Passing with the Stars:

Learning from Coworkers in Collaborative Jobs *

Gábor Békés[1] and Bence Szabó[2]

[1]Central European University, KRTK, CEPR

[2]Corvinus University of Budapest, KRTK

October 29, 2023

Abstract

How do workers learn from their peers in the workplace, especially when collaboration is essential? We investigate the mechanism of learning in a high-skill collaborative environment: elite male soccer. We start by confirming that players with peers of higher human capital (measured as their estimated transfer value potential) tend to experience higher future human capital. This is especially true in skewed environments – being together with stars is essential in learning. However, proximity is not enough, learning happens via interaction. We show that players with high exposure to peers – i.e. who play more minutes in games – will experience higher growth of their human capital. Once again being with stars is additionally helpful. Finally, we show that beyond minutes on the pitch interaction – passing – is where massive gains can come from. All results are especially important for younger players. Importantly, the learning channel is mitigated by a correlation between exposure and team quality: it is harder to get exposure in better teams. This implies that in teams with better peers, learning may be hindered by limited exposure. This suggests that estimates to date might underestimate learning by missing exposure.

Keywords: Learning, collaboration in teams, peer effects, stars, sports data

JEL-codes: E24, J31, O33, D83, Z20

*Corresponding author: Bence Szabó, bence.szabo@uni-corvinus.hu

1 Introduction

In this paper, we study human capital accumulation via learning in the workplace. We focus on a specific industry of high-skilled workers that will allow improved detection and understanding of both the process of learning as well as its impact on perceived human capital. Our industry of relevance will have a workplace where collaboration is essential. Examples could include most high-skilled services jobs such as academia, consulting, marketing, but also mid-level management,

Workplace learning is about transferable human capital accumulation (HCA) – acquisition of new skills and knowledge that can be used elsewhere, and thus, it creates value for the worker and the employer[1]. The literature on learning from co-workers relied mostly on large administrative data to gauge the impact of better peers in the establishment on wages (Cornelissen et al., 2017) (no effect) or wage growth (Jarosch et al., 2021a) (positive effect).

This paper contributes to the literature on human capital accumulation with an emphasis on learning from coworkers. We focus on an industry where collaboration is essential and show that exposure to peers is essential in human capital accumulation and that it also confounds results on peer effects.

Our paper will use a complex set of information from European footballers, we improve the estimation of learning from coworkers with a special focus on young talent. Football in the top European leagues created a very unified labor market, with players moving constantly between teams and countries. It is also a very high-skilled industry, with only the top players making it. This means that our empirical findings will be related to evidence in learning for the most skilled workers (in Jarosch et al. (2021a) this would be the top wage decile, where the learning coefficient is three times the average).

We build a new dataset tracking the career trajectory of six thousand players from the top five European football leagues over 8 seasons. Our data include 6,090 players over a total of 33,148 player*half-season observations. We focus on career trajectory over 3 years or 6 half-seasons. This data is combined with very fine, event-by-event data on 15 thousand football matches to establish the relationship and the intensity between player development and interaction with peers. Interaction is characterized by shared minutes and passing. These data enable the direct observation of the collaboration's intensity, providing new and substantial evidence to the literature.

Human capital accumulation (learning, or skill development) will be captured via players pre-

[1]In addition, there may be learning of workplace-specific information that will increase productivity but is not transferable.

dicted value on the transfer market. The market value aggregates the consensus view of experts with access to public information such as game footage (Poli et al., 2021). As an average view of people, the transfer value is an estimate of work-relevant human capital. In terms of this process, a player initially has some human capital (innate ability) to play football, which is affected by the team, by peers as he develops, resulting in a trajectory of the player's value[2].

We find that players experience higher growth in human capital, when (i) in the same team with better peers, (ii) in the same team with stars, (iii) exposed to peers (spend actual play time), (iv) interact with peers (passing), (v) interact with stars in particular. We will argue that in the estimation of peer effects omitting exposure might lead to substantially underestimating positive learning effects. All our results will be stronger for young players.

Beyond the focus on collaboration, we also contribute to the literature by documenting peer effects overcoming measurement problems. While there have been great advances in identification methodology following Abowd et al. (1999), this literature – and possibly mixed results – suffers from several measurement problems.

2 Related literature

The past and current workplace environments are important contributors to worker performance and compensation. Important related literature uses matched employer-employee data to estimate what drives wages using worker and firm fixed effect with help of worker switching following Abowd et al. (1999), Card et al. (2013). They reveal that a great deal of wage variation may be explained by sorting and firm characteristics. Econometrically looking at sorting is rather difficult due to limited switching (Bonhomme et al., 2023). Closer to us, is related stream of literature investigating the role of firms in explaining the variation in wage *growth* (Gregory, 2021). Finally, several papers looked at the heterogeneity of work environment on learning. Arellano-Bover and Saltiel (2023) studied the learning of young employees with administrative datasets from Brazil and Italy.

Overall, this literature suggests that firm characteristics, and heterogeneity across firms are key drivers of earnings at firms. There are several sources of why performance and earnings may be higher in some firms such as better management yielding better incentives, firm work culture, or in-house training programs.

[2]In semi-structured interviews, Werner and Dickson (2018) study peer learning in the Bundesliga and find that learning from others, knowledge sharing functions via four main channels: imitating, peer communication, labor mobility, and knowledge brokers.

The focus of this paper is another source of variation in wages: the impact of coworkers. Learning will come in many forms such as passive presence and observation of peers, interaction with them, and intensive collaboration.

In this literature the starting point is measuring the wage premium of better coworkers. Cornelissen et al. (2017) argue workers may alter their productivity due to both peer pressure and learning from co-workers. Learning in particular will be stronger from better peers and may thus lead to more inequality. While the evidence for learning is mixed (both lab and observational), there is more evidence for peer pressure (Mas and Moretti, 2009). They use German linked-employer-employee data from Munich covering 26 years and focus on wages rather than productivity. A worker's peer group is defined at the firm and three-digit occupation level, and special attention is paid to sectors with high knowledge (e.g. doctors). In their model, peers have a positive effect on wages either as compensation for more effort or via generating more output (learning). In terms of estimating peer effects, Cornelissen et al. (2017) has a firm-level point estimate with firm fixed effects of 14.8% that falls to 6.6% once firm-occupation fixed effects are added and 1.1% with additional firm-year fixed effects. Occupation and firm-level shocks are indeed pretty important. However, for repetitive tasks, there is actually a 6% compared with basically zero for the most knowledge-intensive ones. This suggests learning is less important, while peer pressure matters.

The closest paper to ours is Jarosch et al. (2021a) which also investigates learning at the workplace with German administrative data but focuses on wage *growth* rather than levels as in Cornelissen et al. (2017). They estimate a model where wage growth between time t and $t + h$ is a function of the quality of peers, and worker characteristics at time t. They find that in firms (firm-occupations) with 100% higher average wage of peers is associated with 7% higher wage *growth* within one year, rising over time, with cumulative growth reaching 21% over ten years. Furthermore, they show that having more highly paid coworkers is strongly associated with future wage growth. Also, unlike for wage levels, learning is there for top performers.

There is less evidence of what aspect of peers help HCA. Education is one: Nix (2020) considers the role of learning important as it manifests as an externality to workers. Using detailed Swedish data, she finds that increasing the average education of coworkers by 10 percentage points increases that worker's wages in the following year by 0.3%.

Beyond human capital accumulation, there could be peer effects. First, better peers may help you do your job better. Arcidiacono et al. (2017) estimates the contribution of peers to perform better on the pitch. Unlike learning or peer pressure, this productivity spillover is related to actual assistance in work. In this paper, the identification of peer effects comes from observing a large variation in peers (of different positions) across games. Second, peer effects can manifest themselves as social pressure, and may be even negative. Peer pressure was estimated to improve repetitive task performance in supermarkets (Mas and Moretti, 2009). Indeed a single relevant peer may have a negative effect: exposure to superstars might actually be detrimental for self-confidence. For instance, in competitive environments, the higher pressure may lead to more mistakes (Bilen and Matros, 2021).

The learning aspect of jobs is especially relevant in the first part of professional careers. Early career exposure to an established and experienced mentoring figure, be it a team member, a coworker, or a relative, might contribute substantially to the development of young professionals. Evidence to date on this relationship in detail in several domains, with mixed results. [3]. A closely related paper, also using sports data is Hoey (2023) which shows that the injury of stars helps young players play more, and thus helps talent discovery.

3 Data

We base our investigation on a recently assembled dataset (Békés and Ottaviano, 2022) that was web scraped from different sources including transfermarkt.com and whoscored.com. The dataset covers each event (each pass, tackle, etc.) from all matches, along with team composition and results of games in the top five men's football leagues (Premier League in England, Ligue 1 in France, Bundesliga in Germany, Serie A in Italy, La Liga in Spain) over eight sporting seasons (2011-12 to 2018-19) [4]. Additionally, we have information on players' histories, their activity in teams, and their valuation, together with information on key individual and team-level characteristics.[5]

The key datasets are:

- Player careers: players and team in every half-season

[3] There is evidence that young-age exposure to experienced and impactful figures might lead to higher future success, or present spillovers in the context of industrial innovation (Bell et al., 2017), the fine arts (Fraiberger et al., 2018), research (Li et al., 2019). Superstars especially drive the differences across firms in many industries (Jarosch et al., 2021a), including football teams (Hoegele et al., 2014)

[4] Data quality and coverage are both very high in our datasets. Nevertheless, small data cleaning steps were needed and we discuss these issues in Appendix (A.3)

[5] With 'European football', or simply 'football' henceforth, we refer to 'association football'. The Premier League, La Liga, Serie A, and Ligue 1 are all composed of 20 teams (playing $20\times19=380$ games), while there are 18 teams ($18\times17=306$ games) in the Bundesliga. In any given season, there are 98 teams in our sample, and we have 98x16=1568 team by half-season units in our dataset. Due to relegation and promotion, we have a total of 154 teams in the sample. Overall, our dataset covers a total of $8\times(380\times4+306)=14{,}608$ games.

- Player valuations: Valuation for a player in every half-season. (Average when multiple values are available.)
- Players and squads: information of squads for each team in every half-season (list of players who were ever listed in the lineup as starter or substitute.)
- Events: Minutes spent together and number of passes between any two players

For every player, we follow the evolution of market value from potentially as early as playing in Under-19 youth teams of football clubs. We retrieve market values from transfermarkt.com, the site's methodology is explained in detail here. Their concept of market value differs from transfer fees as it targets the *expected value of a player in a free market*, ideal for our setup to study the evolution of human capital. Rather than building a statistical model only, they rely on augmenting player pricing models with the judgment of the website's community, taking into account a list of factors such as future prospects, peer groups, injuries, contractual and club-specific conditions (an exhaustive list can be found in Appendix A.5), more suited for the evaluation of complex and particular situations. Therefore their market value calculation combines several tangible and intangible aspects of a football player's market value, closely resembling how economists view human capital. It also has the additional advantage compared to transfer fees that we do not need to consider the issue of business practices in large vs. smaller leagues, legalities, tax and accounting considerations, and other distorting factors. Transfermarkt publishes market values twice in a season, usually after the end of a season, and once during a season after sufficiently many games were played. In our dataset, accordingly, we aggregate player market values to the season/half-season level, reflecting the updates in market value.

We also track the work interaction between players during all matches in terms of completed passes, pass combinations, and shared minutes on the pitch, enabling us to quantify and finely measure the nature of collaboration in their working environment. We then aggregate these measures to the season/half-season level to harmonize with the market value tracking, creating a dynamic directed network of pass interaction between teammates in a half-season. A half-season contains typically 16-20 games in either the Summer/Fall (August to December) or Winter/Spring (January to May). Collaboration is measured as the number of minutes and passes in games during regular season[6].

[6] There is evidence that players who collaborate more in training and drills will be played alongside and collaborate more on the pitch.

As a third pillar, we build on the combined team-level, temporal, and playing-position level distribution of player values to create individual indicators and team-level variables. First, we use the lineup information and combine it with the market values to produce the team's total market value, Herfindahl–Hirschman index, and related measures. Second, we identify star players as those that have been in the top 5% and 25% of the market value distribution in the relevant top 5 leagues conditional on the season/season-half and their respective playing position, in any period of the last two years. This segmentation of the players allows us to consider different types of interactions between the elite players, very good players, and regular players (keeping in mind that players in these football leagues are incredibly talented individuals in their respective fields).

Descriptive statistics are presented in Section A.2 of the Appendix.

A major advantage of our dataset over several usual administrative datasets with linked employer-employee information is that we can differentiate between simply having different types of coworkers in terms of productivity in close proximity or on our team vs. actively collaborating with them during work, which we can measure using the passes and the shared minutes on the pitch. Furthermore, the usual occupation categorizations are often overly general regarding the actual tasks employees have to perform and therefore require several steps to infer whether learning could actually take place meaningfully. In our setup the work activity and the during-work interactions are directly related to the human capital accumulation of the employees. This establishes a solid connection between learning and interpersonal interactions at the workplace.

Beyond this aspect of collaboration, we (i) measure human capital accumulation rather than wage, (ii) measure peer effects in a meaningful team unit of fixed size, (iii) look at peers who might indeed provide source of learning. For more on this, see Section A.4 in the Appendix.

4 Empirical setup

To estimate peer effects, we regress the future value of HC (HC_{t+h}) measured by $ln(value)$ as the function of its current level (HC_t) and the team quality ($HCteam_t$)[7]. The team value excludes the actual player. Control variables are in X_t. All right-hand side variables are measured at period (half-season) t, while future wage is measured $t+h$, h periods later.

As for the key variables, the **Market value** (Transfermarkt): mean market value of the player in the half-season, without evaluation the last valid evaluation is assigned to the player. The

[7]As values are in logs, it is equivalent to having the change in the value as the dependent variable (with the coefficient of current HC changing only)

Teammates' total market value: the sum of the market value of the teammates' of the player based on their mean market value in the half-season. Section A.1 in the Appendix defines the key variables of the analysis in detail.

This is a pooled cross-section OLS regression, standard errors are clustered at the league*half-season*position level. Cross-section in the sense that variation comes from different players across different teams, and not within teams. The reason for this is that while there is churning, team value is correlated over our short time span of 8 seasons.

$$\mathbb{E}[HC_{t+h}|.] = \alpha HC_t + \beta HCteam_t + \gamma X_t$$

As a basic concept, our human capital measure shall capture all relevant information. However, we will use some corrections mostly to correct for less-than-perfect matching between human capital and measured valuation. While they can be matched to Mincerian controls in Jarosch et al. (2021a), we also have industry-specific reasons.

First, valuations are drawn up bearing in mind that valuations reflect value for the team beyond sharing talent. Thus the best defender will have about half the value of the best forward because his skills can be translated less directly to goals. Thus, we added position dummies.

Second, player valuation will include some risks of injury. Older players are more likely to get injured, and very young players may have unknown injury potential. Thus, we add age as a control variable, measured in a non-parametric way (years of age dummies).

Third, contracting issues may also play a role. For now, they are captured by age, but we'll add some more.

Fourth, valuations of players may depend on leagues and seasons, because labor markets are closely but not perfectly integrated. Value inflation, possibly in a country-specific way could also matter. We add league $\times$ half-season dummies to all models.

5 Results

5.1 Establishing evidence on learning

We start by presenting our findings on learning from peers. Table 1 shows the core evidence on learning for Equation 4. This is basically very close to what was estimated in Jarosch et al. (2021a) but has a finer measure of human capital as the dependent variable and time is measured in half-seasons (half years), not years: 6 periods is in fact 3 years.

Table 1: Learning over time and peer quality: full sample

Dependent Variables:	h=1	h=2	h=3	h=4	h=5	h=6
			log(value$_{t+h}$)			
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
log(value)	0.916***	0.898***	0.857***	0.839***	0.810***	0.801***
	(0.006)	(0.006)	(0.007)	(0.008)	(0.009)	(0.009)
log(teammates' total value)	0.049***	0.061***	0.099***	0.106***	0.136***	0.129***
	(0.006)	(0.007)	(0.010)	(0.011)	(0.012)	(0.013)
Fixed-effects						
position	Yes	Yes	Yes	Yes	Yes	Yes
player age	Yes	Yes	Yes	Yes	Yes	Yes
league X half-season	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	32,965	32,763	32,488	30,097	27,698	25,299
R^2	0.917	0.867	0.813	0.769	0.729	0.697
Within R^2	0.885	0.814	0.736	0.671	0.612	0.564

Note: All players, all teams. Standard errors clustered at position × league × half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

As the coefficients on peers suggest, the results are rather close to what was shown in German administrative data. Conditioning on their current HC value, and partialing out key confounders, workers with better peers will develop HC more. The point estimate is growing steadily with time reaching 11% by half season 6 (i.e. over 3 years).

Given that for our main analysis we will look at the interaction between continuous variables, instead of million euros we will look at their standardized values (Z-score), in Table 2. In this setup, the point estimates are actually pretty close, one standard deviation is close to a log unit.

Table 2: Learning over time and peer quality: full sample, Z-score

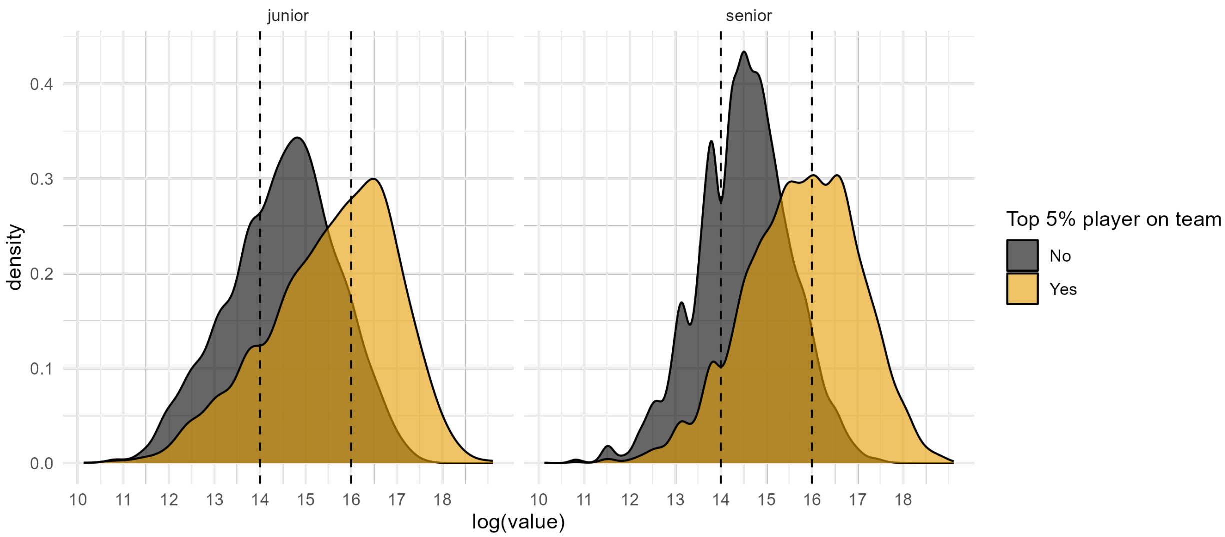
Dependent Variables:	h=1	h=2	h=3	h=4	h=5	h=6
	log($value_{t+h}$)					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
log(value)	0.921***	0.905***	0.871***	0.853***	0.826***	0.813***
	(0.005)	(0.006)	(0.007)	(0.007)	(0.008)	(0.009)
Teammates' total value (Z)	0.037***	0.044***	0.067***	0.074***	0.104***	0.110***
	(0.005)	(0.006)	(0.008)	(0.009)	(0.012)	(0.012)
Fixed-effects						
position	Yes	Yes	Yes	Yes	Yes	Yes
player age	Yes	Yes	Yes	Yes	Yes	Yes
league X half-season	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	32,965	32,763	32,488	30,097	27,698	25,299
R^2	0.917	0.867	0.812	0.769	0.729	0.697
Within R^2	0.885	0.813	0.735	0.670	0.611	0.564

Note: All players, all teams. Standard errors clustered at position × league × half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

From now on, all regression tables will present Z-scores, and interpretations will be based on comparing one SD difference in peer quality.

The core comparison in this project is players who have the same amount of human capital but vary in terms of peer quality. One key econometric problem here (and possibly in other datasets), is the lack of common support at either end of the player distribution. There are very few low-valued players at top teams, and very few very high-valued players at bottom teams. We help create a common support by focusing on the middle section of the distribution excluding players below 1 million and 9 million. Figure 1 displays the densities of log(market values) by seniority and whether the team has a top 5% player or not. We can see that for both age-groups there is a fair share of players between the two cut-off values (dashed vertical lines on the plot) in both types of teams.

Figure 1: Common support cutoffs



Note: Market value densities by types of teams in the sample.

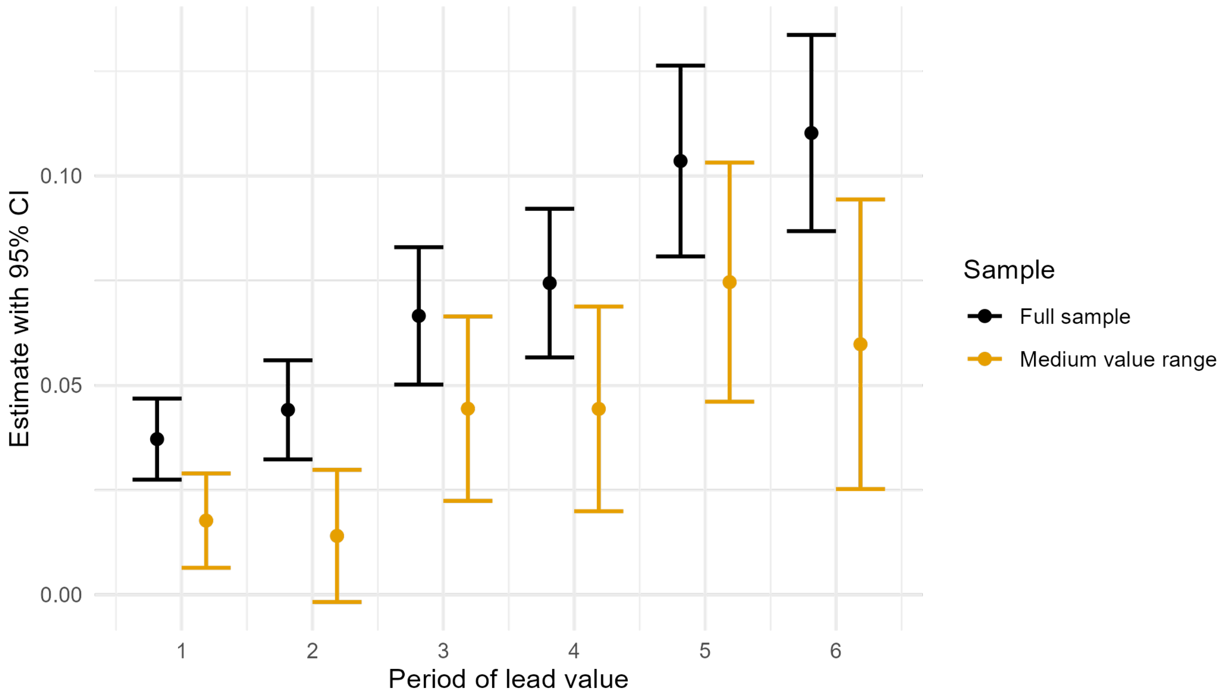
Table 3 and Figure 2 repeat the previous exercise but now, for the subsample of medium-value players. Results are similar qualitatively but smaller in magnitude.

Table 3: Learning over time and peer quality: medium value range players

Dependent Variables:	h=1	h=2	h=3	h=4	h=5	h=6
			$\log(\text{value}_{t+h})$			
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
log(value)	0.956***	0.956***	0.934***	0.926***	0.907***	0.899***
	(0.007)	(0.009)	(0.010)	(0.012)	(0.014)	(0.015)
Teammates' total value (Z)	0.018***	0.014*	0.044***	0.044***	0.075***	0.060***
	(0.006)	(0.008)	(0.011)	(0.012)	(0.015)	(0.018)
Fixed-effects						
position	Yes	Yes	Yes	Yes	Yes	Yes
player age	Yes	Yes	Yes	Yes	Yes	Yes
league X half-season	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	17,907	17,858	17,792	16,674	15,582	14,423
R^2	0.718	0.611	0.544	0.510	0.498	0.492
Within R^2	0.671	0.519	0.408	0.336	0.285	0.246

Note: Players in the medium market value range, all teams. Standard errors clustered at position $\times$ league $\times$ half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 2: Human capital gain over time: full sample vs. medium value players



Note: Association of market values with the team's total value evolving over the career, in the entire sample and the medium value range.

After the core results, let us turn to better understanding the mechanisms of learning. We will first look at the heterogeneity of learning by worker age. Then, we dissect the peer group into stars, high-flyers, and the rest. While we'll look at variation across these players, bear in mind that despite this dissection, we speak about the top 2% of global football players who broke into this elite[8]. This means, that finding in terms of this variation has limited external validity for this industry as a whole.

5.2 Heterogeneity 1: Workers

In this section, we look at who benefits more or less from their peers. We repeat the exercise for the young players defined as age below 23. Learning should evidently matter more for the young, as on average, they are farther away from the frontier and more willing to learn and adapt.

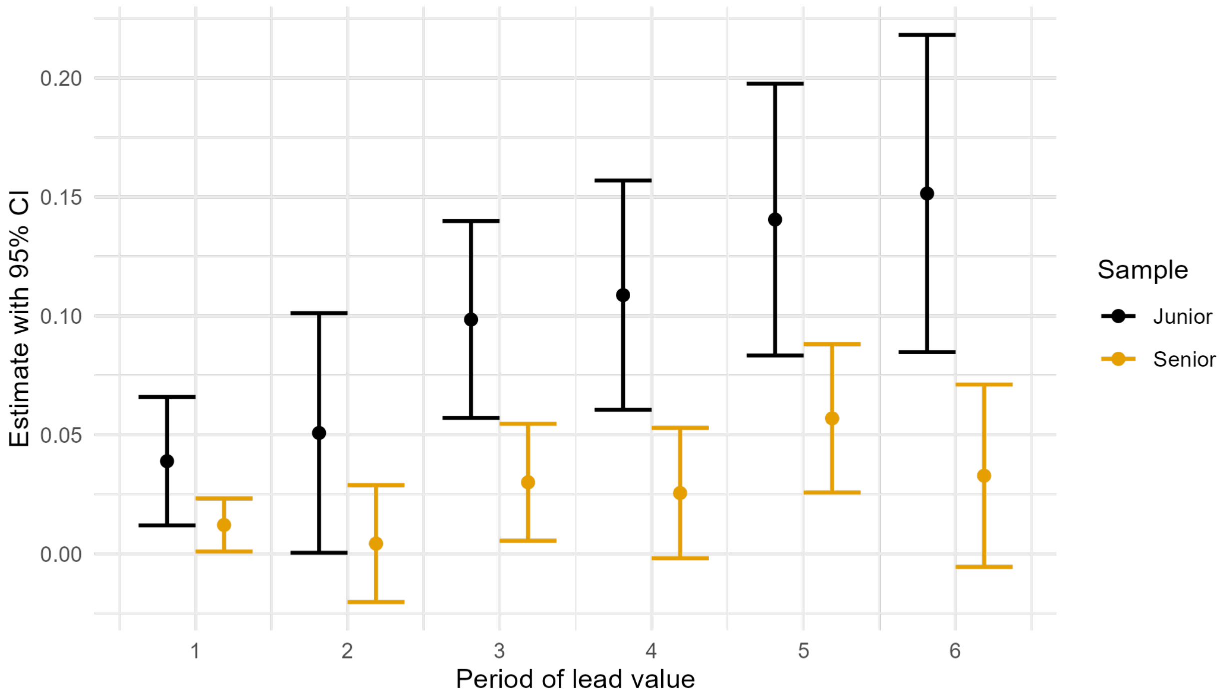
[8] 98 teams in top 5, appr 26 men squads for 130,000 professionals globally. https://www.statista.com/statistics/1283927/number-pro-soccer-players-by-country/

Table 4: Learning over time and peer quality: junior players

Dependent Variables:	h=1	h=2	h=3	h=4	h=5	h=6
			$\log(\text{value}_{t+h})$			
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
log(value)	0.948***	0.947***	0.957***	0.949***	0.946***	0.931***
	(0.015)	(0.020)	(0.024)	(0.030)	(0.036)	(0.038)
Teammates' total value (Z)	0.039***	0.078***	0.098***	0.109***	0.140***	0.151***
	(0.014)	(0.017)	(0.021)	(0.025)	(0.029)	(0.034)
Fixed-effects						
position	Yes	Yes	Yes	Yes	Yes	Yes
player age	Yes	Yes	Yes	Yes	Yes	Yes
league X half-season	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	3,511	3,511	3,511	3,325	3,132	2,924
R^2	0.621	0.485	0.410	0.347	0.306	0.274
Within R^2	0.589	0.437	0.352	0.285	0.243	0.211

Note: Junior players in the medium market value range, all teams. Standard errors clustered at position × league × half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 3: Human capital gain over time: juniors vs. seniors



Note: Association of market values with the team's total value evolving over the career, in the medium value range seniors vs. juniors.

Table 4 and Figure 3 show that young players learn substantially more from peers, one standard deviation of peer value translating into around 10% of market value after six half-seasons.

5.3 Heterogeneity 2: Peers

In this section, we look at what kind of peers help the most. This question relates to the distribution of peers given average quality. One way to gauge it is to add the **Hirschman–Herfindahl Index** (HHI) of the market values. Higher value implies more concentration, i.e. presence of star players – relative to the team average. Another approach is to look for absolute stars. Here we consider superstars, the top 5% of all players, and high-flyers, those in the top 25% but not in 5%.

In Table 5 we show the results, first for all the players (in the middle-value section) followed by young players only.

The point estimate of the HHI index is positive in the full and non-significant in the young player sample: higher HC growth is positively associated with a more skewed peer group for the entire sample (Column 2), but not for the younger players (Column 5). However, the presence of top players contributes to everyone's and to the young players' development (Columns 3 and 6).

Players with a 10pp higher share of superstars will on average experience a 2% higher market value growth.

For junior players, the size is higher, and the coefficient estimate for being with high flyers also has a large point estimate (Column 6). This suggests that beyond the HC accumulation benefits of having better peers in general, we see that being in proximity of stars is additionally helpful, especially for junior players.

Table 5: Learning over time and peer quality: Playing with stars

Dependent Variable:	log(value$_{t+6}$)					
		Full sample			Junior	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
log(value)	0.899***	0.897***	0.898***	0.931***	0.931***	0.916***
	(0.015)	(0.015)	(0.015)	(0.038)	(0.038)	(0.039)
Teammates' total value (Z)	0.060***	0.053***		0.151***	0.151***	
	(0.018)	(0.018)		(0.034)	(0.035)	
Team HHI		1.50**			0.161	
		(0.616)			(1.80)	
Share of top 5% teammates			0.206***			0.543***
			(0.075)			(0.137)
Share of top 25% teammates			0.064			0.363***
			(0.058)			(0.138)
Fixed-effects						
position	Yes	Yes	Yes	Yes	Yes	Yes
player age	Yes	Yes	Yes	Yes	Yes	Yes
league X half-season	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	14,423	14,423	14,423	2,924	2,924	2,924
R^2	0.492	0.492	0.492	0.274	0.274	0.275
Within R^2	0.246	0.247	0.246	0.211	0.211	0.213

Note: Columns (1) to (3), all medium market value range players, Columns (4) to (6) only young players. Teammates' values are measured as the Z-score. HHI is the team level Hirschman–Herfindahl Index. Standard errors clustered at position × league × half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

To recap, we have confirmed Jarosch et al. (2021a) for a super high-skilled segment, with well-identified peer groups and having human capital not wage. We also confirmed an asymmetry in peers, a skewed distribution with stars is indeed helpful. Young players can learn better from peers, especially stars. Next, we'll look more at the mechanisms.

5.4 Mechanims 1: Exposure

Workers will benefit from proximity to peers in general, and stars in particular by being actively engaged with them. Players who have more playing time in games are likely to have also trained more with the best peers.

An important addition is that exposure to peers here (and in any workplace) is also close to the concept of learning by doing. We capture this aspect with two variables: first, we measure exposure by the number of in-match minutes (standardized), and how important of a role the player plays in the team's pass system measured by the eigenvector centrality of the player (also standardized). The latter is calculated based on the network of directed passes aggregated to half-seasons and player-couples.

Our first measure is minutes spent on the pitch during the half-season. As Table 6 Column (2) shows players who spend 0.1 sd unit more time with their peers are expected to experience 1.4% more HC accumulation.

Second, note that as we added minutes, the point estimate of the teammates' total value went up from 0.06 to 0.22. So in peer effect regressions, exposure (activity) measured here with minutes played is actually a confounder. It is *negatively* correlated with team value (conditional on controls). In our case, this means players who play in higher-valued teams tend to play less. Indeed, it is harder to make it into the team when it's filled with stars and high flyers.

Furthermore, we added a measure of how central players are to the team's functions (Column 2). The pass centrality captures if the player is not only on the pitch but is actively engaged with peers. Thus, this is another, more sophisticated metric of exposure. We find (Columns 2, 5) that higher centrality is also positively associated with higher HC accumulation. Players who spend more time on the pitch or even highly engaged with peers will experience a higher HC accumulation.

Table 6: Learning over time and minutes played: medium market value range players

Dependent Variable:	log(value$_{t+6}$)					
	Full sample			Junior		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
log(value)	0.899***	0.708***	0.709***	0.931***	0.692***	0.682***
	(0.015)	(0.017)	(0.017)	(0.038)	(0.040)	(0.040)
Teammates' total value (Z)	0.060***	0.224***	0.219***	0.151***	0.331***	0.420***
	(0.018)	(0.019)	(0.020)	(0.034)	(0.041)	(0.049)
Total minutes (Z)		0.144***	0.141***		0.160***	0.188***
		(0.014)	(0.014)		(0.041)	(0.043)
Pass eigenvector centrality (Z)		0.154***	0.152***		0.237***	0.250***
		(0.012)	(0.013)		(0.038)	(0.038)
Teammates' total value (Z) × Total minutes (Z)			-0.012			0.108**
			(0.013)			(0.043)
Fixed-effects						
position	Yes	Yes	Yes	Yes	Yes	Yes
player age	Yes	Yes	Yes	Yes	Yes	Yes
league X half-season	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	14,423	14,169	14,169	2,924	2,861	2,861
R^2	0.492	0.534	0.534	0.274	0.343	0.345
Within R^2	0.246	0.305	0.305	0.211	0.284	0.286

Note: Players in the medium market value range, all teams. Standard errors clustered at position × league × half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In Columns (4) to (6) we look at young players, where we see the same pattern, point estimates are similar or higher. But we see an additional result in Column (6). Higher peer value is magnified by more exposure (as shown by the interaction term). Young players learn more from better players when they play more, while seniors do not.

5.5 Mechanism 2: Passing with stars

It's possible that playing time is not equally shared with peers. Learning when playing with stars vs. rest is not the same. Next, we add passing intensity and shared minutes on the pitch with stars (top 5%) and high-flyers (top 25%). In this section, we only consider players who have top 5% (and top 25%) peers, i.e. have the chance to learn from them on paper. Pass intensity is measured on a per capita basis, we look at the total number of passes with the types of players normalized by the number of them on the team (a team often has 1 top 5% player, but possibly several). The categories are created based on the distribution of passes in a half-season between all players: if it is low (lowest 25%), it means that compared to how many passes are exchanged between any two players, the player and the stars on the team interact relatively little.

In Table 7 we start with a categorical variable of pass intensity, with the reference category of low of passing intensity (Column 1). We see that high-intensity passing – strong interaction with stars and high flyers – is what explains HC growth a great deal. This is true even when we condition on team value: players with similar quality peers but higher pass count with stars see their HC grow faster. Interestingly, this holds (if with smaller point estimates) when we also compare players with the same minutes on the pitch (Column 2). No shared minutes on the pitch with stars is not different from low amounts (and is also rare). The same results are reproduced with a continuous version of the pass count variable (in Columns 3 and 4) where we measure passes with standardized scores.

In all these cases, we see that learning is stronger from passes with high-flyers than with stars.

In Columns 5 and 6 we examine a related statistic: the shared minutes on the pitch with the stars and the high-flyers. This variable measures that out of the total mass of shared minutes with everyone on the team, what proportion is attributed to the top 5% and top 25% players on the team. For the overall sample, we can see that even without controlling for total minutes on the pitch, we find no significant associations, suggesting that indeed it is the pass-interaction and not the shared minutes in general that drives the growth in human capital of the typical players.

Table 7: Learning over time and passing: medium market value range players

Dependent Variable:	log(value$_{t+6}$)					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
log(value)	0.718***	0.669***	0.719***	0.669***	0.865***	0.675***
	(0.028)	(0.028)	(0.028)	(0.029)	(0.027)	(0.029)
Teammates' total value (Z)	0.153***	0.179***	0.168***	0.189***	0.095**	0.186***
	(0.023)	(0.023)	(0.022)	(0.023)	(0.038)	(0.040)
Pass count with top 5%: no shared minutes	-0.065	-0.063				
	(0.148)	(0.150)				
Pass count with top 5%: in 25-75 percentile	0.126***	0.086**				
	(0.034)	(0.033)				
Pass count with top 5%: in 75+ percentile	0.282***	0.166***				
	(0.044)	(0.044)				
Pass count with top 25%: no shared minutes	-0.008	3.82×10^{-5}				
	(0.130)	(0.130)				
Pass count with top 25%: in 25-75 percentile	0.190***	0.088**				
	(0.040)	(0.043)				
Pass count with top 25%: in 75+ percentile	0.485***	0.257***				
	(0.054)	(0.063)				
Total minutes (Z)		0.160***		0.135***		0.256***
		(0.019)		(0.016)		(0.014)
Pass count with top 5% teammates (Z)			0.091***	0.064***		
			(0.016)	(0.015)		
Pass count with top 25% teammates (Z)			0.218***	0.145***		
			(0.019)	(0.019)		
Minutes shared with top 5% out of total (Z)					-0.012	0.009
					(0.039)	(0.039)
Minutes shared with top 25% out of total (Z)					0.016	0.037
					(0.028)	(0.026)
Fixed-effects						
position	Yes	Yes	Yes	Yes	Yes	Yes
player age	Yes	Yes	Yes	Yes	Yes	Yes
league X half-season	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	5,366	5,366	5,366	5,366	5,273	5,273
R^2	0.558	0.564	0.564	0.569	0.533	0.563
Within R^2	0.281	0.291	0.290	0.298	0.233	0.282

Note: Players in the medium market value range, all teams. Standard errors clustered at position × league × half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

For young players, we see similar patterns with higher point estimates with regard to passing. Table 8 shows that for juniors passing with the stars is incredibly important. Being in the medium passing count group is as important as having a 1 standard deviation higher average team-mate. Being in the top group in terms of passing with high-flyers is associated with an around 130%

higher market value, which still remains around 68% even after controlling for how much the young player spends on the pitch in the half-season. The continuous measurements of the variables corroborate these findings: for young individuals work-related interactions are very important in terms of growth in their human capital.

Moreover, in contrast to the full sample of medium-value range players, the time shared with top players has a remarkably different relation. We can see that even after controlling for the total minutes a young player spends in the games, sharing 1 standard deviation more minutes with high-flying peers is associated with around 17% higher market value in the 6 half-season span. Interestingly this association is not present for the shared minutes with star players. This is in accord with the results in Table 6, but we can see that the association found there is driven especially by the high-flyers: not star peers, but very good ones.

Table 8: Learning over time and passing: junior players

Dependent Variable:	log(value$_{t+6}$)					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
log(value)	0.723***	0.663***	0.716***	0.669***	0.919***	0.665***
	(0.064)	(0.064)	(0.060)	(0.063)	(0.060)	(0.064)
Teammates' total value (Z)	0.270***	0.306***	0.276***	0.301***	0.267***	0.376***
	(0.046)	(0.047)	(0.046)	(0.046)	(0.088)	(0.087)
Pass count with top 5%: no shared minutes	-0.570	-0.576				
	(0.431)	(0.431)				
Pass count with top 5%: in 25-75 percentile	0.125	0.052				
	(0.084)	(0.083)				
Pass count with top 5%: in 75+ percentile	0.326**	0.147				
	(0.131)	(0.133)				
Pass count with top 25%: no shared minutes	0.103	0.105				
	(0.237)	(0.237)				
Pass count with top 25%: in 25-75 percentile	0.298***	0.175*				
	(0.092)	(0.100)				
Pass count with top 25%: in 75+ percentile	0.831***	0.513***				
	(0.146)	(0.172)				
Total minutes (Z)		0.234***		0.165***		0.395***
		(0.060)		(0.054)		(0.045)
Pass count with top 5% teammates (Z)			0.151**	0.103*		
			(0.058)	(0.057)		
Pass count with top 25% teammates (Z)			0.354***	0.261***		
			(0.048)	(0.052)		
Minutes shared with top 5% out of total (Z)					-0.019	0.015
					(0.094)	(0.088)
Minutes shared with top 25% out of total (Z)					0.143**	0.171***
					(0.065)	(0.062)
Fixed-effects						
position	Yes	Yes	Yes	Yes	Yes	Yes
player age	Yes	Yes	Yes	Yes	Yes	Yes
league X half-season	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	1,156	1,156	1,156	1,156	1,131	1,131
R^2	0.391	0.401	0.403	0.408	0.334	0.396
Within R^2	0.294	0.306	0.308	0.313	0.224	0.296

Note: Junior players in the medium market value range, all teams. Standard errors clustered at position × league × half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Examining the intensity of the work-related interactions, we find a strong association of high-intensity interaction with growth in human capital. In addition, this channel is even more pronounced for junior players, whose in-match interactions via passes, or even joint presence during the game, are associated with large and statistically significant gains in market value.

6 Conclusions

In this paper, we study the effect of peer quality on human capital growth, with special emphasis on the impact of top-level coworkers. We build on and extend the literature, primarily the work of Jarosch et al. (2021a), in several aspects such as incorporating the intensity of collaboration, deriving mechanisms, and measuring human capital more precisely. We use sports data that enable us to access in fine detail the human capital growth and collaboration between colleagues. We find that beyond having high-quality coworkers in our nearest surroundings, the intensity and nature of the interaction are also key components in the evolution of players' market value. Having better teammates early on is associated with substantially higher future market value growth, and the connection is even more pronounced for younger players.

In future iterations of the paper, we plan to incorporate skill measurements such as passing or ball control to capture the development of young players in a more detailed way compared to market value. Furthermore, we expect to extend the analysis by adding other influential components to the mechanisms such as the role of coaching, mentorship of players past their prime, or team culture.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Arcidiacono, P., Kinsler, J., and Price, J. (2017). Productivity Spillovers in Team Production: Evidence from Professional Basketball. *Journal of Labor Economics*, 35(1):191–225.

Arellano-Bover, J. and Saltiel, F. (2023). Differences in on-the-job learning across firms. Technical report, CAGE working paper no. 670.

Bell, A., Chetty, R., Jaravel, X., Petkova, N., and Van Reenen, J. (2017). Who Becomes an Inventor in America? The Importance of Exposure to Innovation.

Bilen, E. and Matros, A. (2021). The Queen's Gambit: Explaining the Superstar Effect Using Evidence from Chess.

Bonhomme, S., Holzheu, K., Lamadon, T., Manresa, E., Mogstad, M., and Setzler, B. (2023). How much should we trust estimates of firm effects and worker sorting? *Journal of Labor Economics*, 41(2):291–322.

Békés, G. and Ottaviano, G. I. P. (2022). Cultural Homophily and Collaboration in Superstar Teams. Technical Report DP17618, CEPR.

Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality*. *The Quarterly Journal of Economics*, 128(3):967–1015.

Cornelissen, T., Dustmann, C., and Schönberg, U. (2017). Peer Effects in the Workplace. *American Economic Review*, 107(2):425–456.

Fraiberger, S. P., Sinatra, R., Resch, M., Riedl, C., and Barabási, A.-L. (2018). Quantifying reputation and success in art. *Science*, 362(6416):825–829.

Gregory, V. (2021). Firms as learning environments: Implications for earnings dynamics and job search. Technical report, Federal Reserve Bank of St. Louis.

Hoegele, D., Schmidt, S. L., and Torgler, B. (2014). Superstars as Drivers of Organizational Identification: Empirical Findings from Professional Soccer. *Psychology & Marketing*, 31(9):736–757.

Hoey, S. (2023). One's pain is another's gain - early career exposure and later labor market outcomes. Technical report, University of Liverpool.

Jarosch, G., Oberfield, E., and Rossi-Hansberg, E. (2021a). Learning From Coworkers. *Econometrica*, 89(2):647–676.

Jarosch, G., Oberfield, E., and Rossi-Hansberg, E. (2021b). Supplement to 'learning from coworkers'. *Econometrica Supplemental Material*, 89.

Li, W., Aste, T., Caccioli, F., and Livan, G. (2019). Early coauthorship with top scientists predicts success in academic careers. *Nature Communications*, 10(1):5170.

Mas, A. and Moretti, E. (2009). Peers at Work. *American Economic Review*, 99(1):112–145.

Nix, E. (2020). Learning spillovers in the firm. Technical report, USC.

Poli, R., Besson, R., and Ravenel, L. (2021). Econometric Approach to Assessing the Transfer Fees and Values of Professional Football Players. *Economies*, 10(1):4.

Werner, K. and Dickson, G. (2018). Coworker knowledge sharing and peer learning among elite footballers: Insights from german bundesliga players. *Sport Management Review*, 21(5):596–611.

A Appendix

A.1 Variable descriptions

Let us define the key variables of the analysis.

- **Market value** (Transfermarkt): mean market value of the player in the half-season, without evaluation the last valid evaluation is assigned to the player
- **Teammates' total market** value: the sum of the market value of the teammates' of the player based on their mean market value in the half-season
- **Total minutes**: total minutes spent on the pitch in the half-season by the player, measured in Z-scores (standardized with mean and standard deviation)
- **Pass eigenvector centrality**: based on the total pass network of the team in the half-season, calculates the eigenvector centrality of the player (proxies the importance of the player in the team's overall play), measured in Z-scores (standardized with mean and standard deviation)
- **Top 5% of players**: based on position, the player is or in the last 2 years was in one of the half-seasons in the top 5% of the market value distribution in the top 5 leagues
- **Top 25% of players**: based on position, the player is or in the last 2 years was in one of the half-seasons in the top 25% of the market value distribution in the top 5 leagues
- **Seniors vs. juniors**: players are split by age at 23 years of age
- **Pass count with top 5% or 25% of players categories**: per capita number of passes with top 5% and 25% of players, proxies the direct interaction with the top players on the team. Calculated only for those who have top 5% (and top 25%) peers.
  - No top player: no top player on the team
  - No shared minutes: no shared minute with a top player
  - lowest 25 percentile: the number of passes is in the 25-75th percentile of the pass distribution between couples of players in the full dataset in a half-season
  - middle 25-75 percentile: the number of passes is in the 25-75 percentile
  - top 75+ percentile: the number of passes is in the 75+ percentile

- **Pass count with top 5% or 25% of players** (numeric): number of passes with top 5% or 25% players in the half-season, measured in Z-scores (standardized with mean and standard deviation)
- **Minutes shared with top 5% or 25% of players out of total** (numeric): fraction of minutes shared with top 5% or 25% of players out of total shared minutes with all players, measured in Z-scores (standardized with mean and standard deviation)
- **Hirschman–Herfindahl Index**: HHI index of the market values of the squad members in the half-season
- **Share of top 5% or 25% players on team**: share of players in the top 5% or 25% of the players' market value distribution

Control sets of dummies (Fixed-effects) in the regressions:

- **League × half-seasons**: league indicators interacted with the half-seasons
- **Position**: broad position of the player, goalkeeper, defender, midfielder, forward
- **Player age**: player's age dummies

A.2 Descriptive statistics

Tables A1, A2 describe the key numeric and categorical variables of the analysis. Table A1 examines the distribution in the full sample, while Tables A4 and A5, along with A3 split the sample according to position in the market value distribution by excluding low and very high-valued players, and then also by seniority, splitting at the age of 25. Cutting off the tails of the distribution is a step towards creating common support in the regressions: the highest value individuals in the dataset skew the distribution as we can see in Table A4. The comparisons in Table A5 reveal that after the cut, junior and senior players are fairly comparable in their most relevant observables.

Table A1: Descriptives of the key numeric variables

	N	mean	sd	p1	p5	p50	p95	p99
Value (in mn EUR)	33148	7.88	13.02	0.20	0.40	3.25	30.00	65.00
log(value)	33148	15.06	1.30	12.21	12.90	14.99	17.22	17.99
Teammates' value (in mn EUR)	33148	173.14	189.00	21.34	30.43	96.78	585.00	919.74
Teammates' value (Z-score)	33148	0.00	1.00	-0.80	-0.76	-0.40	2.18	3.95
log(Teammates' value)	33148	18.53	0.90	16.88	17.23	18.39	20.19	20.64
HHI of team	33148	0.07	0.01	0.05	0.05	0.07	0.10	0.12
Share of top 5% players in team	33148	0.11	0.19	0.00	0.00	0.00	0.59	0.74
Share of top 25% players in team	33148	0.36	0.20	0.00	0.04	0.35	0.68	0.77
Total minutes	33148	913.20	579.47	11.00	62.00	897.00	1840.00	2175.53
Total minutes (Z-score)	33148	0.00	1.00	-1.66	-1.48	-0.01	1.58	2.17
Pass eigenvector centrality (Z-score)	32396	0.00	1.00	-1.43	-1.37	-0.12	1.89	1.95
Pass count top 5% (Z-score)	14907	0.00	1.00	-0.95	-0.95	-0.29	1.96	3.63
Pass count top 25% (Z-score)	32512	0.00	1.00	-1.04	-1.02	-0.27	1.96	3.40
Minutes shared with top 5% (Z-score)	32528	0.00	1.00	-0.60	-0.60	-0.60	2.46	2.95
Minutes shared with top 25% (Z-score)	32528	0.00	1.00	-1.78	-1.56	0.03	1.56	1.94

Note: All players, all teams.

Table A2: Distribution of the key categorical variables

		N	%
Pass count with top 5%	1. no player	18241	55.0
	2. no minutes	120	0.4
	3. low	2596	7.8
	4. mid	6154	18.6
	5. high	6037	18.2
Pass count with top 25%	1. no player	636	1.9
	2. no minutes	489	1.5
	3. low	5579	16.8
	4. mid	17611	53.1
	5. high	8833	26.6

Note: All players, all teams.

Table A3: Distribution of the key categorical variables in groups

		tails of value dist.				middle of value dist.			
		junior		senior		junior		senior	
		N	Percent	N	Percent	N	Percent	N	Percent
Pass count with top 5%	1. no player	1641	4.95	5262	15.87	2186	6.59	9152	27.61
	2. no minutes	17	0.05	32	0.10	14	0.04	57	0.17
	3. low	404	1.22	745	2.25	405	1.22	1042	3.14
	4. mid	727	2.19	2336	7.05	665	2.01	2426	7.32
	5. high	770	2.32	3286	9.91	241	0.73	1740	5.25
Pass count with top 25%	1. no player	68	0.21	286	0.86	57	0.17	225	0.68
	2. no minutes	71	0.21	178	0.54	61	0.18	179	0.54
	3. low	978	2.95	2020	6.09	742	2.24	1839	5.55
	4. mid	1683	5.08	6079	18.34	1961	5.92	7888	23.80
	5. high	759	2.29	3098	9.35	690	2.08	4286	12.93

Note: All players, all teams.

Table A4: Descriptives of the key numeric variables by place in value distribution

	tails of dist.				middle of dist.			
	N	mean	sd	p50	N	mean	sd	p50
Value (in mn EUR)	15220	12.83	17.86	9.50	17928	3.67	1.97	3.00
log(value)	15220	15.16	1.83	16.07	17928	14.98	0.53	14.91
Teammates' value (in mn EUR)	15220	233.07	234.72	144.86	17928	122.26	116.77	83.45
Teammates' value (Z-score)	15220	0.32	1.24	-0.15	17928	-0.27	0.62	-0.47
log(Teammates' value)	15220	18.76	1.04	18.79	17928	18.34	0.71	18.24
HHI of team	15220	0.07	0.01	0.07	17928	0.07	0.01	0.07
Share of top 5% players in team	15220	0.17	0.24	0.04	17928	0.06	0.13	0.00
Share of top 25% players in team	15220	0.35	0.20	0.35	17928	0.36	0.20	0.35
Total minutes	15220	934.50	631.66	920.00	17928	895.12	530.51	882.00
Total minutes (Z-score)	15220	0.03	1.09	0.01	17928	-0.03	0.91	-0.02
Pass eigenvector centrality (Z-score)	14824	-0.03	1.03	-0.16	17572	0.03	0.97	-0.08
Pass count top 5% (Z-score)	8317	0.15	1.04	-0.12	6590	-0.19	0.91	-0.49
Pass count top 25% (Z-score)	14866	-0.05	0.98	-0.33	17646	0.04	1.01	-0.22
Minutes shared with top 5% (Z-score)	14894	0.34	1.21	-0.32	17634	-0.28	0.66	-0.60
Minutes shared with top 25% (Z-score)	14894	-0.14	1.01	-0.22	17634	0.12	0.97	0.20

Note: All players, all teams.

Table A5: Descriptives of the key numeric variables by seniority, middle of value distribution

	junior				senior			
	N	mean	sd	p50	N	mean	sd	p50
Value (in mn EUR)	3511	3.80	2.04	3.23	14417	3.64	1.96	3.00
log(value)	3511	15.00	0.55	14.99	14417	14.97	0.53	14.91
Teammates' value (in mn EUR)	3511	131.55	127.89	89.00	14417	119.99	113.79	82.57
Teammates' value (Z-score)	3511	-0.22	0.68	-0.45	14417	-0.28	0.60	-0.48
log(Teammates' value)	3511	18.39	0.74	18.30	14417	18.32	0.71	18.23
HHI of team	3511	0.07	0.01	0.07	14417	0.07	0.01	0.07
Share of top 5% players in team	3511	0.07	0.14	0.00	14417	0.05	0.12	0.00
Share of top 25% players in team	3511	0.36	0.19	0.35	14417	0.36	0.20	0.35
Total minutes	3511	708.42	499.87	644.00	14417	940.58	527.82	946.00
Total minutes (Z-score)	3511	-0.32	0.88	-0.43	14417	0.04	0.90	0.07
Pass eigenvector centrality (Z-score)	3422	-0.22	0.94	-0.42	14150	0.09	0.97	0.00
Pass count top 5% (Z-score)	1325	-0.41	0.71	-0.65	5265	-0.13	0.94	-0.44
Pass count top 25% (Z-score)	3454	-0.17	0.93	-0.44	14192	0.09	1.03	-0.16
Minutes shared with top 5% (Z-score)	3434	-0.25	0.69	-0.60	14200	-0.29	0.65	-0.60
Minutes shared with top 25% (Z-score)	3434	0.12	0.93	0.21	14200	0.12	0.98	0.20

Note: Middle value range players, all teams.

A.3 Data cleaning steps: sample selection

Sample selection steps were the following:

- We start from the histories of player values (N = 1,007,727). We discard all observations without a valid market value: meaning before the first valuation (mostly U17 or U19 player histories) as we complete the market value histories with past values if it is not available. (641,881 rows remain)
- Keep only those players who have at least 11 players with a valid valuation on their team. (433,770 rows remain)
- Keep those that are:
  - in the relevant seasons of 2012/2013-2019/2020
  - top 5 European football leagues
  - at least 18 years of age
  - data cleaning: has valid team id, has less than 50 player to interact with and non-negative minutes

So valid pass-data are not required at this point. 33,148 observations remain of playerid X halfseasons.

A.4 Advantages over admin data

In what follows, let us discuss the literature more in detail and discuss how we improve on key aspects.

First, a worker's wage (even relative to the firm average) reflects HCA only partially. The wage is also a function of bargaining and the acquisition of firm-specific skills. If they are correlated with tenure, it might confound results. Wage growth is rarely fast following HCA, and may even appear after a switch[9].

How will this setting improve on what we know? First, we use direct human capital (HC) estimates rather than wages. Human capital would be impossible to observe in a regular workplace setting, but in sports, production (games) is in plain sight. A change in transfer value will be interpreted as a change in HC. This could be an increase as players evolve and learn from others. It could be a decrease as players age or become injured. To be more specific, we use an adjusted transfer value cleaned from the impact of contracting conditions (such as remaining duration). Unlike wages, this measure captures exactly what we are interested in, and there is no need to model wage setting or wait for wages to adapt to HC accumulation,

Second, identifying peers, based on firm-occupation group is hard and noisy, especially across a large variety of firm types and sizes. Measurement error may attenuate estimates and yield low learning coefficients.

Indeed, many empirical papers suffer from a heterogeneous worker pool in terms of occupation. Jarosch et al. (2021a) and Cornelissen et al. (2017) use German labor data, where workers are known to be part of a firm and occupation. The latter explicitly shows that peer pressure is at work at repetitive tasks, while there is no evidence for learning in higher-skilled jobs. The key problem is that there is no way to know which worker is spending time with which others. Engineers at a large Siemens plant may be spread out. Assembly workers may be sorted by skills or not. Instead in our setting, there is a great deal of homogeneity: all workers are men, high-skilled, and between 18-35 years of age. These are young and very competitive players in a high-pressure environment, with easily comparable and measurable performance. Furthermore, they are all in the same, very narrow occupation: football player. It is actually a 5-digit code (3441/football player).

[9] Your wage might change after good publications immediately, but only after promotions.

The three-digit category (344) will include box referees, fitness instructors or lifestyle consultants[10]. This means, that if there is ever a place where both peer pressure and learning might occur, this is it. In that sense, we shall find an upper estimate of what may be expected outside of sports.

The size of the peer group may be correlated with skills (occupation). First, some occupations (skill categories) may be typically found in small (cooks) or large firms (assembly workers). Second, in a car factory, many auto assembly workers and few engineers, in an engineering company, many engineers, and few secretaries. This may bias the variation of learning by skill types.

In the German admin data, teams of the same establishment and occupation average 18 but the distribution is very skewed (Jarosch et al., 2021b). Most firm-occupation bins are very small, while a few may be very large.

Instead in our data, teams are of very similar composition and size. This helps partial out the role of the number of peers.

A.5 Most important factors for TM market values

Via the website https://www.transfermarkt.com/market-value-definition/thread/forum/357/thread_id/3433 the following factors are considered when producing the market value of a player.

1. Most important factors:

    - Future prospects
    - Age
    - Performance at the club and national team
    - Level and status of the league, both in sporting and financial terms
    - Reputation/prestige
    - Development potential
    - League-specific features
    - Marketing value
    - Number & reputation of interested clubs
    - Performance potential

[10] In the UK using the ONS system, the relevant 3 digit code is 344. In Germany, Jarosch et al. (2021a) uses the KldB-88 system, the three-digit occupation is 838 "Performers, professional sportsmen, auxiliary artistic occupations"

- Experience level
- Injury susceptibility
- Different financial conditions of clubs and leagues
- General demand and ”trends” on the market
- General development of transfer fees
- External factors such as the coronavirus pandemic and its consequences

2. Individual modalities:

- Transfers via an option to buy/obligation to buy
- Loan fee
- Only part of transfer rights acquired
- Exit clause
- Buyback option
- Player swap deal
- Contract length
- Resale participation
- Bonus payments
- Improvement of financial balance

3. Situational conditions:

- Pressure situations such as competitive, success or financial pressure, etc.
- Will/desire/interests of the player
- Club does not sell to highest bidder
- Player goes on strike or similar
- High salary
- Club wants to sell player