which was only slightly below the average value of $M = 27.3$ rps for low performers. Additionally, for the balancing backwards, standing long jump, sit-ups, and 6 min run tests, 84% of the top performers were better than the average low performer. A maximum of 16% of low performers managed to reach the average of the top performers on the standing long jump and ball throw. All descriptive statistics for the test items can be found in Table 3.

Table 3. Descriptive statistics for the two anthropometric, five physical fitness, and four motor competence diagnostics of the U9 participants.

Variables	Groups	*N*	M	SD	SE	95% CL		Min	Max
						LL	UL		
Calendar age * (month)	LP	158	93.8	5.0	0.39	93.0	94.6	83	110
(Time of testing; U9)	TP	16	97.5	6.4	1.60	94.1	100.9	88	112
Test results									
Body height * (cm)	LP	158	129.1	5.7	0.45	128.2	130.0	117	145
	TP	16	132.3	4.5	1.13	129.8	134.7	127	143
Body weight (kg)	LP	158	27.2	4.1	0.33	26.6	27.9	20.0	39.3
	TP	16	27.3	1.9	0.48	26.3	28.3	23.4	30.6
Sideward jumping **	LP	157	27.3	6.1	0.49	26.3	28.2	6.5	40.5
(repeats)	TP	16	33.0	5.1	1.27	30.3	35.7	27.0	45.0
Balance backward **	LP	158	30.3	8.5	0.67	28.9	31.6	8	48
(steps)	TP	16	38.3	5.7	1.42	35.2	41.3	28	48
Standing long jump **	LP	157	135.1	16.2	1.30	132.5	137.6	82	190
(cm)	TP	16	153.0	17.2	4.30	143.8	162.2	125	178
20 m sprint *	LP	158	4.45	0.36	0.03	4.39	4.51	3.10	5.32
(s)	TP	16	4.21	0.34	0.08	4.03	4.39	3.50	4.72
Push-ups * (repeats)	LP	158	14.6	3.6	0.28	14.0	15.1	4	24
	TP	16	17.1	5.2	1.30	14.3	19.9	9	25
Sit-ups * (repeats)	LP	158	19.2	5.1	0.40	18.4	20.0	2	30
	TP	16	23.4	4.1	1.02	21.3	25.6	15	29
Bend forward (cm)	LP	158	1.98	5.94	0.47	1.05	2.92	−11	18
	TP	16	4.16	4.90	1.22	1.55	6.77	−10	12
6 min run **	LP	154	959	130.8	10.5	938	979	545	1259
(m)	TP	16	1078	80.7	20.2	1035	1121	891	1200
Ball throw ** (m)	LP	155	13.5	4.03	0.32	12.9	14.2	3.8	27.6
	TP	16	18.8	5.35	1.34	15.9	21.6	9.2	28.3

M = mean; SD = standard deviation; SE = standard error; CL = confidence limit; LL = lower limit; UL = upper limit; Min = minimum; Max = maximum; LP = low performer; TP = top performer; * $p \leq 0.05$; ** $p \leq 0.01$.

3.2. Test Quality Parameters of the Prediction Methods

3.2.1. Tennis Recommendation Score

Using the TRS (see Table 4), a total of 39 of the 174 test participants (22.4%) were predicted to become talented tennis athletes (selection rate). Overall, 135 participants (77.6%) did not receive a top-level assignment because it was assumed that they could not reach the top level due to their low individual recommendation values. With a selected threshold value of z_{TRS} = 1.3, 12 of the 16 later top performers (75%) were true-positives (sensitivity). One hundred and thirty-one participants (82.9%) who did not reach the top level were also correctly classified as low performers, that is, true negatives (specificity). In total, 143 of 174 (82.2%) children could be correctly identified as later top or low performers via this linear method. If the talent forecasts with the same prediction rate (22.4%) were expressed at random, only 4, as compared with 12, of the top performers and 122, as compared with 131, of the low performers would be correctly predicted, and the overall prediction quality would only be 72.4%, as compared with the value of 82.2% described above. If the percentage values for sensitivity and specificity were considered in combination, the benefit is more obvious. With a cutoff limit value of z_{TRS} = 1.3, the Youden Index amounts to a total of 157.9% (75% + 82.9%), thus representing a 57.9% improvement as compared with a random drawing. The Area Under ROC Curve was 0.852 (standard error = 0.048; 95% confidence interval: min = 0.759, max = 0.946).

Table 4. Sensitivity and specificity of the tennis recommendation score at z_{TRS} = 1.3.

Tennis Recommendation Score (z_{TRS} = 1.3)		Predicted/Recommended		
		Low Performer	Top Performer	Percentage Correct
Observed/Existing	Top Performer	4	12	75%
	Low Performer	131	27	82.9%
	Percentage (total)	77.6%	22.4%	82.2%

3.2.2. Logistic Regression

In addition to the TRS, a (binary) logistic regression also offers a chance to create a talent prognosis for the test participants. The 5-fold cross-validation was performed five times, and the results were averaged. The *omnibus test* of the model's averaged coefficients showed a significant result (chi-square(11) = 33.67, $p < 0.001$). The model quality was determined based on Nagelkerke's R-square, which had a value of 0.476. The Area Under ROC Curve was 0.810 (standard error = 0.054; 95% confidence interval: min = 0.705, max = 0.915). Overall, however, in most analyses, none of the regression coefficients showed a significant result, so this form of analysis should be interpreted with caution.

Approximately 6 of the 16 athletes who were ultimately competitively ranked were correctly identified by their test values (Table 5). The sensitivity was therefore 37.5%. It should be noted, however, that only about twelve children in total received a top-level prognosis in this form of analysis. The specificity was 96%. With the exception of six participants, all low performers could be correctly classified. The overall accuracy of the analysis method was 90.4%. The summed percentage result for the sensitivity and specificity of the logistic regression was 133.5%, which is 33.5% better than that of a random sample. If a random prediction were considered with the same selection rate, the sensitivity would be 7.2%. Thus, it would be 26.3% less than the sensitivity value of 33.5% achieved via the logistic regression analysis. The same applies to a random specificity. This would amount to 92.6%, as compared with the 96% achieved here, and the overall result in terms of correctly identified test participants would only be 84.4%. Among participants predicted to be future top performers via the logistic regression, 50% of such prognoses were correct (positive predictive value). This means that every second child classified as a top performer actually made it into the top group.

Table 5. Sensitivity and specificity of the (binary) logistic regression analysis.

(Binary) Logistic Regression [1]		Predicted/Recommended		
		Low Performer	Top Performer	Percentage Correct
Observed/Existing	Top Performer	10	6	37.5%
	Low Performer [2]	144	6	96%
	Percentage (total)	92.8%	7.2%	90.4%

[1] A 5-fold cross-validation was performed a total of five times, and the results were averaged. For a better evaluation of the results, the average values were multiplied by five. [2] Eight children could not perform at least one test, so they were excluded from this analysis.

3.2.3. Discriminant Analysis

A 5-fold cross-validated discriminant analysis was calculated and repeated five times with various partitions. The corresponding results were averaged and provided another way to make talent predictions. Because there were only two service groups (binary), only one discriminant function was required here. The ability to separate the two groups can be seen in the averaged eigenvalue (this was EV = 0.286, canonical correlation: r = 0.471). Thus, the groups could be separated satisfactorily based on their group centroids. In addition, the discriminant function should now show the best possible separation. In this

analysis, Wilk's lambda was 0.775 (chi-square(11) = 31.81, $p < 0.01$). The Area Under ROC Curve was 0.800 (standard error = 0.069; 95% confidence interval: min = 0.663, max = 0.936). Consequently, significant differences between the two groups could be detected via the discriminant function, but the model had only a limited selectivity due to the high Wilks lambda. Nevertheless, the cross-validation classification showed good results (Figure 2).

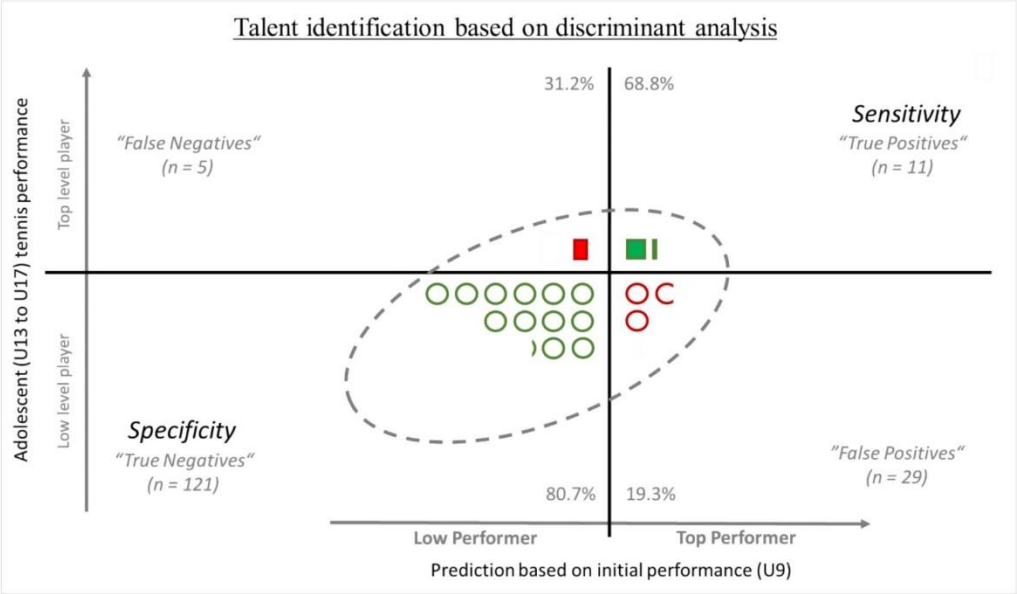

Figure 2. Discriminant analysis to predict later U13-U17 tennis performance group based on initial performances (U9; each full symbol represents ten children).

Using this method, 11 of the 16 later top performers could be correctly identified (Figure 2). This corresponds to a sensitivity of 68.8%. Additionally, 121 of the later 150 low performers could be correctly classified (specificity of 80.7%). The overall prediction quality was thus 79.5% (132/166), and the sum of the two parameters was 149.5%. A random prediction with the same classification rate (24.1%) would result in a sensitivity of 24.1% and a specificity of 75.9%, with an overall prognostic quality of 70.9%. While the discriminant analysis differed only slightly from the random sample in specificity and total result, in sensitivity it showed a clear advantage, with 40% more true-positive hits. This means that, as compared with a random prediction, seven additional top performers could be identified via the discriminant analysis.

3.2.4. Neural Network Analysis

On average, the MLP achieved a sensitivity of 75% and a specificity of 84% (Table 6). It could thus correctly identify approximately 12 of the 16 top performers and 126 of the 150 low performers. Taken together, these two parameters add up to 159%. The overall prognostic quality was thus 91%. If a test participant received a judgement as a later top performer, 33% of the predictions were correct. This means that, out of nine high-potential test participants, three were expected to reach the top level. Among the predicted low performers, only 3.1% of the predicted performance outcomes failed to apply. Thus, 3 out of 100 test participants reached the top level despite a negative individual forecast. Comparing the results with a random draw, the benefits of the MLP become apparent. The random sensitivity and specificity were 21.7% and 78.3%, and thus, the overall classification quality was 72.9%. Both the sensitivity and the specificity of the MLP clearly exceeded these random predictions. The Area Under ROC Curve was 0.831 (standard error = 0.056; 95% confidence interval: min = 0.721, max = 0.941).

Table 6. Sensitivity and specificity of the Neural Network (Multilayer Perceptron).

Multilayer Perceptron [1]		Predicted/Recommended		
		Low Performer	Top Performer	Percentage Correct
Observed/Existing	Top Performer	4	12	75%
	Low Performer [2]	126	24	84%
	Percentage (total)	78.3%	21.7%	91%

[1] A 5-fold cross-validation was performed a total of five times, and the respective test results were averaged. [2] Eight children could not perform at least one test, so they were excluded from this analysis.

3.3. Prognostic Validity of the Prediction Methods

In the linear discriminant analysis (see Figure 2; N = 166; eight children could not perform at least one test, so they were excluded from this analysis), the sensitivity could assume values of 24.1% (random result) to 100%, with a constant selection rate of 24.1%. The specificity reached values ranging from 75.9% (random result) to 84%. Thus, accordingly to Formula (8), the random (total) hit rate was 70.9% (see Formula (13)).

$$\text{Random Hit Rate} = \frac{0.241 \times 16 + (1 - 0.241) \times 150}{166} \approx 0.709 \quad (13)$$

Following Formula (9), the maximum hit rate was 0.855 (Formula (14)):

$$\text{Maximum Hit Rate} = 1 - \frac{|29 - 5|}{166} \approx 0.855 \quad (14)$$

For this calculation, therefore, the counter-probability is used to consider what the actual probability is if only the over-recommendations lead to an incorrect prediction. With 40 top performer predictions, 24 over-recommendations occurred, which corresponds to 14.5%—accordingly, this percentage is deducted from 100%. Finally, accordingly to Formula (11), it remains to calculate the RIOC value (Formula (15)):

$$\text{RIOC Value} = \frac{0.795 - 0.709}{0.855 - 0.709} \approx 0.588 \quad (15)$$

The calculations of the predictors of the other analysis methods are done analogously. All four methods revealed RIOC values in the range of 0.33–0.67 and thus illustrate a good classification result. However, the differences in the results of at least three out of four analyses were small.

The results for sensitivity and specificity are very different (Table 7). The logistic regression shows a sensitivity of only 37.5% but reaches the highest specificity, with 96%. The TRS and MLP prediction methods reach 75% or more in both parameters. Considering the positive predictive values, more than every second prediction of a top performer is correct for the logistic regression (50%). The top performer predictions for the TRS and MLP analysis methods only apply to every third (TRS = 30.8% and MLP = 33.3%). However, their negative predictive values are higher than those of the other two methods. *Youden's* J turns out to be the highest, with J = 0.59 for the MLP. This is closely followed by the TRS, with J = 0.579. Logistic regression performs worst among the four methods. Despite its high *Youden's* J, the TRS only achieves a value of F_1 = 0.437 in terms of the F_1 score, which is probably related to its low positive predictive value (30.8%). Among the four methods, the MLP performs best here, with F_1 = 0.462. This is not surprising because it has the highest sensitivity and also a high positive predictive value (33.3%). The RIOC values show values above 44% for all methods, with the MLP reaching the highest value of 68.1%. If, for example, a method would correctly predict the performance of another person, the RIOC value can increase between 1% and 8% depending on the performance group and method. Thus, another correct top performer prediction would increase the RIOC value of discriminant analysis from 58.8% to almost 67%.

Table 7. Comparison of the prognostic validity of the four analyses.

	Recommendation Score	(Binary) Logistic Regression	Linear Discriminant Analysis	Multilayer Perceptron
Sensitivity	0.750	0.375	0.688	0.750
Specificity	0.829	0.960	0.807	0.840
Selection Rate	0.224	0.072	0.241	0.217
Positive Predictive Value	0.308	0.500	0.275	0.333
Negative Predictive Value	0.970	0.935	0.960	0.969
Random Hit Rate	0.725	0.845	0.709	0.729
Hit Rate	0.822	0.904	0.795	0.831
Maximum Hit Rat	0.868	0.976	0.855	0.880
Youden's J	0.579	0.320	0.495	0.590
F_1 score	0.437	0.429	0.393	0.462
RIOC Value	0.678	0.447	0.588	0.681

Combining the predictions (union set) of the TRS with those of the MLP yields the highest sensitivity of all analyses (81.3%). Of 16 top performers, 13 are recognized as such. The selection rate is 27.7%. However, with a value of 78%, the specificity is below the values of the individual analysis methods. The lower specificity also causes the overall hit rate to drop to only 78.3%. The positive predictive value is 28.3%, and the negative predictive value is 97.5%. However, in comparison with the four individual forms of analysis, this combination of methods performs above average in *Youden's* J (0.593) and also in RIOC (0.741). Only the F_1 score is very low at 42%, which is due to the low positive predictive value. Looking at the intersection of the predictions of the three classification methods, TRS, MLP, and DA (see also Figure 3), the sensitivity reaches a value of 68.8% and the specificity a value of 88.7%. Due to the higher positive predictive value (0.393), the F_1 score (0.5) is the highest among all analysis methods and combinations. *Youden's* J reaches a value of 57.5%, and the RIOC value is 62.4%. A combination of the other analysis methods showed no significant improvements of the results.

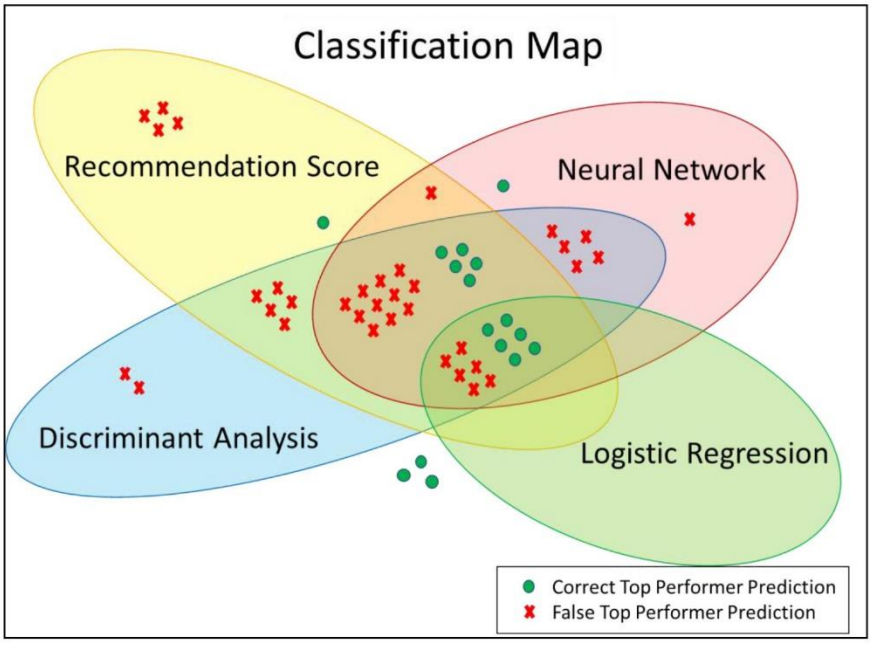

Figure 3. Classification map of correct and false top performer predictions by the four analyses.

3.4. Classification of Individual Tennis Players

Through the various prognostic procedures, 13 out of the total of 16 later top performers were correctly identified, whereas 3 athletes could not be detected correctly by any of the methods (Figure 3). The highest number of correct assignments was produced by the artificial neural network and the TRS. With both instruments, twelve of the later top performers were correctly predicted. However, the TRS also led to the highest number of incorrect classifications (n = 27). Its positive predictive value is correspondingly low (30.8%). The artificial neural network, on the other hand, produced 24 incorrect predictions. Its positive predictive value was 33.3%. Moreover, both methods could correctly predict at least one player who could not be predicted by the other analyses. By combining the two types of analyses, 13 top performers could be classified correctly as true positives.

Six top performers and six low performers were included in the talent prognosis by all four methods. Five other top performers were additionally recognized by all methods except logistic regression. Logistic regression predicted the smallest number of top performers overall, and only those who were also recognized by other methods. However, with only twelve predictions, it has the lowest selection rate, but at the same time, with six out of twelve, it has a high rate of hits in terms of its correct identification of top performers. A combination of the methods could also provide a high positive predictive value. If only those test participants who were jointly identified in the tennis recommendation score, discriminant analysis, and neural network were judged as top performers, 11 future top performers would be correctly classified out of only 28 predictions in total (39.2%).

The probabilities at which the analysis methods made correct top-performer recommendation are plotted in Table 8. The average values of the cross-validation procedures were converted into a percentage ranking system to facilitate comparison. The cutoff value for a top performer recommendation was uniformly set to 50%. Accordingly, given a percentage value above 50%, an athlete was classified as a top performer. Six top performers were recognized by all methods. Their recommendation probabilities ranged from 64% to 99%. Three top performers could not be classified as such. Their recommendation probabilities were in the range of 1% to 40%. If we look more closely at the probabilities with which the respective analysis methods correctly predicted 1 of the 16 top performers, it is noticeable that the logistic regression only provides a very low percentage value for many of the athletes.

Table 8. Probability outputs of the analysis methods.

	Probability of Being a Top Performer (%)			
	Recommendation Score	**Logistic Regression**	**Discriminant Analysis**	**Multilayer Perceptron**
Top Performer 1	99	90	99	99
Top Performer 2	97	86	98	97
Top Performer 3	93	68	97	96
Top Performer 4	96	71	96	90
Top Performer 5	83	64	95	90
Top Performer 6	89	68	90	95
Top Performer 7	91	15	68	71
Top Performer 8	79	33	85	88
Top Performer 9	55	7	65	70
Top Performer 10	66	18	63	68
Top Performer 11	57	9	58	68
Top Performer 12	67	16	41	32
Top Performer 13	44	3	23	52
Top Performer 14	40	5	24	37
Top Performer 15	20	2	9	29
Top Performer 16	33	1	1	15

Legend: The averaged test values of the cross-validation test sets (hold-out) were converted into a percentage ranking in order to compare them. Above a percentage value of 50%, a top performer recommendation was made.

4. Discussion

The aim of this study was to compare four statistical prediction methods used for the calculation of individual talent prognosis as part of talent identification. To investigate the prognostic validity of the talent-detection procedure, the match between the talent forecasts and actual later tennis performance was analyzed by means of various individual variables. In addition to the attempt to achieve as many correct positive talent predictions (sensitivity, recall) as possible, it is equally important to secure as many correct negative forecasts (specificity) as possible [26]. The quality of any prognostic attempt can then be determined based on the combination of both criteria. Given a random talent prediction rate of every second child judged as a future top performer, the sensitivity (8/16) and specificity (79/158) would be 50%. In this case, in total 87 children would receive correct predictions of their future performance, resulting in 50% (87/174) correct prognosis. In another example, with a random prediction of every fourth child to be a future top performer, the sensitivity would be 25% (4/16), and the specificity would be 75% (118/158). However, the total number of correct recommendations would be 70.1% (122/174). Although both examples are purely random ratings, the overall results are different. The larger group of low performers therefore distorts the overall talent prediction quality. Thus, the prognostic validity of talent identification cannot be determined only by the sensitivity and specificity of the chosen analysis. Both parameters are also highly dependent on the number of predictions made (selection rate). For example, if we consider (binary) logistic regression, the sensitivity (37.5%) is lowest. In this method, however, only twelve predictions of future top performers were made. This means that, even if all the predictions were correct, the sensitivity would still only reach a maximum of 75%, and this would still be as high as the sensitivities of the TRS and MLP (Table 7). The same applies to specificity. As soon as an analysis produces 17 or more predictions of future top performers, 1 of the cases must necessarily be wrong (there were only 16 observed top performers). With every overly "optimistic" false forecast made in this way, the highest possible specificity decreases by about 0.6% (1/158). Consequently, the comparison of the four analytical methods purely on the basis of sensitivity, specificity, and thus, the total hit rate is not sufficiently valid. To overcome this problem, it makes sense to consider the Youden Index [56], F_1 score, and RIOC value [57]. All three variables provide a prediction of the validity of prognostic methods. However, the RIOC value has an advantage in that it measures the accuracy of the method used on the basis of the maximum possible accuracy. Thus, the selection rate has almost no influence on the calculation, and therefore, different methods with different selection rates can be better compared. It makes sense to not always consider a method only in relation to a random result (J = 0) but to also include the maximum possible result. In the example of logistic regression, the Youden Index reaches a theoretical maximum of J = 0.75. However, the TRS could reach a value of up to 85.4% with the existing selection rate. This is 10% above the maximum potential *J* for logistic regression. A direct comparison of the two values is therefore sometimes difficult.

In practice, the positive predictive value is important because it indicates how many of the forecasts made actually predict a later top performer. Here, the logistic regression, with the calculation method and selection rate specified by SPSS, shows the largest positive predictive value. This method, therefore, may appeal to financially weak institutions because only a few particularly promising tennis players are recommended, and the available budget can thus be better calculated and channeled. For financially strong companies or clubs, the TRS may be appealing. Although an excessive number of athletes are predicted as future talent, the percentage of correctly predicted low performers is also the highest (97%). This means that this method has detected the most top performers (12 out of 16). Considering the results of the neural network in the classification map (Figure 3), similar results can be seen. In terms of both predictive values (positive predictive value 0.334 and negative predictive value 0.969), the neural network represents a good intermediate solution for talent identification. Therefore, this method seems to be well-suited to campaigns with average funding opportunities or average capital.

The four analysis methods used to calculate the performance prognosis served as examples of the currently predominant strategies in general talent identification campaigns. All four methods showed a high accuracy in their classification results (Table 7). The four calculated RIOC values were in the range between 0.447 and 0.681. Comparable studies on the prognostic validity by Marx and Lenhard [53] showed RIOC values in the range of 0.55. This value was exceeded by three of the four methods. Additionally, in studies by Hohmann t al. [26] and Siener and Hohmann [45] on the prognostic validity of talent predictions in soccer and table tennis, only RIOC values of 0.52 could be proven. The Youden Index of both studies reached values of J = 0.336 and J = 0.204, respectively. In studies by Pion et al. (2016) on the classification of drop-out rates among young gymnasts, RIOC values of 0.278 (MLP) to 0.537 (weighting) could be achieved via similar calculations (with MLP, discriminant analysis, and weighting). With these analysis methods, untalented athletes could be identified at an early stage, and thus funding costs could be reduced by about 33%. In a prognostic validity study of 117 soccer players (U14), significantly higher RIOC values were obtained by means of a logistic regression. A study by Sieghartsleitner et al. [56] showed RIOC values of 0.866 (Holistic Pattern Model) and 0.910 (Coach Assessment/Coaches' eye model). *Youden's* J was in the range of 61% to 77%. However, in addition to motor abilities, a number of psychological components, in-game performance, and familial support were also included in the calculation. When only motor abilities are considered in their study, the logistic regression has values of x_{RIOC} = 0.544 and J = 0.432, which are comparable with the results achieved here. Nevertheless, with the addition of the in-game performance (odds ratio = 12.5*) and familial support (odds ratio = 5.2*) survey parameters, a significantly higher prognostic validity is revealed for the predictions. Thus, in the study described [56], RIOC values could be significantly increased by about 20–35%, and *Youden's* J could be increased by about 20-30%.

Comparing the four prediction methods, differences can be found despite similar RIOC values. A specific recommendation for sports practice can therefore not be derived. All values (sensitivity, specificity, and positive and negative predictive value) of the discriminant analysis method turned out to be weaker than those of the TRS. From a statistical point of view, therefore, discriminant analysis was rather negligible in terms of its worth in this study. However, the other three methods have particular advantages, which may be more or less important depending on the situation. None of the four analysis methods were able to identify all future top performers. Even when all four methods were combined, three high-performance junior tennis players could not be identified. This shows that there are limits to any talent prognosis method [56] and that the future of an athlete cannot always be predicted perfectly. Nevertheless, a combination of prediction methods still seems to be the best solution. For this purpose, the combination of a linear (e.g., TRS) and a non-linear method (neural network) seems to be suitable. Depending on the combination method (union set or intersection set), a combination of methods gives the highest F_1 score (0.5 for intersection of TRS, MLP, and DA) or the highest RIOC value (0.741 for the union set of TRS and MLP) among all analysis options.

The example of the TRS also shows that the limit value chosen (e.g., z_{TRS} = 1.0 versus z_{TRS} = 1.3) to separate top and low performers makes a large difference in sensitivity and specificity. The summation of both values (overall benefit) varies between 100% (*Youden's* J = 0) and 157.9% (*Youden's* J = 0.579). The same applies to the separation parameters of the other three methods. Further studies are required to be able to make a more precise statement about this. Additionally, in the analyses, the selection rate fluctuates between 7.2% and 36.1%. Although this difference has no mathematical relevance for prognostic validity due to the use of the RIOC value, an approximation of the selection rate could provide more information about prediction calculations, and differences could then be shown on a classification map (Figure 3).

This study has certain limitations. In order to improve the learning phase and test phase of the calculation methods in a meaningful way (e.g., artificial neural network), more data collection is needed. However, such large samples are difficult to obtain, especially

in the talent area. On the one hand, there are not many athletes at the highest level of performance, and on the other hand, physical fitness and laboratory tests are sometimes very time consuming. Thus, there is usually a lack of time and participants for large sample numbers. Nevertheless, studies by Silva et al. [48] and Musa et al. [51] show that, even with smaller sample sizes ($N < 150$), valid results can be obtained even with neural networks. However, the size of the dataset must always be considered when interpreting analyses. Studies using neural networks with only a comparatively small number of participants, such as the study presented here, cannot be generalized and could turn out differently with a different sample. This problem can be reduced by averaging multiple computational runs with different hold-out partitions, but it can never be eliminated. The results of the MLP shown here must therefore be interpreted with caution. Additionally, with small datasets, overfitting can often occur with neural networks. This can be seen, for example, in the fact that the recommendation accuracy of the hold-out drops sharply as compared with the accuracy of the training and test datasets. In the analyses here, this was not the case, which may argue against overfitting.

For future analyses, it also seems useful to look beyond the classical classification methods to other promising analysis approaches. Giles et al. [58], for example, have already had good experiences with random forests. In their studies, F_1 values of up to 0.729 were found. A comparison between random forests and classical methods or other neural network solutions (e.g., radial basis functions) would be interesting for future studies.

Although the test batteries used thus far generally allow for a comprehensive performance survey, there is still room for improvement in the sport-specific case of tennis. For example, missing test items on agility in combination with decision-making as well as maximum (isometric) arm strength could be added to improve the prognostics [59,60]. In the future, this may lead to a better explanation of the variance between performance groups. In addition to motor test extensions, psychological test items could also provide further indications of potential future top performers [31,56,61–63]. Zuber et al. [31] have made the first promising attempts to integrate psychological tests into a talent identification campaign in the Swiss Soccer Federation. Nevertheless, despite optimized test tasks and the latest analysis methods, all talented athletes cannot always be found. Contrary to all predictions, some players develop into professional athletes, and very few career paths seem straightforward [64]. Thus, there are very individual pathways to the top of the athletic world, which are determined by many dynamic parameters, and each athlete reacts to these parameters in a unique way [65]. Holistic talent identification is therefore essential.

5. Conclusions

It has been shown that even in a very complex sport such as tennis, which requires motor competence as well as physical fitness [8,23,59,66], statistical analysis methods can be used to make reliable predictions of future success [67] based on the performance profiles of young tennis players. The performance profiles of 8-year-old tennis players can be determined in the context of talent identification by means of sport motor tests. Considering the prognostic validity of the prediction methods, all the results of an analysis must be taken into account, and the focus should not only be on high sensitivity [45]. A low sensitivity and specificity can also represent a good result for the talent identification procedure or selection rate used. To obtain insights into the actual quality of the method, it is worth considering the results obtained with random recommendations. Only here does the true value of the analysis become apparent. The RIOC value includes the random result and the maximum possible result in its calculation and is therefore a good predictor of prognostic validity [34,53]. Considering this value, all four methods show good overall results and stand out clearly from a random method. The prognostic validity of the general talent detection campaign in tennis investigated in this study was therefore of a medium to high quality. However, each calculation method had its particular advantages, which must be considered in practice. A combination of the methods provides the best results. Because methods are only as good as the available training data, it is important to continue to collect

further data and observe test participants over longer follow-up periods. Furthermore, it could be helpful to add more supplemental tests to the existing test batteries and to use new statistical methods in order to better analyze the differences between the later performance groups, as well as between the sexes, and thus increase the validity of early talent-detection campaigns.

Author Contributions: Conceptualization, M.S. and A.H.; methodology and analysis, M.S.; writing—original draft preparation, M.S.; writing—review and editing, I.F. and A.H.; supervision and project administration, A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This publication was funded by the German Research Foundation (DFG) and the University of Bayreuth in the funding program Open Access Publishing.

Institutional Review Board Statement: The study and research design was in line with the Declaration of Helsinki and was approved by the ethics committee of the Municipality of Fulda (Germany), as well as the State Office of School Education of Fulda (Germany).

Informed Consent Statement: Written informed consent was obtained from the participants to publish this paper.

Data Availability Statement: The data associated with the study are not publicly available but are available from the corresponding author on reasonable request.

Acknowledgments: We are grateful to the parents, children, and teachers who participated in the study. Furthermore, the authors thank the Fulda County Administration for their support.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Reilly, T.; Williams, A.M.; Nevill, A.; Franks, A. A multidisciplinary approach to talent identification in soccer. *J. Sports Sci.* **2000**, *18*, 695–702. [CrossRef] [PubMed]
2. De Bosscher, V.; de Knop, P.; van Bottenburg, M.; Shibli, S. A Conceptual Framework for Analysing Sports Policy Factors Leading to International Sporting Success. *Eur. Sport Manag. Q.* **2006**, *6*, 185–215. [CrossRef]
3. De Bosscher, V.; de Knop, P.; van Bottenburg, M. *Sports Policy Factors Leading to International Sporting Success*; VUB Press: Brussels, Belgium, 2007.
4. Baker, J.; Wilson, S.; Johnston, K.; Dehghansai, N.; Koenigsberg, A.; de Vegt, S.; Wattie, N. Talent Research in Sport 1990–2018: A Scoping Review. *Front. Psychol.* **2020**, *11*, 607710. [CrossRef] [PubMed]
5. Hoffmann, M.D.; Colley, R.C.; Doyon, C.Y.; Wong, S.L.; Tomkinson, G.R.; Lang, J.J. Normative-referenced percentile values for physical fitness among Canadians. *Health Rep.* **2019**, *30*, 14–22. [CrossRef]
6. Rowland, T. Counselling the young athlete: Where do we draw the line? *Pediatric Exerc. Sci.* **1997**, *9*, 197–201. [CrossRef]
7. Wiersma, L.D. Risks and benefits of youth sport specialization: Perspectives and recommendations. *Pediatric Exerc. Sci.* **2000**, *12*, 13–22. [CrossRef]
8. Faber, I.R.; Bustin, P.M.J.; Oosterveld, F.G.J.; Elferink-Gemser, M.T.; Nijhuis-van der Sanden, M.W. Assessing personal talent determinants in young racquet sport players: A systematic review. *J. Sports Sci.* **2016**, *34*, 395–410. [CrossRef]
9. Fuchslocher, J.; Romann, M.; Rüdisüli, L.R.; Birrer, D.; Hollenstein, C. Das Talentselektionsinstrument PISTE: Wie die Schweiz Nachwuchsathleten auswählt. *Leistungssport* **2011**, *41*, 22–27.
10. Golle, K.; Muehlbauer, T.; Wick, D.; Granacher, U. Physical Fitness Percentiles of German Children Aged 9–12 Years: Findings from a Longitudinal Study. *PLoS ONE* **2015**, *10*, e0142393. [CrossRef]
11. Pion, J. *The Flemish Sports Compass: From Sports Orientation to Elite Performance Prediction*; University Press: Ghent, Belgium, 2015.
12. Stemper, T.; Bachmann, C.; Diehlmann, K.; Kemper, B. Das Düsseldorfer Modell der Bewegungs-, Sport- und Talentförderung (DüMo). In *Talentdiagnose und Talentprognose: 2. BISp-Symposium: Theorie trifft Praxis*; Bundesinstitut für Sportwissenschaft, Ed.; Strauss: Kölln, Germany, 2009; pp. 139–142.
13. Fernandez-Fernandez, J.; Sanz-Rivas, D.; Mendez-Villanueva, A. A review of the activity profile and physiological demands of tennis match play. *Strength Cond. J.* **2009**, *31*, 15–26. [CrossRef]
14. Doherty, S.A.; Martinent, G.; Martindale, A.; Faber, I.R. Determinants for table tennis performance in elite Scottish youth players using a multidimensional approach: A pilot study. *High Abil. Stud.* **2018**, *29*, 241–254. [CrossRef]
15. Balyi, I.; Hamilton, A. *Long Term Athlete Development: Trainability in Childhood and Adolescence. Windows of Opportunity. Optimal Trainability*; National Coaching Institute British Columbia & Advanced Training and Performance Ltd.: Victoria, BC, Canada, 2004.
16. Knudsen, E.I. Sensitive periods in the development of the brain and behavior. *J. Cogn. Neurosci.* **2004**, *16*, 1412–1425. [CrossRef]

17. Watanabe, D.; Savion-Lemieux, T.; Penhune, V.B. The effect of early musical training on adult motor performance: Evidemce for a sensitive perios in motor learning. *Exp. Brain Res.* **2007**, 332–340. [CrossRef] [PubMed]
18. Anderson, D.I.; Magill, R.A.; Thouvarecq, R. Critical periods, sensitive periods, and readiness for motor skill learning. In *Skill Acquisition in Sport: Research, Theory, and Practice*, 2nd ed.; Hodges, N., Williams, A.M., Eds.; Routledge: London, UK; New York, NY, USA, 2012; pp. 211–228. ISBN 9780203133712.
19. De Bosscher, V.; Bingham, J.; Shibli, S.; van Bottenburg, M.; de Knop, P. *The Global Sporting Arms Race. An International Comparative Study on Sports Policy Factors Leading to International Sporting Success*; Meyer & Meyer: Aachen, Germany, 2008.
20. Bloyce, D.; Smith, A. *Sport, Policy and Development: An Introduction*; Routledge: London, UK; New York, NY, USA, 2010; ISBN 041540407X.
21. Houlihan, B.; Green, M. *Comparative Elite Sport Development: Systems, Structures and Public Policy*; Butterworth-Heineman: London, UK, 2008.
22. Pion, J.; Hohmann, A.; Liu, T.; Lenoir, M.; Segers, V. Predictive models reduce talent development costs in female gymnastics. *J. Sports Sci.* **2016**, *35*, 806–811. [CrossRef] [PubMed]
23. Roetert, E.P.; Garrett, G.E.; Brown, S.W.; Camaione, D.N. Performance Profiles of Nationally Ranked Junior Tennis Players. *J. Appl. Sport Sci. Res.* **1992**, *6*, 225–231. [CrossRef]
24. Kovacs, M. Tennis physiology: Training the competitive athlete. *Sports Med.* **2007**, 189–198. [CrossRef] [PubMed]
25. Filipcic, A.; Filipcic, T. The influence of tennis motor abilities and anthropometric measures on the competition successfulness of 11 and 12 year-old female tennis players. *Acta Univ. Palacki. Olomucensis. Gymnica* **2005**, *35*, 35–41.
26. Hohmann, A.; Siener, M.; He, R. Prognostic validity of talent orientation in soccer. *Ger. J. Exerc. Sport Res.* **2018**, *48*, 478–488. [CrossRef]
27. Baker, J.; Schorer, J.; Wattie, N. Compromising Talent: Issues in Identifying and Selecting Talent in Sport. *Quest* **2018**, *70*, 48–63. [CrossRef]
28. Höner, O.; Votteler, A. Prognostic relevance of motor talent predictors in early adolescence: A group- and individual-based evaluation considering different levels of achievement in youth football. *J. Sports Sci.* **2016**, *34*, 2269–2278. [CrossRef]
29. Höner, O.; Leyhr, D.; Kelava, A. 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* **2017**, *12*, e0182211. [CrossRef]
30. Pion, J. *Sustainable Investment in Sports Talent: The Path to the Podium Through the School and the Sports Club*; HAN University of Applied Sciences Press: Arnhem, Germany, 2017.
31. Zuber, C.; Zibung, M.; Conzelmann, A. Holistic patterns as an instrument for predicting the performance of promising young soccer players—A 3-years longitudinal study. *Front. Psychol.* **2016**, *7*, 1088. [CrossRef]
32. Urbano, D.; Restivo, M.T.; Barbosa, M.R.; Fernandes, Â.; Abreu, P.; Chousal, M.d.F.; Coelho, T. Handgrip Strength Time Profile and Frailty: An Exploratory Study. *Appl. Sci.* **2021**, *11*, 5134. [CrossRef]
33. Cairney, J.; Streiner, D.L. Using relative improvement over chance (RIOC) to examine agreement between tests: Three case examples using studies of developmental coordination disorder (DCD) in children. *Res. Dev. Disabil.* **2011**, *32*, 87–92. [CrossRef]
34. Farrington, D.P.; Loeber, R. Relative improvement over chance (RIOC) and phi as measures of predictive efficiency and strength of association in 22 tables. *J. Quant. Criminol.* **1989**, *5*, 201–213. [CrossRef]
35. Deutscher Tennis Bund. Jugendrangliste: Deutsche Ranglisten der Juniorinnen und Junioren. Available online: www.dtb-tennis.de/Tennis-National/Ranglisten/Jugend (accessed on 31 December 2020).
36. Hohmann, A.; Seidel, I. Scientific aspects of talent development. *Int. J. Phys. Educ.* **2003**, *40*, 9–20.
37. Bös, K.; Schlenker, L.; Albrecht, C.; Büsüch, D.; Lämmle, L.; Müller, H.; Oberger, J.; Seidel, I.; Tittlbach, S.; Woll, A. *Deutscher Motorik-Test 6-18 (DMT 6-18)*; Feldhaus: Hamburg, Germany, 2009.
38. Utesch, T.; Strauß, B.; Tietjens, M.; Büsch, D.; Ghanbari, M.-C.; Seidel, I. Die Überprüfung der Konstruktvalidität des Deutschen Motorik-Tests 6-18 für 9- bis 10-Jährige. *Z. Für Sportpsychol.* **2015**, *22*, 77–90. [CrossRef]
39. Klein, M.; Fröhlich, M.; Emrich, E. Zur Testgenauigkeit ausgewählter Items des Deutschen Motorik-Tests DMT 6-18. *Leipz. Sportwiss. Beiträge* **2012**, *53*, 23–45.
40. Bardid, F.; Utesch, T.; Lenoir, M. Investigating the construct of motor competence in middle childhood using the BOT-2 Short Form: An item response theory perspective. *Scand. J. Med. Sci. Sports* **2019**, *29*, 1980–1987. [CrossRef]
41. Meylan, C.; Cronin, J.; Oliver, J.; Hughes, M. Reviews: Talent identification in soccer: The role of maturity status on physical. physiological and technical characteristics. *Int. J. Sports Sci. Coach.* **2010**, *5*, 571–592. [CrossRef]
42. Carling, C.; Collins, D. Comment on "Football-Specific fitness testing: Adding value or confirming the evidence?". *J. Sports Sci.* **2014**, *32*, 1206–1208. [CrossRef] [PubMed]
43. Hohmann, A.; Fehr, U.; Siener, M.; Hochstein, S. Validity of early talent screening and talent orientation. In *Sport Science in a Metropolitan Area*; Platen, P., Ferrauti, A., Grimminger-Seidensticker, E., Jaitner, T., Eds.; University Press: Bochum, Germany, 2017; p. 590.
44. Willimczik, K. Determinanten der sportmotorischen Leistungsfähigkeit im Kindesalter. In *Kinder im Leistungssport.*; Howald, H., Hahn, E., Eds.; Birkhäuser: Stuttgart, Germany, 1982; pp. 141–153.
45. Siener, M.; Hohmann, A. Talent orientation: The impact of motor abilities on future success in table tennis. *Ger. J. Exerc. Sport Res.* **2019**, *49*, 232–243. [CrossRef]

46. Kolias, P.; Stavropoulos, N.; Papadopoulou, A.; Kostakidis, T. Evaluating basketball player's rotation line-ups performance via statistical markov chain modelling. *Int. J. Sports Sci. Coach.* **2021**, 174795412110090. [CrossRef]
47. Khasanshin, I. Application of an Artificial Neural Network to Automate the Measurement of Kinematic Characteristics of Punches in Boxing. *Appl. Sci.* **2021**, *11*, 1223. [CrossRef]
48. Silva, A.J.; Costa, A.M.; Oliveira, P.M.; Reis, V.M.; Saavedra, J.; Perl, J.; Rouboa, A.; Marinho, D.A. The use of neural network technology to model swimming performance. *J. Sports Sci. Med.* **2007**, *6*, 117–125.
49. Barron, D.; Ball, G.; Robins, M.; Sunderland, C. Artificial neural networks and player recruitment in professional soccer. *PLoS ONE* **2018**, *13*, e0205818. [CrossRef]
50. Lancashire, L.J.; Rees, R.C.; Ball, G.R. Identification of gene transcript signatures predictive for estrogen receptor and lymph node status using a stepwise forward selection artificial neural network modelling approach. *Artif. Intell. Med.* **2008**, *43*, 99–111. [CrossRef]
51. Musa, R.M.; Majeed, A.; Taha, Z.; Abdullah, M.R.; Maliki, A.B.H.M.; Kosni, N.A. The application of Artificial Neural Network and k-Nearest Neighbour classification models in the scouting of high-performance archers from a selected fitness and motor skill performance parameters. *Sci. Sports* **2019**, *34*, e241–e249. [CrossRef]
52. Heazlewood, I.; Walsh, J.; Climstein, M.; Kettunen, J.; Adams, K.; DeBeliso, M. A Comparison of Classification Accuracy for Gender Using Neural Networks Multilayer Perceptron (MLP), Radial Basis Function (RBF) Procedures Compared to Discriminant Function Analysis and Logistic Regression Based on Nine Sports Psychological Constructs to Measure Motivations to Participate in Masters Sports Competing at the 2009 World Masters Games. In *Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS)*; Chung, P., Soltoggio, A., Dawson, C.W., Meng, Q., Pain, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 93–101. ISBN 9783319245584.
53. Marx, P.; Lenhard, W. Diagnostische Merkmale von Screeningverfahren. In *Frühprognose Schulischer Kompetenzen*; Hasselhorn, M., Schneider, W., Eds.; Hogrefe: Göttingen, Germany, 2010.
54. Youden, W.J. Index for rating diagnostic tests. *Cancer* **1950**, 32–35. [CrossRef]
55. Kim, J.O.; Jeong, Y.-S.; Kim, J.H.; Lee, J.-W.; Park, D.; Kim, H.-S. Machine Learning-Based Cardiovascular Disease Prediction Model: A Cohort Study on the Korean National Health Insurance Service Health Screening Database. *Diagnostics* **2021**, *11*, 943. [CrossRef]
56. Sieghartsleitner, R.; Zuber, C.; Zibung, M.; Conzelmann, A. Science or Coaches' Eye?—Both! Beneficial Collaboration of Multidimensional Measurements and Coach Assessments for Efficient Talent Selection in Elite Youth Football. *J. Sports Sci. Med.* **2019**, *18*, 32–43.
57. Beißert, H.; Hasselhorn, M.; Lösche, P. Möglichkeiten und Grenzen der Frühprognose von Hochbegabung (Possibilities and limits of the early prognosis of giftedness). In *Handbuch Talententwicklung (Handbook of Talent Development)*; Stamm, M., Ed.; Hans-Huber: Bern, Germany, 2014; pp. 415–426.
58. Giles, B.; Kovalchik, S.; Reid, M. A machine learning approach for automatic detection and classification of changes of direction from player tracking data in professional tennis. *J. Sports Sci.* **2020**, *38*, 106–113. [CrossRef]
59. Ulbricht, A.; Fernandez-Fernández, J.; Mendez-Villanueva, A.; Ferrauti, A. Impact of fitness characteristics on tennis performance in elite junior tennis players. *J. Strength Cond. Res.* **2016**, *30*, 989–998. [CrossRef] [PubMed]
60. Pion, J.; Segers, V.; Fransen, J.; Debuyck, G.; Deprez, D.; Haerens, L.; Vaeyens, R.; Philippaerts, R.; Lenoir, M. Generic anthropometric and performance characteristics among elite adolescent boys in nine different sports. *Eur. J. Sport Sci.* **2015**, *15*, 357–366. [CrossRef]
61. Epuran, M.; Holdevici, I.; Tonita, F. *Performance Sport Psychology. Theory and Practice*; Fest: Bucharest, Romania, 2008.
62. Mosoi, A.A. Skills and Motivation of Junior Tennis Players. *Procedia Soc. Behav. Sci.* **2013**, *78*, 215–219. [CrossRef]
63. Zuber, C.; Zibung, M.; Conzelmann, A. Motivational patterns as an instrument for predicting success in promising young football players. *J. Sports Sci.* **2015**, *33*, 160–168. [CrossRef] [PubMed]
64. Elferink-Gemser, M.T.; Jordet, G.; Coelho-E-Silva, M.J.; Visscher, C. The marvels of elite sports: How to get there? *Br. J. Sports Med.* **2011**, *45*, 683–684. [CrossRef]
65. Phillips, E.; Davids, K.; Renshaw, I.; Portus, M. Expert performance in sport and the dynamics of talent development. *Sports Med.* **2010**, *40*, 271–283. [CrossRef] [PubMed]
66. Kovacs, M. Applied physiology of tennis performance. *Br. J. Sports Med.* **2006**, *40*, 381–386. [CrossRef]
67. Brouwers, J.; Bosscher, V.; de Sotiriadou, P. An examination of the importance of performances in youth and junior competition as an indicator of later success in tennis. *Sport Manag. Rev.* **2012**, *15*, 461–475. [CrossRef]

Der Autor

Maximilian Siener

Bereits während seines Studiums an der Friedrich-Alexander-Universität Erlangen-Nürnberg arbeitete Maximilian Siener am Institut für Sportwissenschaft der Universität Bayreuth. Nach seinem Bachelor-Abschluss im Fach Mathematik und dem ersten Staatsexamen in den Fächern Mathematik und Sport für das Lehramt an Gymnasien promovierte er an der Universität Bayreuth zum Thema „Validitätsstudie zur Talentforschung – Prüfung der prognostischen Validität von Talentidentifikation“.

Maximilian Siener ist Mitglied des Koordinierungsausschusses zum bayerischen Staatsexamen im Bereich Trainings- und Bewegungswissenschaft und beschäftigt sich als wissenschaftlicher Mitarbeiter mit den Forschungsschwerpunkten trainingswissenschaftliche Leistungsdiagnostik, Talentidentifikation und Spielanalyse. Außerdem ist er als Gutachter für das Journal „PLOS One“ und das „International Journal of Racket Sports Science“ tätig sowie Mitglied der Fachgruppe „Fitness“ zur „Report Card Germany“ (Bewegungszeugnis für Deutschland).

Die langjährige Betreuung des Fuldaer und Bayreuther Bewegungs-Checks sowie sein mathematisches Interesse an statistischen Methoden motivierten ihn zu einer empirischen Forschungsarbeit zum Thema Talentidentifikation.