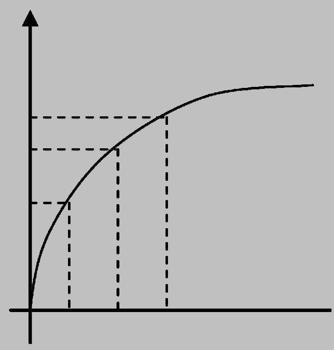
Diskussionspapiere des
Europäischen Instituts für Sozioökonomie e. V.

Working Papers of the
European Institute for Socioeconomics



A machine learning approach to ‘revisit’ specialization and sampling in institutionalized practice

Michael Barth, Eike Emrich & Arne Güllich

Nr. 26

2018

Dr. Michael Barth
Fachbereich: Sportwissenschaft
Adresse: Leopold-Franzens-Universität Innsbruck
Institut für Sportwissenschaft
Fürstenweg 185
6020 Innsbruck
Österreich

E-Mail: michael.barth@uibk.ac.at

Prof. Dr. Eike Emrich
Fachbereich: Sportökonomie und Sportsoziologie
Adresse: Universität des Saarlandes
Sportwissenschaftliches Institut
Campus
66123 Saarbrücken
Tel: 0049 (0)681 302 4170
E-Mail: e.emrich@mx.uni-saarland.de

Prof. Dr. Arne Güllich
Fachbereich: Geistes- u. sozialwiss. Sportwissenschaft
Adresse: Universität Kaiserslautern
Fachgebiet Sportwissenschaft
Erwin Schrödinger-Straße
67663 Kaiserslautern
Tel: 0049 (0)631 205 4966
E-Mail: guellich@sowi.uni-kl.de

Europäisches Institut für Sozioökonomie e. V. / European Institute for Socioeconomics

c/o Universität des Saarlandes

Sportwissenschaftliches Institut
Arbeitsbereich Sportökonomie und Sportsoziologie
Campus, Gebäude 8.1

66123 Saarbrücken

http://www.soziooekonomie.org

A machine learning approach to “revisit” specialization and sampling in institutionalized practice

Michael Barth, Eike Emrich & Arne Güllich

Abstract:
Apart from a broad consensus statement stressing the essential role of practice for achieving success in international senior-level competitions, the nature and scope of developmental participation leading to that extraordinary success in sports have been controversially discussed in international literature for many years.
The aim of this paper is to contribute to the existing body of literature in two respects: first, by reviewing the existing literature comparing the developmental activities of internationally and only nationally successful senior athletes. Second, a new methodical approach combining decision trees and gradient boosting is applied to data from a previous study, the results of which were internationally published. This does not only allow for the realization of a multivariate analysis (robustness check), but also gives reasonable hope of achieving a relatively better explanation than with the procedures applied in the past. The approach is realized by means of Extreme Gradient Boosting (XGBoost under the R environment). The results indicate that some formerly found differences in the volume of structured practice in main and other sports between internationally and only nationally successful athletes may represent rather artifacts of uncontrolled age effects than variables that differentiate the groups. In the context of the specialization-diversification debate, the present results indicate that from today’s perspective there is a debate about a “production function”, the structure of which is unknown. Obviously, practice-related recommendations on developmental practice volume are expressions of highly rationalized myths rather than evidence-based efficient norms.

Keywords:
talent development; international sporting success; machine learning; extreme gradient boosting; deliberate practice

Introduction

Independent of the importance of genetic variation for reaching exceptional performance, authors seem to agree on the essential role of (structured) practice (input side) for success in international senior-level competitions (hereinafter: international senior success (ISS), representing the output side) (Güllich, 2017; Tucker & Collins, 2012).

Apart from this broad consensus statement, the nature and scope of activities and developmental participation that lead to extraordinary success in sports have been controversially discussed in international literature for many years (Côté, Baker, & Abernethy, 2007; Güllich, 2017; Sieghartsleitner, Zuber, Zibung, & Conzelmann, 2018). Given the increasing number of international competitions with growing governmental involvement (Heinilä, 1982; Houlihan & Green, 2008) and the belief (of system participants) that senior sports success is producible (De Bosscher & De Rycke, 2017; De Bosscher, Shibli, Westerbeek, & Van Bottenburg, 2015), there is a clear need for more detailed research on the career of athletes achieving ISS from different theoretical perspectives (Barth, Emrich, & Daumann, 2018; Barth, Güllich, & Emrich, 2018; Güllich & Emrich, 2014).

This paper contributes to the existing body of literature in two respects: first, by reviewing the existing literature on developmental activities – i.e. the volume of domain-specific structured practice (institutionalized practice in organized settings such as sports clubs, extracurricular high-school sports or sport academies in the main sport of the athlete; hereinafter: V-IPMS) as well as outside domain-specific structured practice (institutionalized practice in other sports; hereinafter: V-IPOS) and, additionally, the age at the beginning of IPMS – of athletes achieving at least once ISS (hereinafter: I-Senior) and (internationally) less successful senior athletes (hereinafter: N-Senior), laying special focus on the statistical methods applied in the empirical studies. Second, a new methodical approach, combining decision trees and gradient boosting, will be applied to revisit data from an early study by Emrich and Güllich (2005)[1]. This approach does not only allow for the realization of a multivariate analysis, but also gives reasonable hope of achieving a relatively better explanation than with procedures applied in the past.

This paper is structured as follows: Originating from two well-established theoretical frameworks about developmental activity patterns leading to expertise and/or extraordinary success in sports, the question of operationalization (on the input and output side) for empirical testing is discussed. This enables the restriction of literature to be reviewed. After a brief presentation of the excluded articles and reasons for excluding them, the findings of the literature review are reported. Then, the methods used in the empirical study are described. Afterwards, the results are presented, followed by a discussion of the study's limitations as well as future directions in the study of athlete development in sport. Finally, a conclusion is drawn.

Problem and state of research

The problem and state of research part of this paper is aimed at setting thematic limits: A restrictive approach is applied in order to avoid extrapolating the scope of findings and inappropriate considerations. In this context, reasons for including the examined articles and excluding others are given. In this, both the input side and output side were examined.

Input side

The debate on developmental activity patterns leading to expertise and/or exceptional success in sports was particularly marked by a controversial discussion of the "Deliberate Practice" (Ericsson, 2006; Ericsson, Chase, & Faloon, 1980; Ericsson, Krampe, & Tesch-Römer, 1993) concept (hereinafter: DP) and the "elite performance through sampling and deliberate play" pathway of the "Developmental Model of Sport Participation" (Côté et al., 2007) (hereinafter: SDPL-DMSP) (Sieghartsleitner et al., 2018). Côté et al. (2007) themselves describe this pathway as "elite performance through sampling." However, we feel this might be confusing since sampling is only one feature characterizing this pathway. Therefore, we use SDPL for sampling (S) and deliberate play (DPL). It should be noticed that SDPL is marked by three dimensions (sampling, inherent enjoyment/playfulness, and peer-led/youth-led/loosely supervised). However, DPL already incorporates two of these dimensions (playfulness and peer-led/youth-led) (cf. Côté & Erickson, 2015). "Elite performance through early specialization", the second pathway of the DMSP (Côté et al., 2007), aligns with the DP (Güllich, 2017).

Contrasting the DP with the SDPL-DMSP enables the construction of a three-dimensional space (Figure I). On the first axis we can have the personal value the activity provides to the participant (intrinsic values, i.e. activities done for inherent enjoyment/playfulness vs. extrinsic values, i.e. activities performed to improve skills or performance), on the second axis the social structure of the activity (peer-led/youth-led/loosely supervised vs. coach-led/adults-led/highly supervised) and on the third axis the domain specificity of the activity (sampling/diversification vs. specialization) (for the first two dimensions cf. Côté & Erickson, 2015). Although all dimensions seem to form a continuum with a broad spectrum of different possibilities, many researchers treat them as dichotomous counterparts (Sieghartsleitner et al., 2018).

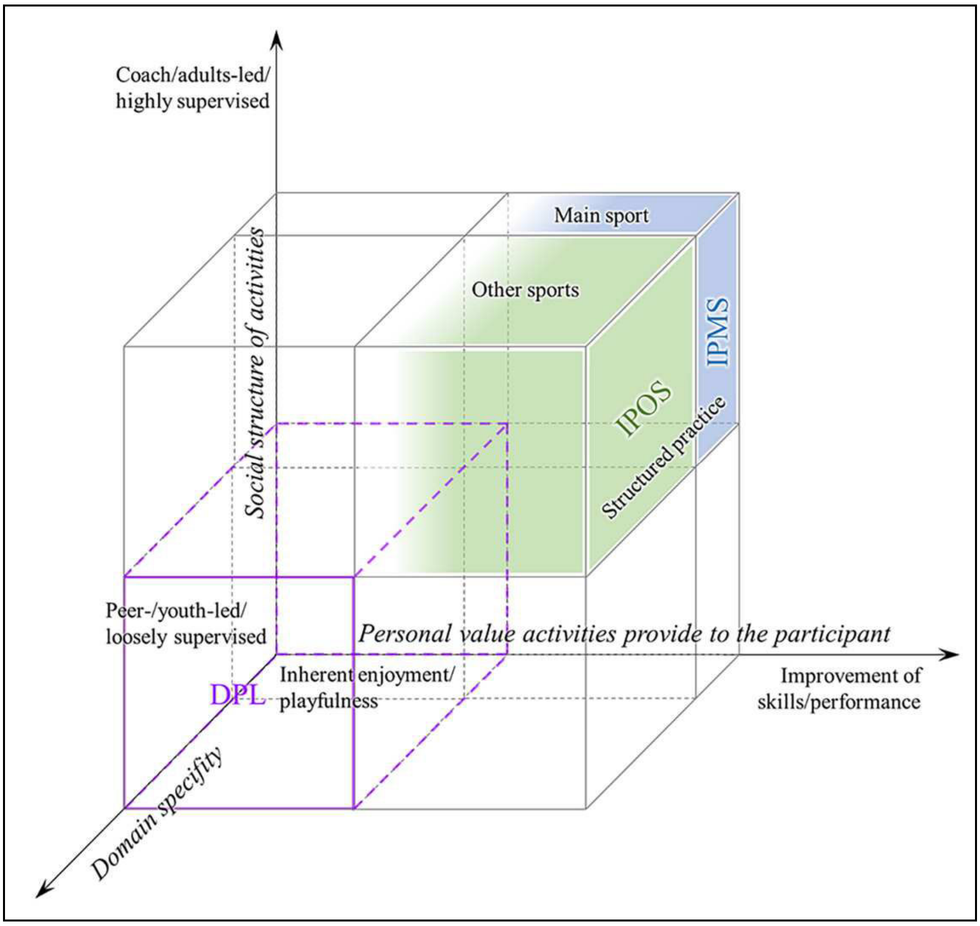


Figure I. Schematic illustration of specialization and sampling in structured/institutionalized practice

Not least because of its strict definition, DP has since been heavily criticized (e.g., Abernethy, Farrow, & Berry, 2003; Helsen, Starkes, & Hodges, 1998). However, in literature there seems to exist a broad consensus that the concept of DP is related to the idea of maximizing domain specificity/specialization and structured (coach-led/adults-led/highly supervised) practice (performed to improve skills/performance) from an early stage on. In contrast, the SDPL-DMSP recommends early involvement in several sports (sampling) in combination with a high amount of DPL (inherent enjoyment/playfulness and peer-led/youth-led/loosely supervised) and a low extent of DP at an early age ("sampling years": 6-12 years; "specializing years": 13-15 years; "investment years": age 16 years and older (Côté et al., 2007, pp. 196–197)).

Testing these two concepts empirically, the first question of operationalization arises on the input side. Particularly in the context of the two dimensions personal value the activity provides to the participant and the social structure of activity different approaches are to be observed: firstly, studies categorizing activities on the basis of their reported nature and purposes (e.g., Berry, Abernethy, & Côté, 2008; Memmert, Baker, & Bertsch, 2010) and, secondly, studies using the organizational structure of the athletes' activities to distinguish between structured practice and DPL. In the context of the latter, authors assume that institutionalized sports participation (i.e. institutionalized practice in organized settings such as sports clubs, extracurricular high-school sports or sport academies) is connected to structured practicing activities, whereas activities outside sports clubs are associated with a playful and loosely supervised form of participation (Côté, Baker, & Abernethy, 2003). Basically, Côté et al. (2003, p. 95) describe structured practice as "activities typical of organized sport". Although this approach seems to be appropriate to distinguish between structured practice (IPMS and IPOS) and DPL (in the main sport and other sports), it seems to be questionable, if a differentiation between IPMS and DP as defined by Ericsson et al. (1993) is possible. Considering Côté et al. (2003, p. 95), who compare „free play", "deliberate play", "structured practice" and "deliberate practice", it seems reasonable to view DP as a special form of structured practice and therefore of IPMS. However, we consider DP as a special form of structured practice, not only because of its sport specificity, but also because of its characteristics on the different dimensions as described by Côté et al. (2003). Consequently, we do not asses IPMS and DP as congruent. Nevertheless, it seems to be appropriate to draw conclusions from IPMS results for the DP concept. However, these conclusions have to be considered with caution.

Furthermore, it seems reasonable to describe IPOS as structured practice in other sports, thus as highly supervised performance oriented activities in other sports and therefore with low domain specificity. Consequently, this paper contributes primarily to the specialization/sampling-debate in structured/institutionalized practice, but not to the DP/SDPL-DMSP-debate. This differentiation is not only useful to draw appropriate conclusions during the course of the subsequent discussion, but also helps to restrict the review of the existing literature and define its variables of analysis (V-IPMS and V-IPOS in childhood and adolescence).

After defining our main variables of interest on the input side we take a closer look on the side of the output, hence the question of defining expertise (and success, respectively).

Output side

Many articles with a broadly diversified thematic spectrum have been published on the topic of talent development in sports, including several articles reviewing the literature on developmental patterns of expert sports performers (e.g., Baker, 2003; Baker, Cobley, & Fraser‐Thomas, 2009; Baker & Young, 2014; Côté et al., 2007; Coutinho, Mesquita, & Fonseca, 2016; Davids & Baker, 2007; Davids, Güllich, Shuttleworth, & Araújo, 2017; Güllich & Cobley, 2017; Güllich & Emrich, 2014; Rees et al., 2016; Vaeyens, Güllich, Warr, & Philippaerts, 2009). In 2014 Macnamara, Hambrick, and Oswald published a meta-analysis on "Deliberate Practice and Performance in Music, Games, Sports, Education, and Professions" (as well as a corrected version of it in 2018); additionally, a meta-analysis on the relationship between DP and performance in sports was presented by Macnamara, Moreau, and Hambrick in 2016.

Examining empirical articles in more detail shows that most of them are based on ‘the relative approach‘, comparing groups with higher and lower performance levels (Coutinho et al., 2016), while, however, using different and inconsistent criteria for what defines an expert athlete. Mostly, the differentiation of experts and non-expert athletes is based on the level of proficiency attained. Furthermore, criteria such as playing in highest division, being part of a national team or achieved competition results are used. Asking national coaches for the “best” athletes in a specific sport is also commonly used to identify the experts (Coutinho et al., 2016). Interestingly, Coutinho et al. (2016) stated that non-expert athletes are usually only classified by their failure to meet expert athletes’ criteria. Due to the fact, that results commonly base on comparisons of experts with non-experts (relative approach), we have to pay the same attention to the “level” of the non-experts group as we do for the group of experts. We clearly follow Coutinho et al. (2016) statement that a detailed description of the criteria defining an athlete’s level of expertise should be given. However, we want to emphasize that these criteria should not only be given to increase the understanding of what an expert is, but also to better understand what a non-expert is, because the way this group was built and possibly restricted downward also influences the findings. In the light of the results of Emrich and Güllich (2005), Güllich and Emrich (2006), and Güllich and Emrich (2014), Güllich (2014, p. 764) states that “conditions for international senior-level success cannot be concluded by simply extrapolating the scope of findings from athletes with moderate success level or from junior athletes.” This means, when determining the level of the athletes and assigning them to groups for comparison in prospective empirical studies, we should at least take two variables into account: first, the athletes’ age level, which is commonly determined in studies by the level of the competition/league athletes are competing in and/or on behalf of the teams athletes are affiliated to. Some studies additionally apply an age-limit, which is oriented at the international competition regulations of the athletes’ respective sports. Secondly, a variable for distinguishing experts and non-experts (e.g., success in competition) and the level each group has to achieve (e.g., medalist at Senior World Championship) should be considered.

Due to the paper’s purposes, the literature review therefore comprises (based on a relative approach) studies comparing developmental activities (i.e. V-IPMS and/or V-IPOS[2] and additionally the age at onset of IPMS) of I-Senior and N-Senior. Due to the fact, that in this context clarifying and operationalizing are normative issues, we decided to inductively develop our classification system and therefore the limits for the determination of I-Senior and N-Senior. I-Senior was defined considering the entire range of articles, namely as athletes having achieved ISS at least once. This means that studies did not compellingly have to have used a cut-off age to be included in our review.

The I-Senior’s peers for comparison – the less successful athletes (N-Senior) – are to be described as athletes belonging to a nation’s best senior athletes (later we will further distinguish between national best and nation’s best athletes), but having never achieved ISS. Determining the senior level was possible through the athlete’s affiliation to a nation’s senior national team or national squad or participation in open-age leagues/competitions. Again, studies have not to have used further limits for athletes' age. We would like to state that this does not seem completely unproblematic because national squads of sports governing bodies commonly encompass junior athletes (and even youth athletes). The occurrence of a situation where youth athletes were members of a sports governing body's national squad and competed in senior-level national and international competitions is conceivable. Therefore, a further restriction of the age level, e.g. as mentioned above, should be used in future studies (for application e.g., Güllich & Emrich, 2014). Furthermore, matching procedures (cf. Güllich, 2017, 2018) or covariates (cf. De Bosscher & De Rycke, 2017) should be applied to control for age. Table I shows the respective (age level and level of achievement) classification of I-Senior and N-Senior.

Table I. I-Senior and N-Senior classification (age level and level of achievement).

[a] Although there exists a correlation between level of competition and age limit for (international) competitions regulations, the one does not automatically determine the other. However, it can be assumed that in view of the criteria used in most of the studies the athletes' age level can be described as senior. Nevertheless, exceptions might exist, which seem to be problematic for our results in the context of the applied statistical procedures.

		Age level[a]	
		Youth/junior/sub-adult	Senior/adult
Success level	World level 1, 2, 3		I-Senior
	Continental level		
	National level		N-Senior
	Below national level		

Beside age level of athletes and achieved success level in competitions, we used a third variable to exactly characterize the groups of the examined studies, called selection level. It distinguishes between athletes selected for a senior national team (= highest level of squads of a national sports governing) and athletes affiliated to a club playing in the highest national league/division or to a senior squad of a national sports governing body, but not to the senior national team. Therefore, less successful athletes have to be at least national best athletes. However, it should be noted, that a description of N-Senior only by their level of selection seems to be problematic, because nations' levels in a certain sport are not congruent. Unfortunately, the same is true for success in national competitions.

Tables II and III present a detailed description of the success level in competitions and the selection level.

Table II. Description of levels of success.

Success level	Description
Senior World level 1	1st to 3rd place at Olympic Games or Senior World Championships
Senior World level 2	4th to 10th place at Olympic Games or Senior World Championships; 1st to 10th place at grand slams, Senior World-ranking competitions or in Senior World-ranking lists
Senior World level 3	11th to 16th at Olympic Games, Senior World Championships, grand slams, Senior World-ranking competitions or in Senior World-ranking lists
Senior Continental level	1st to 10th place at Senior Continental-level competitions (European Championships, Pan American Games, Asian Games)
Senior National level	Success < Senior Continental level BUT athletes have to belong to the group of the nation's best or national best

Table III. Levels of selection[a].

[a] Athletes were classified via the affiliation given in the respective study or via their level of success (e.g., medalists at Senior World Championships have to part of the senior national team). However, this means the affiliation can correspond to the time of the survey or the time at which the athlete achieved the success.

Nation's best	Athletes affiliated to a nation's senior national team
National best	Athletes affiliated to a club playing in the highest national league/division or to the national squad of a sports governing body, but not to the senior national team

These operational definitions were primarily used to select studies appropriate for inclusion in our review. Furthermore, results have to be reported in a peer-reviewed journal.[3] This review is aligned to the works by Güllich and Emrich (2014), Rees et al. (2016), and Davids et al. (2017), but new articles were added. Before the findings of this review are concisely described, the excluded articles with reasons for the decision to omit them are listed. We feel this to be important since reviews commonly compromise a broader range. However, those approaches may run the risk of extrapolating results; thus, we used quite a restrictive approach to avoid inappropriate considerations.

Excluded Articles

Table IV shows the excluded articles[4] and reasons for excluding them from this review.

Table IV. Articles excluded from the review.

No.	Source	Sports	Reasons for exclusion
1	Carlson, 1990	Tennis	The success level of the less successful group of athletes could not be determined. (Condition used in the study: not ranked in the ATP list, which means athletes not being successful or athletes who had already retired.)
2	Hodges & Starkes, 1996	Wrestling	The expert-group athletes' success could not be determined exactly. Some of them have to be considered as not internationally successful. Furthermore, the success of the non-experts could not be identified. Some of them are below provincial/regional level.
3	Helsen et al., 1998	Soccer, Field Hockey	This investigation consists of two empirical studies. Study 1: The level of the expert-group is too low. The group of experts/international players consists of 12 players (out of 17) selected for the 1994 FIFA World Cup. The team were eliminated in the round of 16 and ranked 24th in the FIFA world ranking (according to Fédération Internationale de Football Association; Fussballdaten.de). The level of non-experts (and further group) is below national level. Non-experts/national players were semiprofessionally involved in soccer in the first and second division. The third group was built of players performing in the third and fourth division. Study 2: Only "most" athletes of the expert group fulfill the criteria of the success level (most of them were selected for the European Championships in 1995, where Belgium was ranked 5th (according to Van Rossum, 2009) (p. 24).
4	Van Rossum, 2000	Field hockey	Discrimination of groups for comparison was based on selection for the national team. Hours of practice were a reconstruction of "the typical Dutch career" (p. 454).
5	Baker, Côte, & Abernethy, 2003	Basketball, netball, field hockey	The success level of the group of less successful athletes is lower than national level. (Athletes not having participated beyond state or provincial level.)

6	Duffy, Baluch, & Ericsson, 2004	Darts	The success level of the less successful group of athletes could not be determined exactly. Criteria were no attainment on an international level of performance, to have amateur status and to have played country darts for at least 15 years.
7	Hodges, Kerr, Starkes, Weir, & Nananidou, 2004	Triathlon	No relative approach was applied. Performance and not success was used as responding variable for regression analysis. There was a wide range of “skill level” from athletes who had competed at World Championships to some who had participated in races on local level only. However, success was not mentioned.
8	Baker, Côté, & Deakin, 2005	Triathlon	Groups were distinguished on behalf of their finishing times in races. Experts are best described as national-level athletes.
9	Baker, Côté, & Deakin, 2006	Triathlon	Master-age triathletes were examined. Groups were distinguished on behalf of their finishing times in races.
10	Johnson, Tenenbaum, & Edmonds, 2006	Swimming	3 out of 19 athletes were classified according to their success in junior classes (i.e. age groups of 12 and 13 and junior competition). Information is taken from Johnson (Johnson, 2006).
11	Berry et al., 2008	Australian Football	The responding variable was not success but game-based perception and decision-making skill.
12	Johnson, Tenenbaum, Edmonds, & Castillo, 2008	Swimming	Although the group of experts consisted of successful athletes at World level, some of them were classified due to the fact that they had set at least one swimming world record (in an age-restricted class). No quantitative data on IPMS or IPOS was available.
13	Weissensteiner, Abernethy, Farrow, & Müller, 2008	Cricket	Experts/skilled players were those having attained state and/or national representation. Non-experts/less skilled players were those who had not attained regional representation.
14	Ronbeck, Dunnagan, & Stewart C., 2009	Nordic Ski	Not available. Not published in a peer-reviewed journal.
15	Memmert et al., 2010	Basketball, field hockey, handball, soccer	No comparison of I-Senior and N-Senior.
16	Moesch, Elbe, Hauge, &	Cgs-sports[5]	With the aim to eliminate age-bias, medals at junior championships were used for determining success.

	Wikman, 2011		
17	Gulbin, Weissensteiner, Oldenziel, & Gagné, 2013	Various Olympic and non-Olympic sports	No relative approach.
18	Moesch, Hauge, Wikman, & Elbe, 2013	Team sports	With the aim to eliminate age-bias, medals at junior championships were used for determining success.
19	Huxley, O'Connor, & Larkin, 2017	Track and field	No relative approach.
20	Hendry & Hodges, 2018	Soccer	The group of highest level (experts) was determined as follows: "whether they had been selected to play first-team, adult soccer in the UK, what we termed 'Adult-professional'" (p. 83).
21	Leyhr, Kelava, Raabe, & Höner, 2018	Soccer	Appearance in one of the five highest German soccer divisions was used to define the expert-group.
22	Sieghartsleitner et al., 2018	Soccer	The participants were junior athletes. The highest level was set to getting at least one nomination for one of the U15 to U18 National Teams.

Included Articles

Based on the criteria stated above, we identified ten studies, including one additional study (De Bosscher & De Rycke, 2017) although it does not report V-IPMS or V-IPOS. The rationale behind is that this study has been published recently and includes a huge sample size as well as a definition for the comparison groups applied. Although the scope of our study deals particularly with the question of the success-relevance of V-IPMS and V-IPOS, it seems important to show the results of our analysis, which are limited to comparing the group's success level as presented in Table V.

Table V. Success levels of I-Senior and N-Senior in included studies.

I-Senior
N-Senior
I: individual sport(s), T: team sport(s)
[a] The study distinguished between three levels of success. Here, the highest and lowest levels (national level) are compared.

No.	Source	Success level					I/T
		Senior World level 1	Senior World level 2	Senior World level 3	Senior Continent-al level	Senior National level	
1	Gibbons, Hill, McConnell, Forster, & Moore, 2002						I/T
2	Güllich & Emrich, 2006						I/T
3	Law, Côté, & Ericsson, 2007						I
4	Vaeyens et al., 2009						I/T
5	Güllich, 2014						T
6	Güllich & Emrich, 2014						I/T
7	Hornig, Aust, & Güllich, 2016						T
8	De Bosscher & De Rycke, 2017[a]						I/T
9	Güllich, 2017				Only Top 1		I/T
10	Güllich, 2018				Only Top 3		I

Our analysis shows that there are differences in the determination of the success levels for I-Senior, but also for their national-level counterparts. Due to the fact, that all studies used a relative approach we do not have to take only their differences into consideration, but also the “difference in the differences of groups”.

Additionally to the variation in the absolute success level of both groups as well as “the difference in the differences of groups” between the studies, the analysis shows that some studies have additively applied age limits, while others have not. Whether not applying age limits is problematic, depends on the determination of the athletes by success level and/or selection level.

Furthermore, some studies compare successful national team athletes with less successful ones, whereas others compare successful national team athletes to national best athletes. Summing up, although a restrictive definition was used, studies have compared different groups of I-Senior and N-Senior. In addition, since I-Elite were older than N-Elite in some studies (cf. Table VI), the question arises whether these compared successful and less successful athletes or rather “not yet successful athletes.” In this context, the rather open definition of less successful athletes according to their degree of success in combination with waiving any restriction in the age limit is a concern. With regard to the statistical procedures to be applied, these problems clearly speak in favor of the use of covariates (e.g., the age of the athletes at the time of survey) or the introduction of matching procedures to control, among other factors, the age of the athletes.

Before describing the included studies’ samples and the reported ages at the onset of IPMS, it should be mentioned that four of the ten studies have used data from an identical survey. These are marked in the following Table VI.

Table VI. Description of the sample, age at onset of IPMS and its success relevance within the included studies.

+ = I-Senior younger than N-Senior
– = I-Senior older than N-Senior
O = no sig. difference
n.a. = information not available
a Effect sizes were calculated using an estimation function for hedges (g*) (Fröhlich & Pieter, 2009). It should be noted that, in the absence of further details in the contributions, the effect strengths were calculated using the total sample sizes.
b Results were not reported for the whole group.
c Using data from an identical survey.
d Retrieved from Vaeyens et al. (2009).
e According to the description of the achievement levels. The study distinguished between three levels of success. Here, the highest and lowest level (National level) are compared.
f Unclear, whether the age is related to the onset of structured/institutionalized practice or overall participation.
g No inferential statistical results were reported.

No.	Source	Sample description		Age at onset of IPMS			
		I-Senior	N-Senior	I-Senior	N-Senior	Age difference/ success relevance	effect size[a]
1	Gibbons et al., 2002	N = 283 Sex ratio (m:f) = 157:126 Age = n.a.	N = 533 Sex ratio (m:f) = 318:215 Age = n.a.	n.a.[b]	n.a.[b]	n.a.[b]	n.a.
2[c]	Güllich & Emrich, 2006	N = n.a. (calculated: 680*0.63)[d] Sex ratio (m:f) = n.a. Age = n.a.	N = n.a. (calculated: 680*0.38)[d] Sex ratio (m:f) = n.a. Age = n.a.	11.4 (±4.6) years	10.2 (±4.1) years	–	n.a.
3	Law et al., 2007	N = 6 Sex ratio (m:f) = 0:6 Age = 16.0 (±0.4)	N = 6 Sex ratio (m:f) = 0:6 Age = 18.3 (±1.6)	n.a.	n.a.	O	n.a.

4[c]	Vaeyens et al., 2009	N = n.a. (calculated: 680*0.63) Sex ratio (m:f) = n.a. Age = n.a.	N = n.a. (calculated: 680*0.38) Sex ratio (m:f) = n.a. Age = n.a.	11.4 (±4.6) years	10.2 (±4.1) years	–	n.a.
5	Güllich, 2014	N = 16 Sex ratio (m:f) = 16:0 Age = 26.6 (±3.1) years	N = 19 Sex ratio (m:f) = 19:0 Age = 22.3 (±3.6)	4.9 (±1.4) years	5.5 (±1.9) years	O	0.35
6[c]	Güllich & Emrich, 2014	N = 387 Sex ratio (m:f) = n.a. Age = n.a.	N = 213 Sex ratio (m:f) = n.a. Age = n.a.	11.4 (±4.7) years	10.1 (±4.3) years	–	0.28
7	Hornig et al., 2016	N = 18 Sex ratio (m:f) = 18:0 Age = n.a.	N = 34 Sex ratio (m:f) = 34:0 Age = n.a.	5.3 (±1.1) years	6.4 (±2.2) years	O	0.57
8	De Bosscher & De Rycke, 2017	N = 692[e] Sex ratio (m:f) = n.a. Age = 26.5 (±0.2) years	N = 745[e] Sex ratio (m:f) = n.a. Age = 22.1 (±0.2) years	9.7 (±5.1) years[f] (Level 2: 10.9 (±5.0) years)[g]	10.8 (±4.8) years	+	0.22
9[c]	Güllich, 2017	N = 83 Sex ratio (m:f) = 43:40 Age = 25.0 (±4.7)	N = 83 Sex ratio (m:f) = 43:40 Age = 24.2 (±4.4)	11.8 (±4.5) years	10.3 (±4.0) years	–	0.35
10	Güllich, 2018	N = 17 Sex ratio (m:f) = n.a. Age = n.a.	N = 17 Sex ratio (m:f) = n.a. Age = n.a.	9.9 (±4.0) years	8.8 (±3.1) years	O	0.30

Four of ten studies are to be described as studies with small sample sizes (one of the groups with n < 20) and these are exactly those reporting no sig. difference between I-Senior and N-Senior concerning the age of onset of IPMS. At this point, the importance of reporting effect sizes (as some of the studies do) becomes clear. For this reason, the effect sizes were calculated, if possible, according to an estimation function for hedges g (g*). In four cases, the effect strengths can be described as “small” and in one case as “medium” (Rasch, Friese, Hofmann, & Naumann, 2010).

The four studies using data from an identical survey (Studies 2, 4, 6 and 9 in Table VI) reported that I-Senior start IPMS at a significantly older age compared to N-Senior. However, De Bosscher and De Rycke (2017) came to adverse results. Interestingly, concerning the age at the onset of IPMS both groups of this study (Study 8 in Table VI) seem to vary from the four studies mentioned above. A possible explanation could be a different composition of samples regarding the sports categories. Although Güllich and Emrich (2014) report that no contrary findings were revealed in any sports category, a difference in the relative frequency of sports (categories) may have led to the different results in the studies. In order to avoid an influence due to a difference in the distribution of relative frequencies of sports categories between I-Senior and N-Senior, the sample has to be tested for homogeneity of the respective distributions. Surprisingly, none of the studies involving multi-sport samples reported such a test. This seems to be a matter of concern, not only for results in the context of comparing the age at onset of IPMS. Prospectively, studies should report on the frequency of sports, at least of sports categories, in both groups. Excepted from this problem for multi-sport studies is the study of Güllich (2017) because it applies a matching procedure with the type of sport being one of the matching variables. A further explanation might be found in the formation of success groups. De Bosscher and De Rycke (2017) used competitions such as Grand Slams, other studies did not. Furthermore, these authors applied a limit for N-Senior which might be more restrictive downwards compared to other studies.

Table VII shows the results for the (success-) relevance of V-IPMS and V-IPOS, respectively for ISS. Due to the restriction of the length of the article, we decided to report an aggregation of the results.

Table VII. (Success-) Relevance of V-IPMS and V-IPOS for ISS (referring to inference statistical results).

n.a.: no information available
Relevance for success:
+: sig. positive correlation (I-Senior train more compared to N-Senior)
–: sig. negative correlation (I-Senior train less compared to N-Senior)
O: no correlation
x/y: the majority of the results in this category correspond to x, but y was also found
x//y: x and y were determined the same number of times

No.	Source	V-IPMS		V-IPOS	
		childhood and adolescence	adulthood	childhood and adolescence	adulthood
1	Gibbons et al., 2002	n.a.	n.a.	n.a.	n.a.
2	Güllich & Emrich, 2006	–	n.a.	+	n.a.
3	Law et al., 2007	n.a./+	n.a.	n.a.	n.a.
4	Vaeyens et al., 2009	O	n.a.	+/O	n.a.
5	Güllich, 2014	O/–	O	O	O
6	Güllich & Emrich, 2014	O/–	O	+	+
7	Hornig et al., 2016	O	O//–	O/+	O//+
8	De Bosscher & De Rycke, 2017	n.a.	n.a.	n.a.	n.a.
9	Güllich, 2017	–	O/n.a.	+	O/n.a.
10	Güllich, 2018	O	O	+	n.a./O

Regarding the effect of V-IPMS on ISS in juvenile age categories/during childhood and adolescence (≤ 18 years), the results are inconsistent with studies finding no and negative effects. Interestingly, only one study found a positive effect of V-IPMS. However, it must be said that the sample of the study consists of rhythmic gymnastics, which means athletes of a sport which can be described as an "early specialization sport" (cf. De

Bosscher & De Rycke, 2017). For V-IPOS no study found a negative effect. Interestingly, six studies reported a positive effect of V-IPOS on ISS.

Analyzing the evaluation of V-IPMS and V-IPOS reveals that different variables have been used to measure the volume of training and test for differences. Differences are to be found not only in the way measurement was done (hours vs. numbers of sessions), but also in the way data was analyzed (expressed p.a., accumulated within age categories).

We are particularly interested in whether authors have used multivariate inference statistics to analyze the training history of athletes. As already mentioned, no data related to this aspect were collected except in one study; therefore, nine studies were analyzed in this context. Interestingly, only three studies applied multi-factorial (but univariate) methods: Two studies used the factors success level and sports category and then performed univariate analysis on V-IPMS and V-IPOS.[6] The third study used a univariate MANOVA (ANOVA with repeated measurement, whereas the two other studies used ANOVAs only; interestingly, with no use of differences between age categories) and thus stages in age and success levels as factors.

Two further additions should be made before summing up the findings. First, the study not reporting any data on the athletes' practice history (De Bosscher & De Rycke, 2017) introduced age as important covariate. Second, as already mentioned above, two of the studies applied a matching-procedure (variables taken together for both studies: gender, age, sports category, discipline, and performance at age 19 years) to control for confounding.

Literature review summary

Summing up, our literature review reveals that

- experts are defined in the literature both by their performance in specific aspects/skills and directly by their competitive success. In both cases this is a normative setting;
- several studies cited in other studies and/or reviews examining the relevance of V-IPMS (often interpreted in the broader context of DP) and V-IPOS for success (not seldom interpreted in the sense of ISS) have not compared I-Senior with N-Senior, but athletes at lower age levels or senior athletes across lower and/or more heterogeneous success levels;
- investigations on the most accomplished performers are still relatively scarce (Güllich & Emrich, 2014);
- although, in comparison to the existing literature, we used a rather restrictive definition for studies to be included in our review, our analysis reveals not only that studies have compared groups at different success levels, but also that there exists a "difference in the differences of groups";

- no study has differentiated between single success and multiple success (on the concentration of success and the resulting precariousness of this categorization cf. Barth, Güllich et al., 2018);
- too little attention has been paid to the description of N-Senior;
- the selection and description of groups should include at least three variables:[7]
  - the athlete's "age": age category (e.g., above the junior age limit of the sport's respective international competition regulations; senior/adult) and age level of most competitions on national and international level (e.g., internationally: youth; nationally: senior) in the last 12 months,
  - the athlete's "affiliation": age category (e.g., senior) and performance level (e.g., National Team, B-squad) at the time of the survey,
  - the athlete's (greatest) success at national and international level: age category (e.g., not restricted; senior), level (e.g., World Championships) and rank (e.g. 2nd place); international competitions results are to prefer, because the national level may vary between nations in sports
- studies are inconsistent in their findings regarding the relevance of V-IPMS and V-IPOS for ISS. On the one hand, a positive effect of V-IPMS or negative effect of V-IPOS in childhood and adolescence has not yet been reliably demonstrated. Although IPMS and DP may not be assessed as congruent (cf. problem and state of research), the findings obtained to date have called into question the prediction of the DP framework assuming that "early specialization"[8, 9] is overrepresented in I-Senior and that they have accumulated more V-IPMS (cf. Güllich, 2017). On the other hand, from seven studies reporting about the relevance of IPOS for ISS, six mentioned at least partially about a positive effect. However, the empirical verification of a positive effect of sampling (in structured practice) is – especially before the background of the shown weaknesses in the context of data analysis in the studies – still questionable.
- none of the studies involving multi-sport samples has reported testing on homogeneity of the relative frequency of sports or sports categories between the groups of I-Senior and N-Senior. This may not only confound previous studies' results[10], but also impede the comparability of the studies. Prospectively, studies should report the (relative) frequency of sports or at least sports categories in both groups;
- different variables have been used to represent V-IPMS;
- only three studies have applied multi-factorial statistical procedures to analyze the success-relevance of V-IPMS and V-IPOS; only one of those used a MANOVA;
- only two studies applied a matching-procedure to control for confounding;
- no study analyzing the success-relevance of V-IPMS and/or V-IPOS used age as covariate;
- no study used multivariate inference statistics to jointly analyze V-IPMS and V-IPOS[6];
- no study analyzed an interaction effect between V-IPMS and V-IPOS[6];
- with exception of the above-mentioned study analyzing data gathered in a longitudinal design (Güllich & Emrich, 2014), no study used hierarchical discriminant function analyses or regression analysis; However, this has been recommended by several authors (e.g., Coutinho et al., 2016; Ericsson, 2013, 2016);

- there exists a clear lack of empirically verified theories, and therefore research in this field has to be described as explorative.

Given the fact that most studies in the broader context of talent development in sports have used a relative approach, the results of our literature research seem alarming in several ways: first, the small sample size in some studies, which may be justified by the fact that the population of I-Senior is small per se. However, considering the power of inference statistical procedure, it seems to be a matter of concern since the tests applied will produce significant results only if the effect size is high enough (increasing beta error with smaller sample size). Second, only few studies have applied statistical procedures able to control for confounding variables; and third, no study analyzed a possible interaction effect of V-IPMS and V-IPOS.[6] Fourth, it must be generally said that the relative approach very rarely allows causal conclusions (Furley, Schul, & Memmert, 2016).

Therefore, the aim of this study is to analyze the relevance of V-IPMS and V-IPOS absolved in childhood and adolescence for ISS by applying a new methodical approach, combining decision trees and gradient boosting, to data from the early study by Emrich and Güllich (2005). This approach allows the realization of multivariate analysis and thus controlling for possible cofounders. Furthermore, the method used permits us to discover non-linear relationships and interaction effects (cf. method section). Therefore, hoping to achieve a relatively better explanation with this approach than with procedures applied in the past seems reasonable. To the best of our knowledge, the application of gradient boosting to talent development in sports has not been fully documented to date. For a first application of the Gradient Boosting Machine cf. Barth and Emrich (2018).

Methods – in search of an appropriate procedure

The statistical procedures applied in the analyzed studies have hardly fulfilled the recommended application of more complex methods for data analysis, such as multi-level modelling, structural equation modelling, or regression analysis (e.g., Coutinho et al., 2016; Ericsson, 2013, 2016). Furthermore, against the background of our literature review, showing a clear lack of empirically verified theories, research in this field has to be described as explorative. Therefore, the proposed statistical approaches, starting by assuming an appropriate data model and estimating parameters for this model from the data, seem to be less appropriate. To avoid starting with a data model and use general-purpose learning algorithms, instead, to learn about the relationship between the response and its predictors (Bzdok, Altman, & Krzywinski, 2018; Elith, Leathwick, & Hastie, 2008), approaches from machine learning seem to be more suitable.

Therefore, we decided to apply a method on existing data, which does not only allow carrying out a multivariate analysis of data and therefore controlling for a potential age-bias, but is also able to discover more complex/non-linear relationships. "Among the machine learning methods used in practice, gradient tree boosting is one technique that shines in many applications" (Chen & Guestrin, 2016, p. 785).

Boosted decision trees are an alternative approach to parametric modeling. A specific strength of boosting is that nonlinearities and interactions need not be explicitly specified beforehand, making boosted decision trees a flexible approach (Miller, McArtor, & Lubke, 2017; Schonlau, 2005). Although using these new approaches might require some reorientation in thinking, the clear evidence of their strong predictive performance and reliable identification of relevant variables and interactions (Elith et al., 2008) make them an extremely promising approach for the research problem at hand.

Data from Emrich and Güllich (2005) were reanalyzed by using an advancement of Friedman's Gradient Boosting Machine (GBM) (Friedman, 2001), called Extreme Gradient Boosting (XGBoost). Implementation was done by the freely available package XGBoost (version 0.71.2) under the R environment (version 3.5.1) (cf. Chen & Guestrin, 2016). Except from maximum depth of a tree (application of 4) default values were used. The number of rounds for boosting was 200. For documentation of XGBoost cf. Chen et al. (2018). This seems particularly relevant as four of the ten studies discussed in this review have used data from an identical project, which were also used in this work. For a description of the methods for data collection cf. Emrich and Güllich (2005) and Güllich and Emrich (2014).

By following a relative approach, we differentiated between I_s-Senior and N_s-Senior[11]; therefore, our task can be described as binary classification problem. We define I_s-Senior and N_s-Senior in accordance with Emrich and Güllich (2005): I_s-Senior are athletes having reached a 1st to 10th place at Olympic Games, Senior World Championships or European Championships.[12] After deleting, among others, the records of athletes not having exceeded the junior-age limit according to the international competition regularities of the respective sport, the number of participants in the dataset decreased from n = 1,566 to n = 595 athletes. The senior-age level sample includes 211 N_s-Senior and 484 I_s-Senior athletes. Concerning the success level of N_s-Senior, it can be said that around 75% of the athletes achieved a medal at senior national-level competitions and/or had (not successfully) participated in the mentioned international competitions. Therefore, describing this group as national-level senior athletes, in brief N_s-Senior, seems justified. The percentage distribution of N_s-Senior and I_s-Senior within the individual sports groups[13] does not differ significantly from one sports group to another ($\chi^2 = 7.33$; df = 4; p = .119; n = 595) The results are shown in Table VIII.

Table VIII. Distribution of N_s-Senior and I_s-Senior within different sports categories.

	Cgs-sports	Game sports	Martial arts sports	Artistic composition sports	Others
N_s-Senior	32.2%	37.6%	36.5%	52.2%	33.9%
I_s-Senior	67.8%	62.4%	63.5%	47.8%	66.1%

Furthermore, we tested for each group if it differs from all other groups taken together. The analysis reveals a sig. result only for artistic composition sports compared to all other groups taken together ($\chi^2 = 5.32$; df = 1; p = .021; n = 595).

An inspection of the athletes' mean age at the time of their interview shows that N_s-Senior (Mdn = 22.0; Iqr = 4.8; n = 210) were significantly younger compared to I_s-Senior (Mdn = 24.0; Iqr = 7.0; n = 384) (U = 28667, p < .01). This speaks in favor of the need to control for age. To classify the groups, we used the following predictors: age of athletes at the time of the survey, gender (dummy for male), categorization of sports (dummies), cumulative structured/institutionalized practice sessions of IPMS (= V-IPMS) as well as IPOS (= V-IPOS) in each age category during childhood and adolescence (up to 10 years, 11 to 14 years, and 15 to 18 years), age at onset of training in the athletes' main sports and in other sports, age at first admission to a national squad and to an Olympic Training Center (OTC). The volume of structured/institutionalized practice in a sport reflected the total accumulated volume, but did not differentiate the "micro-structure" within the practice sessions (e.g. proportions of technical skills exercises, physical conditioning, playing forms, stretching etc.).

The model's accuracy is assessed by means of a tenfold cross-validation and the subsequent averaging of the respective areas under the receiver operating characteristic curves (AUROC). Furthermore, the relative importance of the model's features as well as two-dimensional partial dependency plots (implementation was done by the "pdp" package version 0.6.0 (Greenwell, 2017) under the R environment) for four selected variables are presented. Last but not least, the interaction effect of IPMS and IPOS in the age category 15 to 18 years is investigated by using a three-dimensional partial dependency plot. [14] We acknowledge that the investigation of an interaction effect would require further analysis. Therefore, we want to emphasize that the results of this study in the mentioned context are a first step. The intention was more to demonstrate the potential of the procedures used for data analyses and less to investigate this question in detail.

Results

The mean AUROC is M = 0.60 (SD ± 0.06). Thus, the model hardly succeeds in making a prediction better than chance (0.50). The relative importance of the model's features is presented in Figure II.

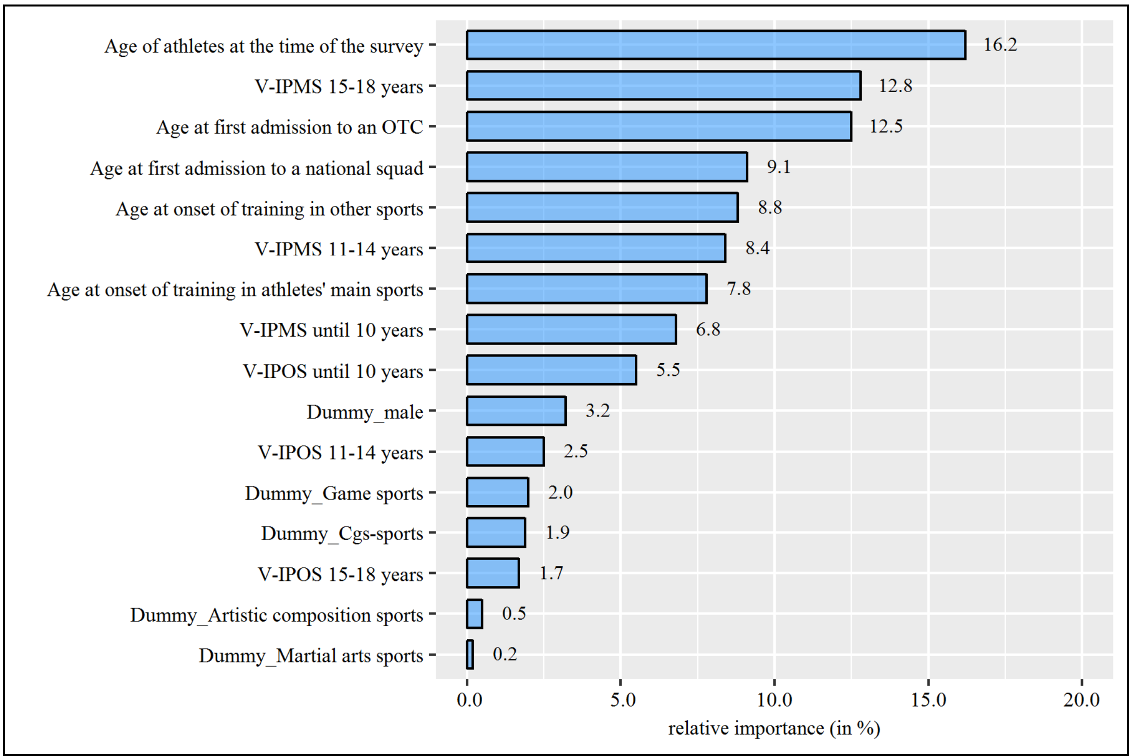


Figure II. Importance of features in the model.

Interestingly, age of athletes at the time of the survey is the model's most important feature. This clearly indicates the necessity of controlling for age when analyzing data; otherwise, results may be age-biased. The two-dimensional partial dependency plot of this feature (Figure III) shows the older the athlete at this time, the higher his probability to be classified as successful/I_s-Senior. There seems to be a weak indication for the existence of a negative correlation between the age at first admission to an OTC and the probability to be classified as successful/I_s-Senior. In particular, the findings for V-IPMS 15-18 years demonstrate the need to take non-linear relationships into account.

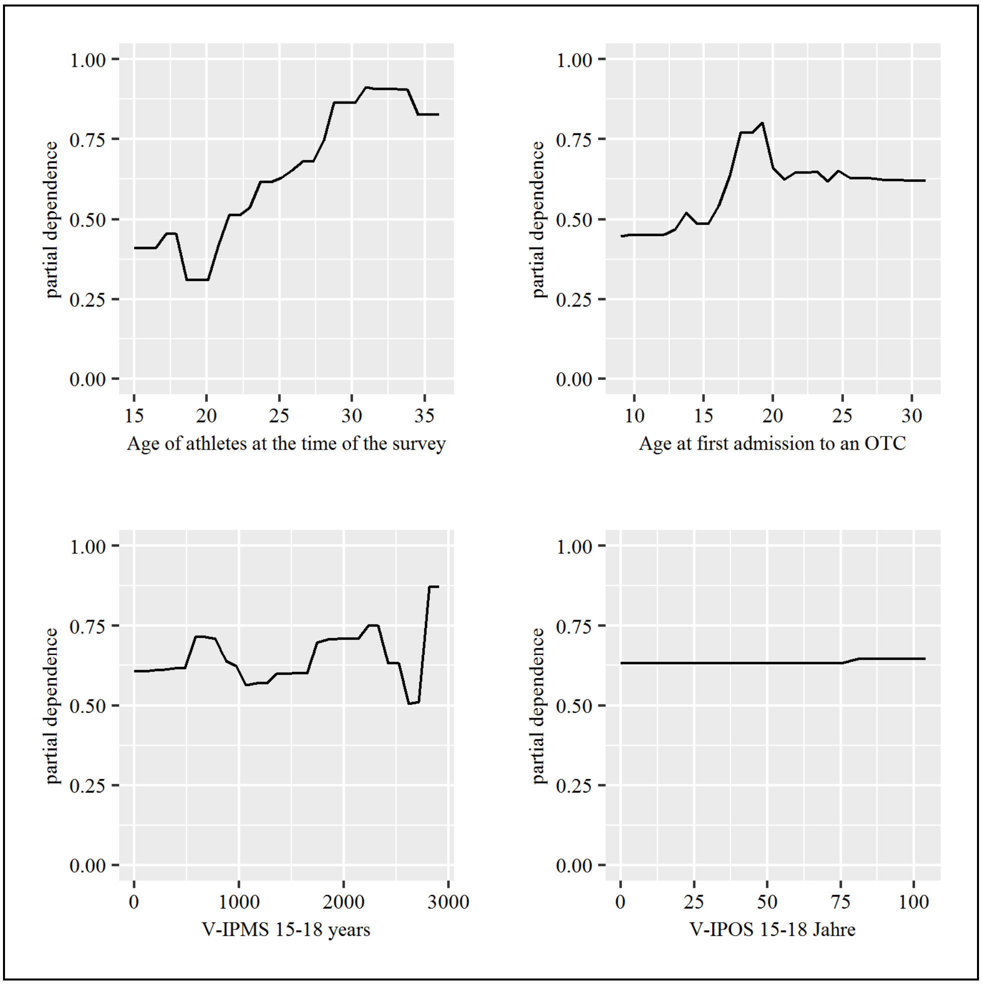


Figure III. Two-dimensional partial dependency plots.

Although the model's performance is very weak, a three-dimensional plot for investigating an interaction effect of V-IPMS 15-18 years and V-IPOS 15-18 years is presented in Figure IV. This will show the potential of the machine learning approach applied for data analysis in the context of talent development in sports.

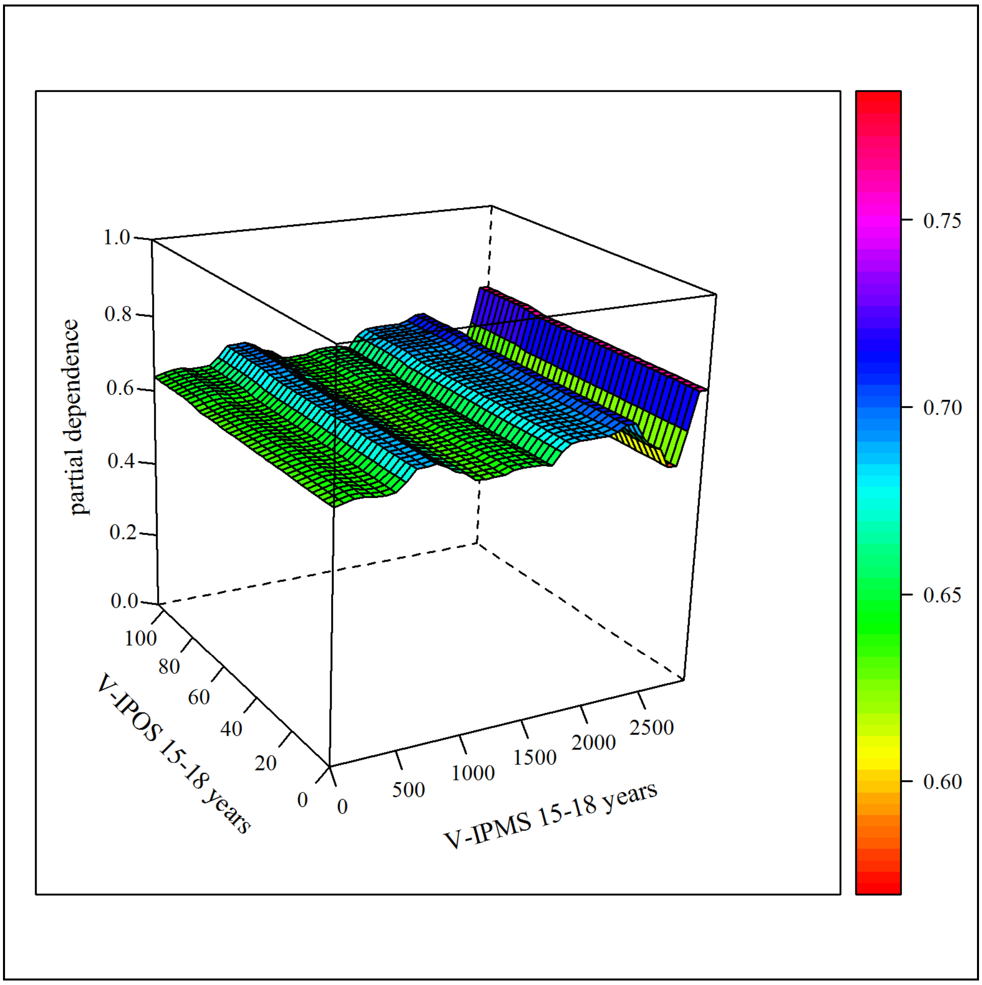


Figure IV. Three-dimensional partial dependency plot for investigating an interaction effect of V-IPMS 15-18 years and V-IPOS 15-18 years.

Keeping in mind the model's weakness, Figure IV shows that there does not seem to exist any interaction effect of V-IPMS 15-18 years and V-IPOS 15-18 years. This does not seem surprising, considering that V-IPOS in the age category 15 to 18 years has a relative importance of 2.4% in the (weak) model. V-IPOS 15-18 years seems to have no influence; this is also demonstrated by the flat course of the feature in the two-dimensional plot (cf. Figure III).

Discussion

This article addresses one of the central questions in connection with talent development in sports, namely how relevant V-IPMS and V-IPOS absolved in childhood and adolescence are for ISS, and reanalyzes an already existing data set with a more efficient procedure in order to answer this question.

As our review of literature reveals, there seems to exist a kind of terminological confusion at the output as well as the input side of the function for producing sporting success. This might increase the risk of extrapolating results or drawing unjustified conclusions in terms of the validity of theoretical models due to insufficient operationalization. The detailed analysis of literature in conjunction with our restrictive approach clearly shows that there is a lack of empirical studies questioning the role of the relevance of practicing in childhood and adolescence for sports talent development and achievement of ISS. Remarkably, out of the few studies only some applied statistical procedures able to control for confounding variables (e.g., age), and no study applied procedures such as hierarchical discriminant function analyses or regression analysis.[6] Furthermore, it has to be said that there exists a lack of empirically verified theories and therefore research in this field has to be described as explorative. These considerations demand not only for a multivariate approach for data analysis, but also for procedures not starting with assuming an appropriate data model. That is why we have adopted a machine learning approach.

Despite the application of the XGBoost – seemingly one of the most promising procedures in supervised machine learning models (Song, Chen, Deng, & Li, 2016) – it must be noted that on the basis of the features used in this study the model is hardly capable of making a correct classification and thus differentiation between I_s-Senior and N_s-Senior. Interestingly, the age of the athletes at the time the survey was conducted is the feature with the relatively highest influence.

The fact that age is the most important predictor in a model hardly able to correctly classify I_s-Senior and N_s-Senior seems to indicate that the significant differences between I_s-Senior and N_s-Senior in the context of V-IPMS (e.g., Güllich, 2014; Güllich & Emrich, 2006, 2014; Law et al., 2007) and V-IPOS (Güllich & Emrich, 2006, 2014; Hornig et al., 2016; Vaeyens et al., 2009) may represent rather artifacts of uncontrolled age effects than variables that differentiate the two groups. In the case of this study, the mean age of I-Senior was higher than that of N-Senior. Notes of this kind can also be found in other studies. The fact that in five of the ten studies included in the literature analysis the data were not sufficient to determine this seems to be extremely problematic. Furthermore, the literature review of this study showed that out of ten existing studies only one reports a positive effect of V-IPOS in childhood and adolescence for ISS. However, the analysis clearly showed that this study is problematic. The study's comparison groups consisted of only six athletes each. Furthermore, athletes of two different nations (cf. Güllich, 2018) were compared, which means that we have to consider a possible effect of differences in the nations' sport systems. Therefore, also the question whether Law et al. (2007) results that the successful athletes were just a bit over two years younger than the comparatively less successful athletes at the time of the survey is due to the respective sports or not must remain open at this point. Interestingly this study is cited in several studies without considering its methodological weaknesses.

However, the results should not be interpreted in a way that assumes training has no effect; rather, the empirical results of this study indicate that V-IPMS and/or V-IPOS absolved during childhood and adolescence are not variables discriminating I_s-Senior and N_s-Senior. In addition, the results illustrate the need to take non-linear relationships into account in modeling.

Interestingly and in contrast to the results of this study, Güllich (2017) found significant differences in V-IPMS and V-IPOS absolved during childhood and adolescence, when comparing successful and relatively less successful athletes with small to medium effect sizes for V-IPMS and medium to large effect sizes for V-IPOS. What makes these results so interesting is the fact that the author applied a matching-procedure (with age being one variable for the procedure) and therefore has already controlled for a possible age-bias.

A potential source of the divergent results might be that success was operationalized in different ways. Compared to Güllich's study (2017), we used a quite broad definition of success. In addition, the applied matching procedure reduced the sample size to n = 166, which makes it difficult to compare the two data sets – despite the fact that they originate from identical surveys. Furthermore, Güllich's (2017) data-analysis was restricted to a statistical comparison of groups, which means that neither potential non-linear effects nor interaction effects were taken into account.

Given this first-time investigation of a (non-linear) interaction effect, the importance of the results should not be exaggerated. We did not discover an interaction effect; however, it has to be said that only an effect within the same age category was investigated. There seems to be a weak indication for a negative correlation between the age at first admission to an OTC and the probability to be classified as successful/I_s-Senior.

This study does have limitations. First of all, it has to be said that the model's response variable is the athletes' individual success and not the success at the collective level of a nation. Due to the fact that the type and volume of scholastic physical education lessons and of DPL in leisure (cf. Côté et al., 2007) were not recorded in the data set, implementing these variables in the model was not possible. However, a positive effect of DPL on senior success has not been demonstrated reliably to date (Güllich & Emrich, 2014).

Further predictors should be considered in future projects. However, the intention of the present paper was not to create the best possible model, but to work on the question of how relevant juvenile engagements in IPMS and IPOS are for ISS, taking into account further (control) variables (robustness check). In addition, it has to be said that, in this first step, no further cross-validations were made to improve the parameters of the model.

In addition to possible (small) further model improvements by parameter optimization, the question arises as to whether a narrower version of "success" would engender better results in classification. A next step could be to compare this model with a model based on a "narrower version" of success, comparable to that of Güllich (2017), as a response variable. A further model could be calculated based on Güllich's (2017) data set, which was created by a matching-procedure (without using the matching variables as predictors). We should also apply different approaches to our classification problem, compare them and maybe combine them in an ensemble to possibly improve overall prediction (cf. Oppel et al., 2012). Perhaps, the procedure applied in this paper is not optimal (basically, for "no free lunch theorems for optimization" cf. Wolpert & Macready, 1997). Other projects seem to suggest that the relative performance of different procedures depends on the data at hand. Thinking in bigger steps, multifactorial research on "nature and nurture" as well as the incorporation of other factors such as contextual factors (e.g., socioeconomic factors) (Coutinho et al., 2016) seem to be promising; even more with new approaches from machine learning to analyzing data, making it possible to develop more interactive and dynamic models. Furthermore, we should not only rely on "habitual factors" but implement "situative factors" (e.g., competition anxiety immediately before the event). Moreover, further features enable us to better exploit the potential of machine learning methods. In our first implementation (Barth & Emrich, 2018) we used more variables. However, we find it more valuable to limit our variables to those in childhood and adolescence as we wanted to work on the question of whether these variables have the potential to make predictions.

Moreover, against the background of the significant age differences found in this study, the question of whether this approach of operationalization does actually compare experts and non-experts (relative approach) or rather experts and "not-yet-experts" arises. In general, the categorization in experts and non-experts does not seem to do justice to the multidimensional, bipolar structure of a relative model. Completing athletes' biographies by means of document analysis will not always be possible. In terms of the validity of the approach, however, this attempt seems desirable.

Conclusion

In the context of the specialization-diversification debate, the present results indicate that from today's perspective there is a debate about a "production function", the structure of which is still unknown to date. Even by implementing one of the seemingly most promising procedures in supervised machine learning (Song et al., 2016), we have not (yet) been able to discover a model with an acceptable detection rate.

Obviously, the institutionalized programs for talent development – in essence, expansion of the available training time and more intensive usage of the individual units of time (extensive and intensive time economy) (for economics of time in training see Emrich & Güllich, 2005; Fröhlich, Emrich, & Büch, 2007) – are an expression of highly rationalized myths rather than evidence-based efficient norms. The function of these myths for sports organizations thus seems to lie in the area of the legitimization function rather than the production function (for differentiation see, among others, Emrich & Güllich, 2005). Although, as stated above, we have to be careful in drawing conclusions on DP, the results seem to be in line with Hambrick, Burgoyne, Macnamara, and Ullén (2018, p. 287), stating that "key claims of the deliberate practice view are not well supported by the available evidence."

It became evident that with every methodological progress there must be a willingness in the sense of organized skepticism to fundamentally revise previously provisionally confirmed correlations. Research, and especially research in a field that is, as shown above, widely theory-less and therefore explorative, must therefore be aware that generally only the contemporary state of error is reported. It also became clear, however, that the production of sporting success through the production of high performance for the purpose of standing out in international competitions is not only highly uncertain in its results; also there is obviously no really reliable knowledge of the means that can be used to achieve these goals.

References

Abernethy, B., Farrow, D., & Berry, J. (2003). Constraints and issues in the development of a general theory of expert perceptual-motor performance: a critique of the deliberate practice framework. In J. L. Starkes & K. A. Ericsson (Eds.), *Expert performance in sports: advances in research on sport expertise* (pp. 349–369). Champaign, Ill. [i.a.]: Human Kinetics.

Baker, J. (2003). Early specialization in youth sport: a requirement for adult expertise? *High Ability Studies*, *14*, 85–94.

Baker, J., Cobley, S., & Fraser‐Thomas, J. (2009). What do we know about early sport specialization? Not much! *High Ability Studies*, *20*, 77–89.

Baker, J., Côte, J., & Abernethy, B. (2003). Sport-specific practice and the development of expert decision-making in team ball sports. *J Appl Sport Psychol*, *15*, 12–25.

Baker, J., Côté, J., & Deakin, J. (2005). Expertise in ultra-endurance triathletes early sport involvement, training structure, and the theory of deliberate practice. *J Appl Sport Psychol*, *17*, 64–78.

Baker, J., Côté, J., & Deakin, J. (2006). Patterns of early involvement in expert and nonexpert masters triathletes. *Res Q Exerc Sport*, *77*, 401–407.

Baker, J., & Young, B. (2014). 20 years later: deliberate practice and the development of expertise in sport. *Int Rev Sport Exerc Psychol*, *7*, 135–157.https://doi.org/10.1080/1750984X.2014.896024

Barth, M., & Emrich, E. (2018, May). *Spitzensportliche Erfolgsbedingungen revisited mit Boosted Regression Trees - Erfolgsbedingungen in einem ergebnisunsicheren Bereich.* Arbeitskreis Sportökonomie. 22. Jahrestagung des Arbeitskreises Sportökonomie, München, DE.

Barth, M., Emrich, E., & Daumann, F. (2018). Approaches and methods used for measuring organizational performance in national sport governing bodies from 1986 to 2014. A systematized review. *Current Issues in Sport Science*, *3*, e010. doi: 10.15203/CISS_2018.010

Barth, M., Güllich, A., & Emrich, E. (2018). The rich get richer and the poor get poorer – the Matthew mechanism as an approach to explain selection effects and the occurrence of multiple medalists in the "production" of international success in alpine ski racing. *Current Issues in Sport Science*, *3*, e008. doi: 10.15203/CISS_2018.008

Berry, J., Abernethy, B., & Côté, J. (2008). The contribution of structured activity and deliberate play to the development of expert perceptual and decision-making skill. *J Sport Exerc Psychol*, *30*, 685–708.

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus machine learning. *Nat Methods*, *15*, 233–234.

Carlson, R. (1990). Zur Sozialisation schwedischer Spitzenspieler im Tennis – eine empirische Untersuchung. In H. Gabler & H.-J. Mergner (Eds.), *Modelle zur Talentförderung im Tennis* (pp. 107–129). Ahrensburg, DE: Czwalina.

Chen, T., & Guestrin, C. (2016). XGBoost: a scalable tree boosting system. In B. Krishnapuram, M. Shah, A. Smola, C. Aggarwal, D. Shen, & R. Rastogi (Eds.), *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794). New York, NY: ACM.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., . . . XGBoost contributors (2018). *Package 'xgboost'*. Retrieved from https://cran.r-project.org/web/packages/xgboost/xgboost.pdf

Côté, J., Baker, J., & Abernethy, B. (2007). Practice and play in the development of sport expertise. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of sport psychology* (3rd ed., pp. 184–202). Hoboken, NJ: John Wileys & Sons.

Côté, J., & Erickson, K. (2015). Diversification and deliberate play during the sampling years. In J. Baker & D. Farrow (Eds.), *Routledge handbook of sport expertise* (pp. 305–316). London, UK/New York, NY: Routledge.

Côté, J., Baker, J., & Abernethy, B. (2003). From play to practice: a developmental framework for the acquisition of expertise in team sports. In J. L. Starkes & K. A. Ericsson (Eds.), *Expert performance in sports: advances in research on sport expertise* (pp. 89–113). Champaign, Ill. [i.a.]: Human Kinetics.

Couronné, R., Probst, P., & Boulesteix, A.-L. (2018). Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics*, *19*, 270.

Coutinho, P., Mesquita, I., & Fonseca, A. M. (2016). Talent development in sport: a critical review of pathways to expert performance. *Int J Sports Sci Coach*, *11*, 279–293.

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

Davids, K., Güllich, A., Shuttleworth, R., & Araújo, D. (2017). Understanding environmental and task constraints on talent development. In J. Baker, S. Cobley, J. Schorer, & N. Wattie (Eds.), *Routledge handbook of talent identification and development in sport* (pp. 192–206). London, UK/New York, NY: Routledge.

De Bosscher, V., Shibli, S., Westerbeek, H., & Van Bottenburg, M. (2015). *Successful elite sport policies: an international comparison of the sports policy factors leading to international sporting success (SPLISS 2.0) in 15 nations*. Maidenhead, GB: Meyer & Meyer Sport.

De Bosscher, V., & De Rycke, J. (2017). Talent development programmes: a retrospective analysis of the age and support services for talented athletes in 15 nations. *European Sport Management Quarterly*, *17*, 590–609.

Duffy, L. J., Baluch, B., & Ericsson, K. A. (2004). Dart performance as a function of facets of practice amongst professional and amateur men ana women players. *Int J Sport Psychol*, *35*, 232–245.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *J Anim Ecol*, *77*, 802–813.

Emrich, E., & Güllich, A. (2005). *Zur „Produktion" sportlichen Erfolgs: Organisationsstrukturen, Förderbedingungen und Planungsannahmen in kritischer Perspektive.* Köln, DE: Strauß.

Emrich, E., Pitsch, W., & Papathanassiou, V. (2001). *Die Sportvereine: Ein Versuch auf empirischer Grundlage*. Schorndorf, DE: Hofmann.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffmann (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 685–706). Cambridge, UK [i.a.]: Cambridge University Press.

Ericsson, K. A., Chase, W., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*, 1181–1182.

Ericsson, K. A. (2013). Training history, deliberate practice and elite sports performance: an analysis in response to Tucker and Collins review-what makes champions? *Br J Sports Med*, *47*, 533–535.

Ericsson, K. A. (2016). Summing up hours of any type of practice versus identifying optimal practice activities: commentary on Macnamara, Moreau, & Hambrick (2016). *Perspect Psychol Sci*, *11*, 351–354.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychol Rev*, *100*, 363–406.

Fédération Internationale de Football Association (2018). FIFA/Coca-Cola World Ranking: Belgium. Retrieved from https://www.fifa.com/fifa-world-ranking/associations/association=bel/men/index.html

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Ann Stat*, *29*, 1189–1232.

Fröhlich, M., Emrich, E., & Büch, M.-P. (2007). Grenzerträge auch im Sport! *German Journal of Exercise and Sport Research (formerly: Sportwissenschaft)*, *37*, 296–311.

Fröhlich, M., & Pieter, A. (2009). Cohen's Effektstärken als Mass der Bewertung von praktischer Relevanz – Implikationen für die Praxis. *Swiss Sports & Exercise Medicine (formerly: Schweizerische Zeitschrift für Sportmedizin und Sporttraumatologie)*, *57*, 139–142.

Furley, P., Schul, K., & Memmert, D. (2016). Positionspapier: Das Experten-Novizen-Paradigma und die Vertrauenskrise in der Psychologie. *Zeitschrift für Sportpsychologie*, *23*, 131–140.

Fussballdaten.de. (2018). WM 1994, Achtelfinale, Deutschland gegen Belgien. Retrieved from https://www.fussballdaten.de/wm/1994/achtelfinale/deutschland-belgien/

Gibbons, T., Hill, R., McConnell, A., Forster, T., & Moore, J. (2002). The path to excellence: a comprehensive view of development of U.S. Olympians who competed from 1984 – 1998. Retrieved from https://www.teamusa.org/-/media/TeamUSA/AthleteDevelopment/Coaching-Education/Path-to-Excellence_1984-1998.pdf?la=en&hash=D424D4415F42E7A216BFEF9D669BFF6FA464AB01

Greenwell, B. M. (2017). Pdp: an R package for constructing partial dependence plots. *R J*, *9*, 421–436. Retrieved from https://journal.r-project.org/archive/2017/RJ…/RJ-2017-016.pdf

Gulbin, J., Weissensteiner, J., Oldenziel, K., & Gagné, F. (2013). Patterns of performance development in elite athletes. *Eur J Sport Sci*, *13*, 605–614.

Güllich, A. (2014). Many roads lead to Rome: developmental paths to Olympic gold in men's field hockey. *Eur J Sport Sci*, *14*, 763–771.

Güllich, A. (2017). International medallists' and non-medallists' developmental sport activities - a matched-pairs analysis. *J Sports Sci*, *35*, 2281–2288.

Güllich, A. (2018). Sport-specific and non-specific practice of strong and weak responders in junior and senior elite athletics - a matched-pairs analysis. *J Sports Sci*, *36*, 2256–2264.

Güllich, A., & Cobley, S. (2017). On the efficacy of talent identification and talent development programmes. In J. Baker, S. Cobley, J. Schorer, & N. Wattie (Eds.), *Routledge handbook of talent identification and development in sport* (pp. 80–98). London, UK/New York, NY: Routledge.

Güllich, A., & Emrich, E. (2006). Evaluation of the support of young athletes in the elite sports system. *European Journal for Sport and Society*, *3*, 85–108.

Güllich, A., & Emrich, E. (2014). Considering long-term sustainability in the development of world class success. *Eur J Sport Sci*, *14 Suppl 1*, S383-397.

Hambrick, D. Z., Burgoyne, A. P., Macnamara, B. N., & Ullén, F. (2018). Toward a multifactorial model of expertise: beyond born versus made. *Ann N Y Acad Sci*, *1423*, 284–295.

Heinilä, K. (1982). The totalization process in international sports. *German Journal of Exercise and Sport Research (formerly: Sportwissenschaft)*, *12*, 235–254.

Helsen, W. F., Starkes, J. L., & Hodges, N. J. (1998). Team sports and the theory of deliberate practice. *J Sport Exerc Psychol*, *20*, 12–34.

Hendry, D. T., & Hodges, N. J. (2018). Early majority engagement pathway best defines transitions from youth to adult elite men's soccer in the UK: a three time-point retrospective and prospective study. *Psychol Sport Exerc*, *36*, 81–89.

Hodges, N. J., Kerr, T., Starkes, J. L., Weir, P. L., & Nananidou, A. (2004). Predicting performance times from deliberate practice hours for triathletes and swimmers: what, when, and where is practice important? *J Exp Psychol Appl*, *10*, 219–237.

Hodges, N. J., & Starkes, J. L. (1996). Wrestling with the nature expertise: a sport specific test of Ericsson, Krampe and Tesch-Römer's (1993) theory of "deliberate practice". *Int J Sport Psychol*, *27*, 400–424.

Hornig, M., Aust, F., & Güllich, A. (2016). Practice and play in the development of German top-level professional football players. *Eur J Sport Sci*, *16*, 96–105.

Houlihan, B., & Green, M. (2008). Comparative elite sport development. In B. Houlihan & M. Green (Eds.), *Comparative elite sport development: systems, structures and public policy* (pp. 1–25). Oxford, UK/Burlington, MA: Butterworth-Heinemann.

Huxley, D. J., O'Connor, D., & Larkin, P. (2017). The pathway to the top: key factors and influences in the development of Australian Olympic and World Championship track and field athletes. *Int J Sports Sci Coach*, *12*, 264–275.

Johnson, M. B. (2006). Climbing to the top: uncovering processes that facilitate and inhibit attaining excellence (Doctoral Thesis). Florida State University, Tallahassee.

Johnson, M. B., Tenenbaum, G., & Edmonds, W. A. (2006). Adaptation to physically and emotionally demanding conditions: the role of deliberate practice. *High Ability Studies*, *17*, 117–136.

Johnson, M. B., Tenenbaum, G., Edmonds, W. A., & Castillo, Y. (2008). A comparison of the developmental experiences of elite and sub-elite swimmers: similar developmental histories can lead to differences in performance level. *Sport Educ Soc*, *13*, 453–475.

Law, M. P., Côté, J., & Ericsson, K. A. (2007). Characteristics of expert development in rhythmic gymnastics: a retrospective study. *Int J Sport Exerc Psychol*, *5*, 82–103.

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*, e0196324. doi: 10.1371/journal.pone.0196324

Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: a meta-analysis. *Psychol Sci*, *25*, 1608–1618.

Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2018). Corrigendum: deliberate practice and performance in music, games, sports, education, and professions: a meta-analysis. *Psychol Sci*, *29*, 1202–1204.

Macnamara, B. N., Moreau, D., & Hambrick, D. Z. (2016). The relationship between deliberate practice and performance in sports: a meta-analysis. *Perspect Psychol Sci*, *11*, 333–350.

Memmert, D., Baker, J., & Bertsch, C. (2010). Play and practice in the development of sport‐specific creativity in team ball sports. High Ability Studies, 21, 3–18.

Miller, P. J., McArtor, D. B., & Lubke, G. H. (2017). A gradient boosting machine for hierarchically clustered data. *Multivariate Behav Res*, *52*, 117.

Moesch, K., Elbe, A.-M., Hauge, M.-L. T., & Wikman, J. M. (2011). Late specialization: the key to success in centimeters, grams, or seconds (cgs) sports. *Scand J Med Sci Sports*, *21*, e282-90. doi: 10.1111/j.1600-0838.2010.01280.x

Moesch, K., Hauge, M.-L. T., Wikman, J. M., & Elbe, A.-M. (2013). Making it to the top in team sports: start later, intensify, and de determined! *Talent Development & Excellence*, *5*, 85–100.

Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O'Connell, A. F., Miller, P. I., & Louzao, M. (2012). Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biol Conserv*, *156*, 94–104.

Rasch, B., Friese, M., Hofmann, W., & Naumann, E. (2010). *Quantitative Methoden: Band 1: Einführung in die Statistik für Psychologen und Sozialwissenschaftler* (3rd ed.). Heidelberg, DE: Springer.

Rees, T., Hardy, L., Güllich, A., Abernethy, B., Côté, J., Woodman, T., . . . Warr, C. (2016). The great British medalists project: a review of current knowledge on the development of the world's best sporting talent. *Sports Med*, *46*, 1041–1058.

Ronbeck, N. F., Dunnagan, T., & Stewart C. (2009). *Early specialization in elite nordic racers: fact or fiction* (Coachesinfo. Online not available).

Schonlau, M. (2005). Boosted regression (boosting): an introductory tutorial and a Stata plugin. *Stata J*, *5*, 330–354.

Sieghartsleitner, R., Zuber, C., Zibung, M., & Conzelmann, A. (2018). "The early specialised bird catches the worm!" - A specialised sampling model in the development of football talents. *Front Psychol*, *9*, e188. doi: 10.3389/fpsyg.2018.00188

Song, R., Chen, S., Deng, B., & Li, L. (2016). eXtreme gradient boosting for identifying individual users across different digital devices. In B. Cui, N. Zhang, J. Xu, X. Lian, & D. Liu (Eds.), *Web-age information management: 17th international conference, WAIM 2016: Nanchang, China, June 3–5, 2016: Proceedings, part I* (pp. 43–54). Cham, CH: Springer Switzerland.

Tucker, R., & Collins, M. (2012). What makes champions? A review of the relative contribution of genes and training to sporting success. *Br J Sports Med*, *46*, 555–561.

Vaeyens, R., Güllich, A., Warr, C. R., & Philippaerts, R. (2009). Talent identification and promotion programmes of Olympic athletes. *J Sports Sci*, *27*, 1367–1380.

Van Rossum, J. H. A. (2000). Deliberate practice and dutch field hockey: an addendum to Starkes. *Int J Sport Psychol*, 452–460.

Van Rossum, J. H. A. (2009). Giftedness and talent in sport. In L. V. Shavinina (Ed.), *International handbook of giftedness: part one* (pp. 751–792). Dordrecht, NL: Springer Netherlands.

Weissensteiner, J., Abernethy, B., Farrow, D., & Müller, S. (2008). The development of anticipation: a cross-sectional examination of the practice experiences contributing to skill in cricket batting. *J Sport Exerc Psychol*, *30*, 663–684. https://doi.org/10.1123/jsep.30.6.663

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Trans Evol Comput*, *1*, 67–82.

Notes

[1] Athletes were interviewed as part of a project funded by the BISp. The survey was approved by the BISp as well as by the Goethe University Frankfurt. The ethics committee of the Saarland University confirmed that it is not necessary for the re-analysis of existing and approved data to get new ethics approval.
[2] We exceptionally included one further article; for reasons cf. included articles.
[3] Again, we made an exception by adding the contribution of Gibbons et al. (2002) to the study's sample because this report is often cited. However, the inspection of the statistical procedures applied will clearly show the weakness of the reported results.
[4] Further studies can be found in Macnamara et al. (2016). However, most of these studies did not use success as response variable. We therefore forgo to list these studies.
[5] Cgs-sports: „Performance is measured in centimetres, grams or seconds (cgs). The task is to minimize time or to maximize distance or weight" (Güllich & Emrich, 2014, S387).
[6] It should be mentioned that Güllich and Emrich (2014) applied a mixed-longitudinal design (combination of a cross-sectional and longitudinal design). We referred to the results from data gathered in the cross-sectional design. However, in analyzing the data based on the 3-year longitudinal design, the authors applied several regression analyses with success at measuring repetition as response variable, first time measured success and different variables of main sport and other sports practice volume (e.i., IPMS and IPOS) as independent predictors.
[7] We are aware that possible "problems" may occur, for instance if an athlete's age lies below the junior-age limit, but the athlete has already achieved international success in senior championships. Furthermore, the selection variables depend on the design and method of the study, probably making some of them redundant.
[8] Therefore, we would expect that I-Senior accumulate a greater amount of IPMS in childhood and adolescence and start IPMS at a younger age.
[9] De Bosscher and De Rycke (2017) reported that I-Senior start IPMS at a significantly younger age compared to N-Senior. However, the authors added that the effect size is small (comparing three levels, ANOVA was applied with effect size $\eta^2_{part} = 0.012$).
[10] As mentioned above, Güllich (2017) used a matched-pairs design controlling for this source of confounding. Güllich and Emrich (2014) also carried out analyses separately for each sports category.
[11] We use "s" for sample because although the criteria are comparable to the inductively developed classification system above, they are not fully congruent.
[12] According to the authors' information, they also took "according to coaches comparable" competitions to European Championships into account. We followed their approach.
[13] For a categorization of types of sports cf. Emrich, Pitsch, and Papathanassiou (2001) or Güllich and Emrich (2014).
[14] Partial dependence plots visualize how each feature influences the prediction (here: the predicted probability of observation belonging to I_s-Senior) while averaging with respect to all the other features. This offers insight into any black box machine-learning model (Couronné, Probst, & Boulesteix, 2018).