

Mechanisms of Learning in Collaborative Jobs*

Gábor Békés¹ and Bence Szabó ^{†2}

¹Central European University, KRTK, CEPR

²Corvinus University of Budapest, Magyar Nemzeti Bank

August 1, 2025

Abstract

How do workers learn from their coworkers in the workplace, especially when collaboration is essential? We investigate the mechanisms of learning in a high-skill collaborative environment: top teams in European men's football. We start by confirming the positive effect of coworker quality found in administrative data. Our setting is conducive to finding evidence of learning as it can compare players in very similar units of soccer squads. To reduce sorting fears, we augment wages by market values as our human capital estimate, and we show that we capture human capital evolution with less bias. We first confirm existing evidence on peer effects: wage growth is faster in better teams. Second, we show that given average peer quality, talent composition does not play a role. Third, having a more important role on the team – more minutes – correlates strongly with learning. Moreover, it is a confounder leading to *underestimated* learning in its absence. Fourth, our detailed data allows for examining details of collaboration. We find that intensive collaboration with high-quality peers matters the most. Additionally, we corroborate the wage growth findings by examining the evolution of relevant skill attributes as well: learnable skills improve due to better peers while non-learnable skills are not related. Finally we separate individual development from team complementarities, showing that higher accumulated skills persist even when workers change jobs.

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

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

*We thank Daniel Prinz, our discussant at the 2023 Conference of The Hungarian Society of Economics, for the valuable feedback on the work. Furthermore, we thank the participants of the 2023 workshop of Research in Economics using Sports Data, the 2024 Szirák Conference on Labor Economic Research, the Seminar of the KRTK Institute of Economics, and the Empirical Research Seminar of the Corvinus University, for providing useful comments helping our research. All mistakes are our own.

[†]Corresponding author: bence.szabo@uni-corvinus.hu

1 Introduction

Learning from colleagues is an essential source of human capital growth. While convincing evidence shows that long-run wage growth is strongly related to the quality of peers (Jarosch et al., 2021; Herkenhoff et al., 2024; Pastorino, 2024), the mechanisms that allow this relationship to emerge are less explored empirically. Most studies utilize linked employer–employee datasets where peers are often defined by having the same occupation category at the same firm or establishment at the same time, without the ability to identify and characterize actual working relationships (the actual links) between individuals.

In this paper, we open up the black box of peer effects and study how human capital accumulates via learning from colleagues in a highly competitive and collaborative profession: European football. We contribute by providing a detailed examination of the following mechanisms: exposure to top-rated coworkers, joint participation in production, direct interactions, the development of wages and relevant skill attributes, and the persistence of human capital growth. While a sports environment is specific in some regards, it can serve as a model industry for other high-skilled, competitive, and efficient labor markets with intense collaboration between colleagues such as academia, R&D, or consulting (Fonti et al., 2023).

We construct a fine-grained dataset covering a decade of professional European football, combining several datasources to track football players’ human capital (wage, market value, and skill attributes), their role in teamwork, and direct collaboration through event-level interactions shared playing time. Our approach eliminates key issues of linked administrative data. First, generally wages react with delay to human capital changes, and if better firms tend to match up with workers who have higher growth potential, peer quality will be positively correlated with higher growth potential, hence peer effects will be overestimated in usual settings. For football players, we can control for growth potential using frequently updated market value information, eliminating this source of bias. Second, we observe the production process of the team’s output (learning on the job), which is crucially affected by peer quality: better peers might hinder meaningful participation in the team’s production. Third, we also observe direct peer interactions, and utilize the quality of the direct relationships workers are involved in. And fourth, as the rules of the game regarding team size and production are strictly given, the endogeneity of firm size and peer exposure does not play a role in our setup, allowing for better-controlled comparisons.

After confirming baseline peer effect results of the literature, and we present four key findings

on the mechanisms of learning. First, how human capital growth is related to the composition of peers and exposure to stars. We find that conditional on the average, the distribution of peer quality (e.g. share of top-tier colleagues) is not important. Second, we consider the player’s role and collaboration in production. If the most instructive tasks are done by the most highly skilled colleagues then higher quality peers could hinder learning-by-doing via diminishing the individual’s role. Indeed we confirm that with better peers the role in production decreases (minutes played and centrality in the passing network), hence the learning-from-peers channel’s association with wage growth is underestimated. Third: using pass intensity, we find that more interaction with top colleagues is associated with higher wage growth. And finally, we find that the correlation between wage growth and peer quality persists even after leaving the initial team, suggesting that the estimates are not driven only by production complementarity. In several robustness checks, we focus on the issue of selection bias along unobserved human capital growth potential (‘hidden gems’) that better organizations might be able to detect more effectively, and present evidence that our results should not be substantially affected by this bias.

The paper continues with a synthesis of the related literature in Section 2, followed by a detailed description of the data we use in Section 3 and the empirical strategy in Section 4. In Section 5, we present our results step-by-step on exposure to top colleagues, contribution to production, direct interactions, and persistence. Finally we examine the most important dangers to our empirical strategy, and conclude.

2 Related literature

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 Past and current workplace environments (including peer quality) are key contributors to worker performance and compensation.

Important related literature uses matched employer–employee data to estimate what drives wages using worker and firm fixed effects with the 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 our work, there is a related stream of literature investigating

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

the role of firms in explaining the variation in wage *growth* (Gregory, 2021). Finally, several papers looked at the heterogeneity of the work environment on learning. Arellano-Bover and Saltiel (2023) studied the learning of young employees with administrative datasets from Brazil and Italy.

This paper focuses on the quality of coworkers as a source of variation in wages. In this literature, the starting point is measuring the wage premium of better coworkers. While the evidence for learning is mixed (both lab and observational), there is more evidence for peer pressure (Mas and Moretti, 2009). Cornelissen et al. (2017) use German linked employer–employee data from Munich covering 26 years and focus on wage levels rather than productivity with special attention sectors with high knowledge intensity (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). They argue workers may alter their productivity due to both peer pressure and learning from co-workers. For repetitive tasks, they estimate a strong peer effect of 6%, but near zero for the most knowledge-intensive jobs, suggesting that peer pressure dominates learning in their setup.

The closest paper to ours is Jarosch et al. (2021) 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 10% higher average wage of peers is associated with 0.7% higher wage growth within one year, rising over time, with cumulative growth reaching 2.1% over ten years. Furthermore, they show that having more highly-paid coworkers is strongly associated with future wage growth. Also, unlike for wage levels, wage growth can be detected for top performers.

There is less evidence on what aspect of peers help human capital accumulation. 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%. Hong (2022) dissects peer effects as the externality a workplace may generate. Production complementarity arises when a peer is a good match so much as so that it yields mutual gains in earnings. Second, learning appears as knowledge spillover from peers. Importantly, in his model, there is positive assortative matching within teams due to production complementarity, leading to unequal learning.

Another finding in the literature is related to heterogeneity in peers. Jarosch et al. (2021) documents large returns on interacting with high-earning colleagues: superstars might especially drive the differences across firms in many industries. As the complexity of jobs increases, learning

specialized know-how from coworkers becomes even more important in career advancement (Caicedo and Lucas Jr, 2019). Superstars, therefore, are expected to have substantial effects on coworker 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 across several domains shows mixed results.²

Several studies use sports data to examine mechanisms of peer effect. The role of stars was shown by Hoegele et al. (2014). Hoey (2023) found that injury of stars helps young players play more, and thus helps talent discovery. Better peers may also help by elevating your game: Arcidiacono et al. (2017) estimates the contribution of peers to perform better on the pitch from via observing a large variation in peers (of different positions) across games. Cohen-Zada et al. (2024) also uses sports data, Israeli soccer, to estimate effort peer effects in terms of making an effort. Even a single player’s improvement can have substantial benefits for team activity and performance. As for example for the negative effect: exposure to superstars might actually be detrimental to self-confidence. For instance, in competitive environments, the higher pressure may lead to more mistakes (Bilen and Matros, 2023).

3 Data

3.1 Overview

We start from a recently assembled dataset (Békés and Ottaviano, 2025) web scraped from different sources including [transfermarkt.com](https://www.transfermarkt.com), [whoscored.com](https://www.whoscored.com), that covers each event (each pass, tackle, etc.) from all matches, along with team composition and results of games in the top seven men’s football leagues (Premier League in England, Ligue 1 in France, Bundesliga in Germany, Serie A in Italy, La Liga in Spain, the Eredivisie in the Netherlands, and the Portuguese Primeira Liga) over eight sporting seasons (2012-13 to 2018-19, except for Portugal that starts in 2015-2016).³ We augment them with information on skill attributes and wages gathered from [fifa.com](https://www.fifa.com), and salaries from [capology.com](https://www.capology.com). Linking these datasources, we can observe the evolution of players’ estimated market

²There 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., 2019), the fine arts (Fraiberger et al., 2018), research (Li et al., 2019).

³Data 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.5). By ‘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 $20 \times 19 = 380$ games), while there are 18 teams in the Bundesliga, in the Eredivisie, and the Primeira Liga ($18 \times 17 = 306$ games). Due to relegation and promotion, we have players from a total pool of 181 teams in the sample.

values, wages, skill attributes, along with their in-game (at-work) interactions and other individual and team characteristics. The details of the data linking procedure can be found in the Appendix at [A.5](#).

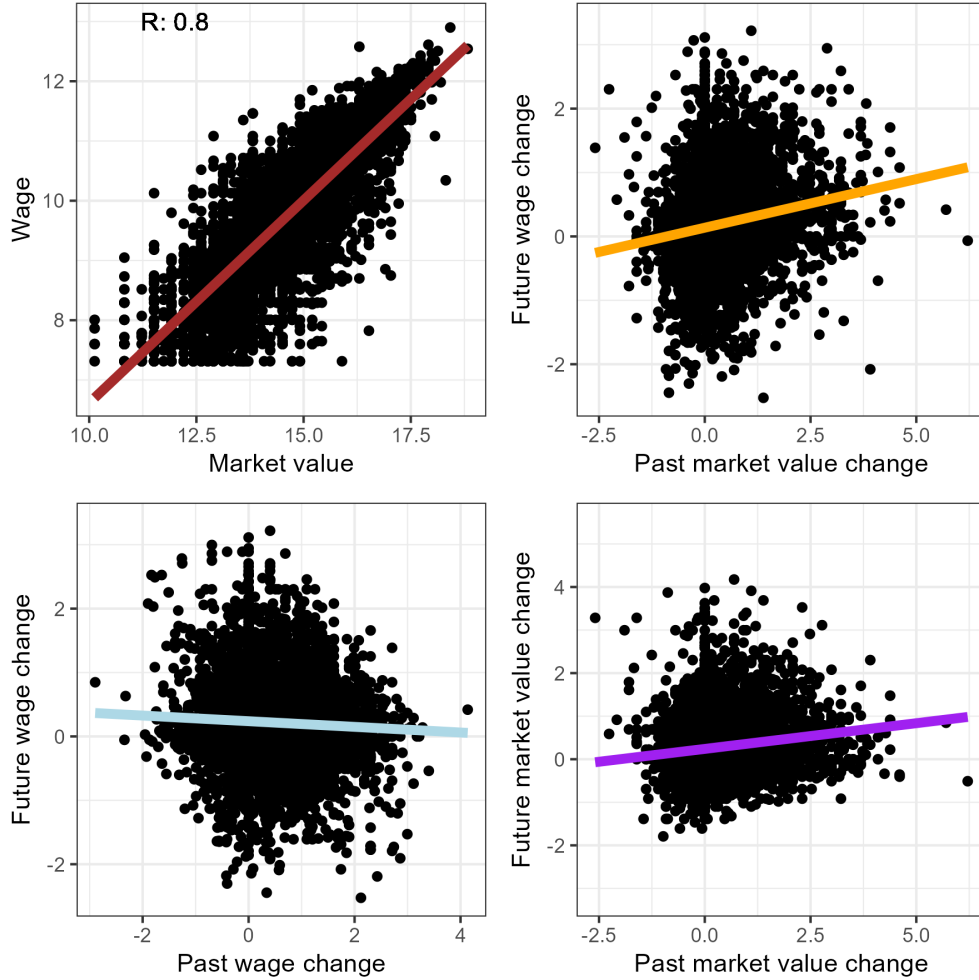
Using sports data in general, and player career data in particular has strong advantages over traditional administrative data, even if it comes with some shortcomings as well. A major advantage of our dataset over several usual administrative datasets with linked employer–employee information is that 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, work activity and 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. Our approach is supported by [Werner and Dickson \(2018\)](#) studying peer learning using semi-structured interviews in the Bundesliga. They find that learning from others, knowledge sharing functions via four main channels: imitating, peer communication, labor mobility, and knowledge brokers.

We use three metrics of human capital: wage, market value, and skill attributes highlighting different aspects of human capital. We retrieve market values from [transfermarkt.com](#)⁴ Their concept of market value targets the *expected value of a player in a free market*, they augment player pricing models with the community judgement, taking into account factors such as future prospects or contracting (an exhaustive list can be found in Appendix [A.6](#)). Using market value beside wages also allows us to control for the growth potential that might not appear in wages, along with issues of bargaining, tax and accounting considerations ([Poli et al., 2021](#)). While market value reacts more sensitively and quickly to current changes in a player’s performance and status by constructions, wages incorporate them with delay. As Figure [1](#) shows, market values and wages are highly correlated but they capture different aspects of human capital: an increase in current market value predicts higher future wages while an increase in current wages predicts the opposite.

⁴The site’s methodology is explained in detail at https://www.transfermarkt.com/market-value-definition/thread/forum/357/thread_id/3433. Transfermarkt publishes them twice in a season (for frequency see Figure [A4](#)), usually after the end of a season, and once during a season after sufficiently many games were played.

Figure 1: Comparing market value and wage dynamics



Note: The figure shows the correlation between different aspects of market value and wage log-levels and dynamics (log-differences) for the final analysis sample.

Finally, as a third measure, we track skill attributes that provide a more direct assessment of how the individual's job-related skills evolve over the lifetime of a player. These are core competences which for other industries remain mostly unobserved, highlighting an additional advantage of using sports data. Skill ratings contain 28 elements scaled 1–100, determined for each player by experts based on advanced stats and scouting information⁵. The correlation structure of the skill ratings can be found in the Appendix at A2, we selected five main attributes for analysis: short

⁵A large team of data editors, reviewers and scouts at Electronic Arts used advanced stats along with extensive scouting and manual adjustments to produce a detailed rating system that represents players well in their computer game. For more details on their approach, visit <https://www.espn.com/soccer/story/-/id/37456940/fifa-17-player-ratings-system-blends-advanced-stats-subjective-scouting>, accessed 2025/01/23.

pass, reactions, interceptions, finishing, and sprinting speed, reflecting relatively uncorrelated but important aspects of players ability.⁶ We expect some skills to develop with collaboration such as short passes, but some are expected to be more unaffected by peers such as sprinting speed.

To capture peer effects, 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 and wage information to produce the average peer quality, along with a Herfindahl—Hirschman index, and related measures. Second, we identify star players as those that have been in the top 5% and 25% of the wage distribution in the relevant top 7 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). Third we track the direct 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. Collaboration is measured as the number of minutes and passes in games during regular season, while besides minutes we measure centrality in the production via the eigenvector centrality of a player in the dynamic directed network of pass interactions between teammates.

In our final dataset, we aggregate all variables to a player \times half-season level to reflect the the market valuation update frequency. A half-season contains typically 16-20 games in either the Summer/Fall (August to January) or Winter/Spring (February to May).

3.2 Descriptives

Detailed descriptive statistics, variable descriptions, and the steps of creating the final dataset are presented in Sections A.1 and A.2 of the Appendix. Each player appears in the dataset once per spell at a team for spells that lasted at least two half-seasons. We required at least 6 half-seasons of observability and valid market values, wages and pass information. The differences between the entire panel with some baseline restrictions, and the final sample is reported in A23.

Table A10 presents the descriptive statistics of our key variables. The final dataset contains 5,227 observations. The distribution of wages, averaging €1.43 million, is right-skewed with a median of

⁶‘Short passing’ refers to passing ability in closer range, an overall important skill attribute; ‘reactions’ captures the mental capacity to read situations; ‘interceptions’ refers to the defensive ability to intercept the passes of opponents; ‘finishing’ refers to the offensive ability of scoring; and ‘sprinting speed’ captures the physical ability of running speed.

€0.96 million. The values range significantly from as low as €0.08 million (p1) to as high as €8.06 million (p99). There are a handful of players in the top 1% with values close or even above 20 million. The average teammates' wage also has a right-skewed distribution, with a mean of €1.55 million and a median of €1.18 million, while it ranges from €0.14 million (p1) to €7.09 million (p99). The average yearly wage in our dataset is around third of the average market value. Skill attributes are on a 1–100 scale, in our sample they are centered around 50–70 and have a quite symmetric distribution without reaching the maximum rating point.

In terms of concentration of quality, the HHI index has a smaller variation, the top 1% has 0.11 which shows the presence of stars in a squad of about 25 players. The share of top 5% players in the team (the 5% taken for the whole sample) indicates a skewed distribution where most teams rely very little on their top 5% of players, but some teams rely heavily, up to 78% (p99), with a broad dispersion ($SD = 0.17$). These are the star teams like Real Madrid or Manchester City. We also identify manually a select group of elite teams that are the traditional top teams in their respective leagues and regularly participate in international tournaments.⁷ Share of the top 25% players in the team is more democratic, many teams have such players (on average, 27% of the sample has such a teammate). Players' total minutes on the field average 1,077 with a symmetric distribution but a very wide range.

4 Empirical strategy

4.1 Human capital growth and peer quality

The starting point is Equation (1) after Jarosch et al. (2021), with this specification implemented separately for various horizons h (in our case, half-seasons):

$$hc_{i,t+h} = \alpha hc_{i,t} + \beta \bar{hc}_{-i,t} + \gamma X_{i,t} + \omega_{\text{player chars}} + \omega_{\text{league} \times t} + e_{i,t} \quad (1)$$

where future human capital $hc_{i,t+h}$ is captured by an individual i 's log wage in half-season $t+h$, regressed on team quality measured as the log mean wage in half-season t , $\bar{w}_{-i,t}$, controlling for i 's own log wage in year t , $w_{i,t}$, along with individual characteristics such as age. In the regression, observations are pooled. Jarosch et al. (2021) argues that such a model will show

⁷These teams are the following: Real Madrid, Barcelona, Atletico Madrid, Manchester City, Manchester United, Liverpool, Chelsea, Tottenham, Arsenal, Bayern München, Dortmund, Schalke, Juventus, AC Milan, Inter Milan, Paris Saint-Germain, Olympique Lyonnais. The selection was done on the basis of historic performance, close to Deloitte's top 20 (an annual report), but is somewhat arbitrary.

evidence of learning from coworkers. However, while labor economics focuses on wage, wage itself is an overloaded variable intending to capture human capital, as a wage contract itself depends on several other factors.

A key concern is sorting: if workers with high wage growth profiles are matched with better firms, it may confound the results. We address this issue by using market values as well: a perceived (average) value of human capital. The benefit of using market values beside wages is that it can capture the increase in human capital when it happens, reflecting the match quality between the team and the player, and the human capital growth of young people with higher predicted outside option value emerge sooner. In comparison, in standard employer-employee datasets we only see wages, which may adjust with delay, close to the end of the contract, or linked to seniority.

So we start out by replicating the base peer effect results of [Jarosch et al. \(2021\)](#) with Table 1, augmented by the variable measuring the log-difference between market value and wages. We find that controlling for individual characteristics and relevant confounders, on a three-year horizon having 10% better teammates is associated with around 3.3% higher growth on average. This is lower by around 0.5% compared to the estimate without market values on the right hand side, suggesting that indeed usual peer effects might be overstated without controlling for growth potential. As football players are top earners and we look at players at the start of their spell at the team, we should compare our estimates to the corresponding top-earner regressions of [Jarosch et al. \(2021\)](#) (Panel A of their Table IV). They find that amongst top decile of earners, 10% higher average peer salary of those who earn more than them is associated with a 3.5% higher wage growth. Our estimates are reassuringly close in magnitude to theirs, validating that our results indeed relate closely to high-skill professions.

Table 1: Human capital evolution and peer quality over time

| | h=2 | | h=4 | | h=6 | |
|-------------------------------------|---------------------|---------------------|----------------------------------|---------------------|---------------------|---------------------|
| | (1) | (2) | $\log(\text{wage}_{t+h})$ (3) | (4) | (5) | (6) |
| $\log(\text{wage})$ | 0.538*** (0.022) | 0.583*** (0.022) | 0.419*** (0.022) | 0.471*** (0.022) | 0.374*** (0.022) | 0.440*** (0.021) |
| $\log(\text{teammates' mean wage})$ | 0.277*** (0.032) | 0.242*** (0.031) | 0.354*** (0.036) | 0.313*** (0.034) | 0.386*** (0.035) | 0.333*** (0.033) |
| $\log(\text{value} / \text{wage})$ | | 0.162*** (0.015) | | 0.190*** (0.017) | | 0.241*** (0.019) |
| Observations | 5,227 | 5,227 | 5,227 | 5,227 | 5,227 | 5,227 |
| R^2 | 0.692 | 0.705 | 0.559 | 0.575 | 0.472 | 0.495 |
| Within R^2 | 0.376 | 0.400 | 0.237 | 0.265 | 0.172 | 0.208 |

Note: Standard errors clustered at position \times league \times half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league \times halfseason, age, position, injury, elite team, spell count, contract maturity.

4.2 Mechanisms of peer effect

In this paper, we set out to study the mechanisms via which peer effects turn into a growth of human capital. Peer quality can affect human capital growth, hence wage growth of individuals via two main channels. The first is learning: by interacting directly with colleagues, or by observing and being around them, productivity might increase leading to becoming more valuable on the labor market. However, as a second channel, the quality of our peers can affect our role in the team's production. If workers are relegated to less instructive tasks when having great peers around, the channel of learning from them could be underestimated without controlling for the negative effect on learning-by-doing as the two are captured jointly. We show that in our setup having more valuable peers indeed hinder the participation in meaningful tasks (minutes played and centrality in passing). In admin data, the actual production of a worker is generally not observable, but not blocking this channel of association results in the underestimation of learning from peers. As a last component, it could also happen that beside learning from peers, workers could simply 'look better' next to high quality colleagues. ('stardust'). We are going to check how much of this influence can be detected in our data.

To capture these different channels, we expand the set of X variables to include composition of peers, the minutes played, centrality in the passing structure of the team, and the direct interaction with top colleagues, while we examine different measures of human capital. We estimate pooled cross-section OLS regressions, with standard errors clustered at the league \times half-season \times position level. Cross-section in the sense that variation comes from different players across different teams,

and not within teams (similarly to Jarosch et al. (2021)). As common in the literature, establishment fixed effects are not included as they capture almost all variation in peer quality, hindering identification. Around 95% of the variation in peer quality is explained by team and league \times half-season fixed effects.⁸

An important threat to our identification strategy is the unobserved growth potential that could possibly relate to both better peer quality and realized human capital growth, introducing selection bias. Let us note that this bias affects not only our setup, but any peer effect estimation without a specific strategy to control for it. We address this threat in three ways. First, we use the post-arrival market value at the new team as baseline, which should reflect the match quality and the elimination of some of the ex ante uncertainty about growth potential. Second, we show that peer effects of their initial teams persist for switchers as well. And third, we run several robustness checks including short-term future value in the right hand side to show that even after potential ‘overcontrolling’, most of the peer effects remain.

We use a rich set of controls. Instead of occupations, we use position fixed effects, and we add league \times half-season fixed effects to all models to control for country-specific and time-dependent differences as labor markets are closely but not perfectly integrated. We also control for age, injury, contract maturity, and tenure captured by spell counts to improve comparability, and instead of team fixed effects we use an indicator for elite teams who regularly participate in the most prestigious international tournaments.

5 Results

5.1 Composition: exposure to stars

In this section, we look at what composition of peers have the most influence on a player’s human capital development. This question relates to the distribution of peers given the average quality. One way to gauge it is to add the Hirschman–Herfindahl Index (HHI) of the wages to measure concentration on the team. A 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 terms

⁸In Appendix Table A11, we present how much of the peer quality variation is explained by different sets of fixed effects.

of wage any time in the last two seasons, as share of the lineup.⁹.

Table 2 reports that given the average level of peers, composition does not matter significantly. While 10% higher average peer wage is conditionally associated with 3% higher wage growth (4% for juniors), the relationship with team composition is not statistically significant. This claim holds with either measurement of team composition, having top players on the team (conditional on the overall talent) is not associated with higher wage growth.

Table 2: Wage evolution and exposure to stars

| | log(wage _{t+6}) | | | | | |
|------------------------------|---------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | (1) | Full sample (2) | (3) | (4) | Junior (5) | (6) |
| log(wage) | 0.440*** (0.021) | 0.440*** (0.021) | 0.439*** (0.021) | 0.398*** (0.032) | 0.397*** (0.032) | 0.399*** (0.032) |
| log(value / wage) | 0.241*** (0.019) | 0.241*** (0.019) | 0.239*** (0.019) | 0.218*** (0.025) | 0.218*** (0.025) | 0.218*** (0.025) |
| log(teammates' mean wage) | 0.333*** (0.033) | 0.333*** (0.033) | 0.290*** (0.054) | 0.404*** (0.047) | 0.405*** (0.048) | 0.406*** (0.083) |
| Team HHI of wages | | 0.310 (0.994) | | | -0.644 (1.86) | |
| Share of top 5% teammates | | | 0.226 (0.180) | | | -0.099 (0.293) |
| Share of top 5-25% teammates | | | 0.099 (0.116) | | | 0.028 (0.177) |
| Observations | 5,227 | 5,227 | 5,227 | 1,957 | 1,957 | 1,957 |
| R ² | 0.495 | 0.495 | 0.495 | 0.472 | 0.472 | 0.472 |
| Within R ² | 0.208 | 0.208 | 0.208 | 0.211 | 0.211 | 0.211 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

As a second component, we also study how the growth in relevant skill attributes relates to peer quality in Table 3. The table reports that the share of top players is not associated significantly with any of the examined skill attributes, while the average peer quality does relate to short passing and reactions abilities: 10% higher peer quality is associated with 0.16, and 0.21 higher rating points on average. Additionally, we find no significant relationships with the other three attributes, partially corroborating that our measurements reflect skill development along 'learnable' skills and not merely selection bias regarding unmeasured growth potential. We return to this point in the next section where two key channels of peer effect is examined in detail.

⁹One extra top 5% player means that with a standard lineup of 23 men, the share of top players increases by around 4.3%.

Table 3: Skill evolution and exposure to stars

| | Short pass _{t+6} (1) | Reactions _{t+6} (2) | Interceptions _{t+6} (3) | Finishing _{t+6} (4) | Sprint speed _{t+6} (5) |
|------------------------------|----------------------------------|---------------------------------|-------------------------------------|---------------------------------|------------------------------------|
| Own skill | 0.537*** (0.019) | 0.374*** (0.017) | 0.705*** (0.016) | 0.779*** (0.015) | 0.877*** (0.015) |
| Team avg. skill | 0.013 (0.040) | 0.152*** (0.047) | -0.062 (0.040) | -0.037 (0.040) | -0.054 (0.041) |
| log(value / wage) | 0.500*** (0.123) | 0.632*** (0.118) | 0.173 (0.167) | 0.434*** (0.148) | -0.049 (0.151) |
| log(teammates' mean wage) | 1.24*** (0.287) | 1.04*** (0.304) | 0.499 (0.418) | 0.415 (0.381) | -0.056 (0.414) |
| Share of top 5% teammates | -0.133 (1.04) | 0.065 (1.03) | 0.165 (1.32) | 1.07 (1.42) | 2.37 (1.53) |
| Share of top 5-25% teammates | 0.876 (0.666) | -0.025 (0.697) | 0.816 (0.850) | -0.866 (0.861) | -0.390 (0.862) |
| Observations | 5,227 | 5,227 | 5,227 | 5,227 | 5,227 |
| R ² | 0.607 | 0.415 | 0.870 | 0.850 | 0.701 |
| Within R ² | 0.392 | 0.219 | 0.545 | 0.596 | 0.581 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

5.2 Learning by doing and learning from peers

After establishing that the mere presence of stars is not related to wage growth, we can turn to how peer quality is transformed into human capital growth via two channels. The first channel is the learning channel: working with more talented colleagues is associated with higher human capital growth. The second channel is a negative one via the diminished role in production. In football and several other production environments, the number of meaningful tasks on the job is limited, hence with better peers an individual can benefit from less 'learning by doing', so there is substitution between the two channels. This is crucial as depending on the job, the trade-off between learning from doing or from peers can lead to over or underestimation of the peer effect in terms of learning. In cases where there is substitution, the learning aspect could be underestimated, while in a complementary relationship the role of learning is actually overestimated.

Table 4 shows what happens when we condition on the contribution to the production of the team, measured by the number of minutes and the centrality in the team's passing network. The association between peer quality and wage growth increases from 0.33 to 0.38 (and for juniors, from 0.41 to 0.53), consistent with better peers reducing growth opportunities (for the estimates see Table A12). Conditional on average peer quality, the time spent with top 5% or 5-25% talent is not significantly related to wage growth. This result speaks to the 'stardust' channel: playing alongside stars, conditional on overall quality and contribution to the team, is not associated with

better wage growth. Playing more minutes, having a more central role, and better peers are more important for young players' growth. The table also shows that for growth, peer quality is of similar importance in terms of magnitudes as contributing meaningfully to the team's production.

Table 4: Wage evolution and contribution to production

| | log(wage _{t+6}) | | | | | |
|-------------------------------|---------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | (1) | Full sample (2) | (3) | (4) | Junior (5) | (6) |
| log(wage) | 0.344*** (0.021) | 0.343*** (0.021) | 0.341*** (0.021) | 0.291*** (0.030) | 0.287*** (0.030) | 0.286*** (0.030) |
| log(value / wage) | 0.179*** (0.018) | 0.179*** (0.018) | 0.176*** (0.018) | 0.151*** (0.023) | 0.153*** (0.023) | 0.153*** (0.023) |
| log(teammates' mean wage) | 0.427*** (0.034) | 0.435*** (0.034) | 0.383*** (0.052) | 0.519*** (0.050) | 0.544*** (0.051) | 0.529*** (0.081) |
| Total minutes (Z) | 0.266*** (0.015) | 0.229*** (0.024) | 0.226*** (0.024) | 0.310*** (0.023) | 0.203*** (0.032) | 0.202*** (0.032) |
| Pass centrality (Z) | | 0.043* (0.022) | 0.045** (0.022) | | 0.126*** (0.034) | 0.126*** (0.034) |
| Minutes shared with top 5% | | | 0.276* (0.145) | | | 0.074 (0.230) |
| Minutes shared with top 5-25% | | | 0.082 (0.089) | | | 0.026 (0.132) |
| Observations | 5,227 | 5,227 | 5,227 | 1,957 | 1,957 | 1,957 |
| R ² | 0.530 | 0.530 | 0.531 | 0.515 | 0.517 | 0.517 |
| Within R ² | 0.263 | 0.264 | 0.265 | 0.275 | 0.278 | 0.278 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

We can decompose peer quality's association with human capital growth via these two channels based on Table A12. Both minutes played and pass centrality is strongly negatively associated with peer quality. Interpreting the estimates causally, a 10% increase in peer quality would lead to 3.33% higher wage growth overall. While future wages would increase by 4.35% due to learning from peers on average, they would also decrease by 0.80% due to fewer total opportunities and by 0.22% due to a diminished overall importance to the team. Another interpretation is that if we could elevate peer quality by 10% while keeping workers' productive contribution constant, it would lead to a 30% higher growth in their human capital (4.35 vs. 3.33).

Looking at skill attributes as outcomes, we can form an even better understanding of how human capital accumulation manifests. It is a danger, as we mentioned earlier, that some of the previous results reflect not the growth in human capital but some selection bias due to unobserved growth potential. Therefore, finding a strong association between peer quality and the learning-relevant skills, and no association with other useful skills that are more 'innate' might eliminate some of

those fears. Table 5 supports that we might be close to estimating causal effects without the selection issue. While more learnable skills such as passing and reacting are significantly associated with peer quality, minutes, or pass centrality, less learnable skills such as sprinting speed or finishing ability have no significant associations with peer quality. The improvement in the most directly related skill attribute, short passing, is indeed the most sensitive to peer quality and pass centrality, as we would expect *ex ante*, while improvement in sprinting speed is not related significantly to any of the variables. Additionally, sharing more playing time with top players does not seem to lead to any improvements in skills. This is strong evidence that our estimates capture the learning channel rather than selection on unobservables.

Table 5: Skill evolution and role on the team

| | Short pass _{t+6} (1) | Reactions _{t+6} (2) | Interceptions _{t+6} (3) | Finishing _{t+6} (4) | Sprint speed _{t+6} (5) |
|-------------------------------|----------------------------------|---------------------------------|-------------------------------------|---------------------------------|------------------------------------|
| Own skill | 0.504*** (0.020) | 0.328*** (0.016) | 0.698*** (0.017) | 0.775*** (0.015) | 0.876*** (0.015) |
| Team avg. skill | 0.051 (0.039) | 0.190*** (0.047) | -0.054 (0.040) | -0.032 (0.038) | -0.054 (0.042) |
| log(value / wage) | 0.335*** (0.120) | 0.389*** (0.114) | 0.040 (0.169) | 0.336** (0.146) | -0.112 (0.150) |
| log(teammates' mean wage) | 1.52*** (0.283) | 1.08*** (0.302) | 0.755* (0.413) | 0.637 (0.385) | 0.039 (0.394) |
| Total minutes (Z) | 0.318** (0.160) | 1.51*** (0.163) | 0.632*** (0.223) | 0.449* (0.263) | 0.148 (0.227) |
| Minutes shared with top 5% | 0.275 (0.840) | 0.764 (0.817) | 0.221 (1.05) | 0.392 (1.17) | 1.53 (1.24) |
| Minutes shared with top 5-25% | 0.516 (0.526) | 0.159 (0.476) | 0.209 (0.671) | -1.06* (0.624) | -0.198 (0.685) |
| Pass centrality (Z) | 1.10*** (0.174) | 0.257* (0.154) | 0.339 (0.207) | 0.254 (0.246) | 0.316 (0.240) |
| Observations | 5,227 | 5,227 | 5,227 | 5,227 | 5,227 |
| R ² | 0.630 | 0.461 | 0.871 | 0.851 | 0.702 |
| Within R ² | 0.426 | 0.280 | 0.550 | 0.599 | 0.582 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

5.3 Direct collaboration with top peers

Our granular data allow us to leverage the information on the passing network to study the strength of the link between any two players. Passing captures the most intensive direct collaboration between football players as those who pass the most to each other should practice together, communicate the most with each other, providing a fertile ground for learning and teaching. We are going to differentiate in the amount of collaboration between three types of players: top 5%, top 5-25%, and the rest. We expect that direct collaboration with top talent would lead to higher human capital

growth reflected in higher wage, and skill attributes. We restrict the sample to teams containing all player types in their lineup, reducing our sample to around one-third of the original size.

Table 6 reports the regressions for this subset of observations with top talent in the lineup. Looking at Column 1, these individuals are somewhat different from the full sample as for them peer quality is more strongly associated with wage growth, while centrality in the team's passing network is not important. When we include direct interactions in the regression (Column 2), the estimate on peer quality is mostly unchanged. Conditionally, one standard deviation more passing with the top players is associated with an around 10% higher wage growth, while passing with regular players does not seem to be related. However, as Column 3 suggests, minutes played on these top teams capture all the important variation regarding direct interactions with top players. The results for younger players suggest similar conclusions, however, the most important direct collaboration happens not with superstars but the very good (top 5-25%) players.

Table 6: Wage evolution and direct collaboration with top peers

| | log(wage _{t+6}) | | | | | |
|---------------------------------|---------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | (1) | Full sample (2) | (3) | (4) | Junior (5) | (6) |
| log(wage) | 0.285*** (0.035) | 0.313*** (0.035) | 0.283*** (0.035) | 0.225*** (0.044) | 0.255*** (0.045) | 0.225*** (0.044) |
| log(value / wage) | 0.202*** (0.031) | 0.239*** (0.032) | 0.201*** (0.031) | 0.176*** (0.040) | 0.216*** (0.041) | 0.177*** (0.041) |
| log(teammates' mean wage) | 0.444*** (0.065) | 0.437*** (0.071) | 0.430*** (0.067) | 0.495*** (0.118) | 0.465*** (0.116) | 0.506*** (0.121) |
| Pass centrality (Z) | -0.041 (0.033) | | -0.044 (0.039) | 0.057 (0.062) | | 0.043 (0.064) |
| Total minutes (Z) | 0.280*** (0.034) | | 0.267*** (0.035) | 0.245*** (0.062) | | 0.239*** (0.064) |
| Pass count with top 5% (Z) | | 0.098*** (0.033) | 0.039 (0.030) | | 0.053 (0.067) | -0.019 (0.057) |
| Pass count with top 5-25% (Z) | | 0.113*** (0.034) | 0.012 (0.040) | | 0.144** (0.068) | 0.011 (0.056) |
| Pass count with top 25-100% (Z) | | 0.010 (0.033) | -0.025 (0.032) | | 0.103* (0.054) | 0.047 (0.050) |
| Observations | 1,644 | 1,644 | 1,644 | 653 | 653 | 653 |
| R ² | 0.471 | 0.449 | 0.472 | 0.427 | 0.406 | 0.428 |
| Within R ² | 0.239 | 0.208 | 0.241 | 0.233 | 0.204 | 0.234 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

Direct collaboration should lead to higher growth in learnable skills such as short passing ability, while unrelated skills such as sprinting speed should not be affected. Table 7 reports these estimates, corroborating our previous results. When we measure human capital development by these skill attributes, direct interaction with the top players is significantly related to better short passing

and reactions, while the other less relevant skill attributes are not significantly related. Our results suggest that in the end, direct collaboration via passing with the best peers leads to better future skills that are relevant to passing, turning into higher wages as well.

Table 7: Skill evolution and direct collaboration with top peers

| | Short pass _{t+6} (1) | Reactions _{t+6} (2) | Interceptions _{t+6} (3) | Finishing _{t+6} (4) | Sprint speed _{t+6} (5) |
|---------------------------------|----------------------------------|---------------------------------|-------------------------------------|---------------------------------|------------------------------------|
| Own skill | 0.536*** (0.040) | 0.351*** (0.023) | 0.737*** (0.032) | 0.804*** (0.024) | 0.810*** (0.025) |
| Team avg. skill | 0.031 (0.067) | 0.241** (0.093) | -0.003 (0.061) | 0.070 (0.080) | -0.031 (0.094) |
| log(value / wage) | 0.314 (0.228) | 0.659*** (0.193) | -0.278 (0.268) | 0.134 (0.272) | -0.018 (0.275) |
| log(teammates' mean wage) | 1.01** (0.472) | 0.226 (0.767) | 0.981* (0.589) | 0.435 (0.556) | 0.633 (0.661) |
| Pass count with top 5% (Z) | 0.598*** (0.210) | 0.772*** (0.217) | 0.395* (0.235) | 0.194 (0.343) | 0.013 (0.290) |
| Pass count with top 5-25% (Z) | 0.631*** (0.205) | 0.728*** (0.223) | 0.255 (0.266) | 0.627* (0.355) | 0.608* (0.313) |
| Pass count with top 25-100% (Z) | 0.164 (0.157) | 0.257 (0.181) | 0.185 (0.254) | 0.207 (0.298) | -0.020 (0.307) |
| Observations | 1,644 | 1,644 | 1,644 | 1,644 | 1,644 |
| R ² | 0.628 | 0.422 | 0.893 | 0.864 | 0.698 |
| Within R ² | 0.414 | 0.247 | 0.608 | 0.628 | 0.515 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

5.4 Persistence of peer effects: human capital growth of switchers

Examining players leaving the organization soon after arriving there allows us to distinguish learning from a particular production complementarity at the initial team. These individuals did not match well enough to remain, however, they had the opportunity to learn and benefit from the experience somewhere else.

Table 8 displays the regression results with the specifications from Section 5.2, reporting separate estimates for those who left in Columns 3 and 4. Peer effects are strongest for those who remained with the team but those who left are not far behind. 10% higher peer quality in the first half-season is associated with around 5% higher wage growth for stayers on average, while for those who left before the third year it is around 4%.

Table 8: Wage evolution of stayers vs. switchers after one and two years

| | log(wage _{t+6}) | | | |
|---------------------------|---------------------------|---------------------------|--------------------------|---------------------------|
| | Full sample (1) | Stayed for 3 years (2) | Left after 1 year (3) | Left after 2 years (4) |
| log(wage) | 0.343*** (0.021) | 0.309*** (0.028) | 0.378*** (0.032) | 0.307*** (0.047) |
| log(value / wage) | 0.179*** (0.018) | 0.149*** (0.022) | 0.207*** (0.032) | 0.177*** (0.039) |
| log(teammates' mean wage) | 0.435*** (0.034) | 0.499*** (0.047) | 0.414*** (0.048) | 0.438*** (0.066) |
| Total minutes (Z) | 0.229*** (0.024) | 0.146*** (0.032) | 0.313*** (0.048) | 0.196*** (0.050) |
| Pass centrality (Z) | 0.043* (0.022) | -0.005 (0.033) | 0.065 (0.051) | 0.057 (0.047) |
| Observations | 5,227 | 2,081 | 1,742 | 1,404 |
| R ² | 0.530 | 0.629 | 0.487 | 0.521 |
| Within R ² | 0.264 | 0.274 | 0.293 | 0.224 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

Skill evolution for switchers tells a similar story. Table 9 shows the results for short passing, for all examined skill attributes see Table A14. The estimates regarding peer quality are within one standard error of each other. The results suggest that in football, learning occurs mostly in the first period of the tenure, and the experience gathered then remains a persistent factor in the growth of human capital even for team switchers. Additionally, it helps us against the idea that selection bias drives our previous results: even in those cases where long-run prospects turn out to be worse and the match between player and team is suboptimal, the quality of the original team leaves a positive mark in the overall growth of the player.

Table 9: Short-pass evolution of stayers vs. switchers after one and two years

| | Short pass _{t+6} | | | |
|----------------------------|---------------------------|---------------------------|--------------------------|---------------------------|
| | Full sample (1) | Stayed for 3 years (2) | Left after 1 year (3) | Left after 2 years (4) |
| Short pass | 0.504*** (0.020) | 0.527*** (0.035) | 0.479*** (0.027) | 0.491*** (0.026) |
| Teammates' avg. short pass | 0.053 (0.039) | 0.089 (0.061) | 0.082 (0.060) | -0.019 (0.077) |
| log(value / wage) | 0.340*** (0.119) | 0.157 (0.202) | 0.531*** (0.147) | 0.348* (0.178) |
| log(teammates' mean wage) | 1.67*** (0.211) | 1.58*** (0.408) | 1.77*** (0.338) | 1.43*** (0.417) |
| Total minutes (Z) | 0.324** (0.160) | 0.112 (0.240) | 0.537** (0.262) | 0.119 (0.316) |
| Pass centrality (Z) | 1.09*** (0.173) | 0.878*** (0.255) | 1.15*** (0.247) | 1.34*** (0.297) |
| Observations | 5,227 | 2,081 | 1,742 | 1,404 |
| R ² | 0.630 | 0.651 | 0.632 | 0.638 |
| Within R ² | 0.426 | 0.427 | 0.425 | 0.422 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

6 Robustness

The main identification threat is that better teams might be more successful at uncovering 'hidden gems': players with great unobserved growth potential. They could be detected only by teams with more advanced scouting resources – which might be the teams with better peer quality. While this is a common issue in peer effect studies, our setup is less vulnerable as we observe the perceived market value incorporating growth potential. As we use market value measured just after the player arrives to a new team, the growth potential might be appropriately captured already. Furthermore, we have demonstrated the persistence of peer effects for switchers, and how they relate only to learnable skills.

Still, we perform five additional robustness checks to bound the problem and corroborate our results. First, we compare peer effects for players from countries with a top league vs. the rest of the world. If selection bias is related strongly to differential scouting success, we expect heterogeneity in this regard. Second, we study the differences by playing positions: defenders, midfielders, and forwards require different sets of abilities to succeed, of which some can be easier to scout than others. Third, we look at somewhat 'overcontrolled' regressions with one-halfseason forward-looking future value growth (capturing any sudden previously unrealized gains in market value) and team success as controls, to show that even with those variables in the regression potentially inducing

downward bias in the estimates, the core results are qualitatively similar. Fourth, we reestimate the main regressions for elite and non-elite teams. Fifth, we split the sample by player wage tertiles to see whether the relationship crucially depends on some parts of the distribution.

The estimates can be found in the Appendix, corroborating robustness along these dimensions. We find no large differences between the top 7 countries and the rest of the world (Table A15). The regressions by positions (Tables A16, A17) show that the point estimates for the three non-goalscorer positions are close to the pooled sample, however, for goalscorers we do not find any significant partial correlation between wage growth and peer quality. If there was a strong selection along unobserved growth potential, we expect that it would affect goalscorers as well, strengthening the validity of our results. The regressions with team success and one-period forward-looking future market value change (Table A18) suggest that peer effects are bounded by 0.3 from below, the remaining relationship is still large and statistically significant. Peer effects are similar and strong both within elite and non-elite teams (Table A19), and along the player’s wage distribution as well (Table A20).

7 Conclusions

This paper investigates the mechanisms through which workers learn, using elite male European football to study the impact of peer interactions on skill development in a high-skilled and competitive environment. We highlighted the significance of exposure to high-quality peers and the role of collaboration in improving human capital accumulation.

First, we showed that wage as an outcome might suffer from important limitations as a measure of human capital. Due to the lack of frequent updates and the complexity of a wage bargaining they might not adequately reflect human capital growth. We used external information on market value to reduce these concerns. Second, players do not benefit much just from the proximity of top coworkers and that a high average quality with very good performers drives human capital growth. We show that more work responsibilities, active engagement with the team, and direct collaboration relate strongly to higher human capital accumulation measured both by wages and skill attributes. We also found that for switchers, the skills acquired earlier persist in their future careers. This last point supports policies of sending people to learn from the best early on even if eventually they will move to lower-ranked jobs.

Overall, the findings of this study showcase the critical role of peer interactions and collaboration

in fostering skill development and enhancing individual performance. While our setting is special, it plausibly represents high-skilled, competitive, and at the same time, collaborative professional job environments, evidenced by the similarities to the top decile earning group of [Jarosch et al. \(2021\)](#). By highlighting the mechanisms through which workers learn from their peers, we provide valuable insights to understand the dynamics of skill acquisition and human capital accumulation. When designing job training programs, or selecting the team members at a firm or a school, policy makers and corporate leaders can consider these findings on the relationship between peer quality, interactions, and responsibilities to improve human capital growth at their organizations.

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. (2019). Who Becomes an Inventor in America? the Importance of Exposure to Innovation. *The Quarterly Journal of Economics*, 134(2):647–713.
- Bilen, E. and Matros, A. (2023). The queen’s gambit: Explaining the superstar effect using evidence from chess. *Journal of Economic Behavior & Organization*, 215:307–324.
- 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. (2025). Cultural homophily and collaboration in superstar teams. *Management Science*.
- Caicedo, S. and Lucas Jr, R. E. (2019). Learning, Career Paths, and the Distribution of Wages. *American Economic Journal: Macroeconomics*, 11(1):49–88.

- 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.
- Cohen-Zada, D., Dayag, I., and Gershoni, N. (2024). Effort peer effects in team production: Evidence from professional football. *Management Science*, 70(4):2355–2381.
- Cornelissen, T., Dustmann, C., and Schönberg, U. (2017). Peer Effects in the Workplace. *American Economic Review*, 107(2):425–456.
- Fonti, F., Ross, J.-M., and Aversa, P. (2023). Using sports data to advance management research: A review and a guide for future studies. *Journal of Management*, 49(1):325–362.
- 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.
- Herkenhoff, K., Lise, J., Menzio, G., and Phillips, G. M. (2024). Production and learning in teams. *Econometrica*, 92(2):467–504.
- Hoegel, 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.
- Hong, L. (2022). Coworker sorting, learning, and inequality. Technical report, University of Wisconsin - Madison.
- Jarosch, G., Oberfield, E., and Rossi-Hansberg, E. (2021). Learning From Coworkers. *Econometrica*, 89(2):647–676.
- 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.

- Pastorino, E. (2024). Careers in firms: The role of learning about ability and human capital acquisition. *Journal of Political Economy*, 132(6):1994–2073.
- 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.
- Winkler, W. E. (1990). String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. In *Proceedings of the Section on Survey Research Methods*, pages 354–359, Alexandria, VA. American Statistical Association.

A Appendix

A.1 Variable descriptions

Our key datasets are the following:

- Player careers: players and team in every half-season with relevant demographics
- Player valuations: TM market valuation for a player in every half-season. (Average when multiple values are available.)
- Player wages and skill ratings: EA Sports' FIFA games include wage information and in addition the scores of rating points of players describing offensive and defensive skill levels such as short passes, tackles, sprinting speed etc.
- Players and squads: information of squads for every match for each team in every half-season (list of players who were ever listed in the lineup as starter or substitute.), providing time spent together, injuries etc.
- Events: number of passes between any two players

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' mean wage**: the mean of the wages of the teammates' of the player based on the data registered in EA Sport's FIFA in the half-season, measured in log-units
- **Teammates' mean market value**: the mean of the market value of the teammates' of the player based on their mean market value in the half-season, measured in log-units
- **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:** within the position, the player is or in the last 2 years was in the top 5% of the wage distribution in the top 7 leagues
- **Top 25% of players:** within the position, the player is or in the last 2 years was in the top 25% of the wage distribution in the top 7 leagues
- **Seniors vs. juniors:** players are split by age at 23 years of age
- **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
- **Hirschman–Herfindahl Index:** HHI index of the wages of the squad members in the half-season
- **Share of top 5% or 25% players on team:** share in terms of number of individuals based on the wage distribution

Fixed-effects in the regressions:

- **League \times half-seasons:** league indicators interacted with the half-seasons
- **Position:** broad position of the player such as goalkeeper, defender, midfielder, forward
- **Player age:** in years
- **Elite team:** manually identified a select group of elite teams that are the traditional top teams in their respective leagues and regularly participate in international tournaments. These teams are the following: Real Madrid, Barcelona, Atletico Madrid, Manchester City, Manchester United, Liverpool, Chelsea, Tottenham, Arsenal, Bayern München, Dortmund, Schalke, Juventus, AC Milan, Inter Milan, Paris Saint-Germain, Olympique Lyonnais. The selection was done on the basis of historic performance, close to Deloitte’s top 20 (an annual report).
- **Injury:** the player suffering an injury, identified by not listed as available on the squad

- **Spell count:** the count of teams the player has played on
- **Contract maturity:** years left on the contract

Skill attribute rating variables: we select from a set of highly correlated variables one that is representative of the overall skill group.

- **Short pass:** short pass ability, proxies passing ability
- **Finishing:** scoring ability, proxies shooting
- **Interceptions:** defensive attribute, ability to intercept a ball
- **Reactions:** general attribute, how fast a player responds to a situation around him
- **Sprinting speed:** how fast a player can sprint

A.2 Descriptive statistics

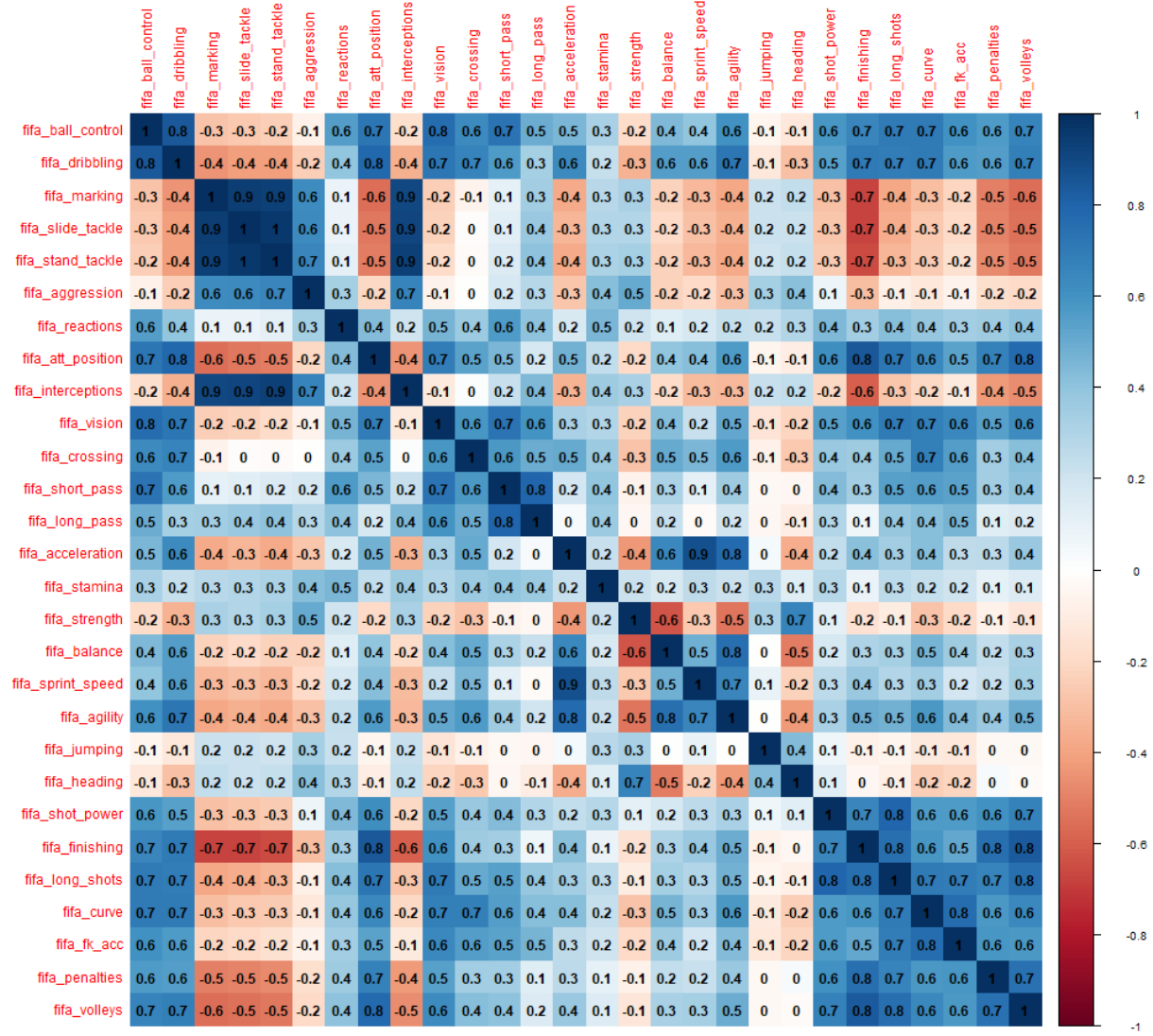
Table A10: Descriptive statistics of the key numeric variables

| | N | Mean | SD | Min | P1 | P50 | P99 | Max |
|---|------|---------|--------|-------|-------|---------|---------|---------|
| Yearly wage (in mn EUR, EA-FIFA) | 5227 | 1.43 | 1.63 | 0.08 | 0.08 | 0.96 | 8.06 | 20.80 |
| log(yearly wage) | 5227 | 13.61 | 1.12 | 11.26 | 11.26 | 13.78 | 15.90 | 16.85 |
| TM market value (in mn EUR) | 5227 | 4.46 | 7.48 | 0.03 | 0.10 | 2.00 | 35.00 | 150.00 |
| log(TM market value) | 5227 | 14.49 | 1.32 | 10.13 | 11.51 | 14.51 | 17.37 | 18.83 |
| log(market value/wage) | 5227 | 4.83 | 0.80 | 1.39 | 2.88 | 4.83 | 6.91 | 8.70 |
| Short pass | 5227 | 67.92 | 8.62 | 23.00 | 44.00 | 68.00 | 85.00 | 91.00 |
| Reactions | 5227 | 67.82 | 7.98 | 35.00 | 48.00 | 68.00 | 84.00 | 96.00 |
| Interceptions | 5227 | 53.16 | 20.11 | 10.00 | 14.00 | 60.00 | 84.00 | 91.00 |
| Finishing | 5227 | 54.27 | 17.39 | 11.00 | 19.00 | 58.00 | 84.00 | 94.00 |
| Speed | 5227 | 71.75 | 10.25 | 31.00 | 42.00 | 73.00 | 92.00 | 96.00 |
| Teammates' mean wage (in mn EUR) | 5227 | 1.55 | 1.43 | 0.10 | 0.14 | 1.18 | 7.09 | 10.98 |
| log(teammates' mean wage) | 5227 | 9.98 | 0.82 | 7.60 | 7.92 | 10.03 | 11.82 | 12.26 |
| Teammates' mean value (in mn EUR) | 5227 | 4.48 | 5.72 | 0.24 | 0.33 | 2.42 | 31.49 | 52.55 |
| log(teammates' mean value) | 5227 | 14.80 | 0.99 | 12.38 | 12.71 | 14.70 | 17.27 | 17.78 |
| HHI of team (wage) | 5227 | 0.06 | 0.01 | 0.04 | 0.04 | 0.05 | 0.11 | 0.27 |
| Share of top 5% players in team (wage) | 5227 | 0.07 | 0.17 | 0.00 | 0.00 | 0.00 | 0.78 | 0.96 |
| Share of top 5-25% players in team (wage) | 5227 | 0.27 | 0.22 | 0.00 | 0.00 | 0.23 | 0.79 | 0.95 |
| Total minutes | 5227 | 1076.55 | 649.03 | 1.00 | 13.00 | 1054.00 | 2465.96 | 3520.00 |
| Pass eigenvector centrality | 5227 | 0.44 | 0.29 | 0.00 | 0.00 | 0.43 | 1.00 | 1.00 |
| Pass count top 5% (wage) | 1718 | 43.59 | 48.00 | 0.00 | 0.00 | 29.37 | 225.04 | 425.00 |
| Pass count top 5-25% (wage) | 4516 | 38.16 | 45.00 | 0.00 | 0.00 | 26.20 | 199.73 | 1298.00 |
| Pass count top 25-100% (wage) | 5162 | 25.60 | 24.10 | 0.00 | 0.00 | 19.57 | 109.79 | 232.00 |
| Minutes shared with top 5% (wage) | 5227 | 0.09 | 0.21 | 0.00 | 0.00 | 0.00 | 0.91 | 1.00 |
| Minutes shared with top 5-25% (wage) | 5227 | 0.31 | 0.26 | 0.00 | 0.00 | 0.27 | 0.88 | 1.00 |
| Player age | 5227 | 23.97 | 3.52 | 18.00 | 18.00 | 24.00 | 32.00 | 37.00 |
| Elite team | 5227 | 0.10 | 0.31 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 |
| Injury | 5227 | 0.06 | 0.24 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 |
| Spell rank | 5227 | 1.78 | 1.00 | 1.00 | 1.00 | 1.00 | 5.00 | 8.00 |
| Contract maturity | 5227 | 3.06 | 1.32 | 0.00 | 1.00 | 3.00 | 6.00 | 7.00 |

Note: All players, all teams. Selected variables.

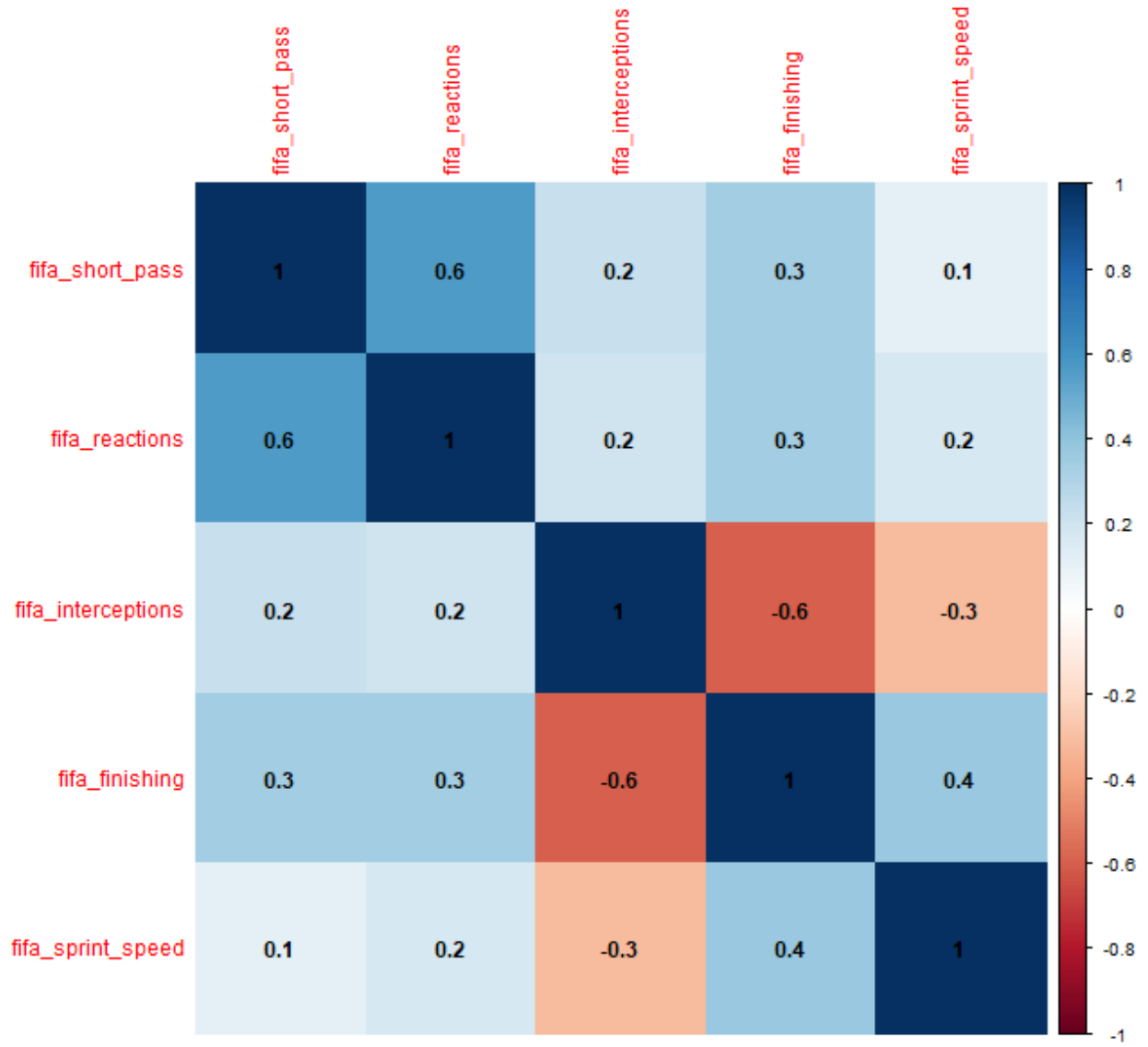
The correlation matrix of the FIFA skill ratings in the sample is the following:

Figure A2: Correlation between EA FIFA skill ratings



Note: The figure shows the correlation structure of FIFA attributes for the final sample.

Figure A3: Correlation between the selected EA Sport's FIFA skill ratings



Note: The figure shows the Pearson correlations of the selected FIFA attributes for the final sample.

A.3 Additional results

A.3.1 Fixed effects

Table A11: Fixed effects and explained variation in peer quality

| | log(teammates' mean wage) | | | |
|------------------------------------|---------------------------|-------|-------|-------|
| | (1) | (2) | (3) | (4) |
| Observations | 5,227 | 5,227 | 5,227 | 5,227 |
| R ² | 0.500 | 0.719 | 0.909 | 0.946 |
| league X half-season fixed effects | ✓ | ✓ | ✓ | ✓ |
| position fixed effects | | ✓ | | |
| player age fixed effects | | ✓ | | |
| elite team fixed effects | | ✓ | | |
| spell count fixed effects | | ✓ | | |
| own injury fixed effects | | ✓ | | |
| contract maturity fixed effects | | ✓ | | |
| managename fixed effects | | | ✓ | |
| teamid_tm fixed effects | | | | ✓ |

A.3.2 Contribution to production

Table A12: Minutes played and centrality in the passing network vs. peer effects in wages

| | log(wage _{t+6}) | | Total minutes (Z) | Pass centrality (Z) |
|---------------------------|---------------------------|---------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (4) |
| log(wage) | 0.440*** (0.021) | 0.343*** (0.021) | 0.361*** (0.021) | 0.333*** (0.023) |
| log(value / wage) | 0.241*** (0.019) | 0.179*** (0.018) | 0.235*** (0.017) | 0.188*** (0.016) |
| log(teammates' mean wage) | 0.333*** (0.033) | 0.435*** (0.034) | -0.353*** (0.031) | -0.495*** (0.032) |
| Total minutes (Z) | | 0.229*** (0.024) | | |
| Pass centrality (Z) | | 0.043* (0.022) | | |
| Observations | 5,227 | 5,227 | 5,227 | 5,227 |
| R ² | 0.495 | 0.530 | 0.315 | 0.296 |
| Within R ² | 0.208 | 0.264 | 0.090 | 0.083 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

Decomposition of the total difference into the positive effect of learning and the negative of less opportunities:

$$\frac{\Delta \mathbb{E}[W_{t+6}|PQ, X]}{\Delta PQ} = 0.333 = \underbrace{\beta}_{0.435} + \underbrace{\gamma}_{0.229} \underbrace{\frac{\Delta \mathbb{E}[M|PQ, X]}{\Delta PQ}}_{-0.353} + \underbrace{\delta}_{0.043} \underbrace{\frac{\Delta \mathbb{E}[C|PQ, X]}{\Delta PQ}}_{-0.495}, \quad (2)$$

where W denotes wage, PQ denotes peer quality, X denotes control variables, M denotes minutes, and C denotes pass centrality.

A.3.3 Direct collaboration with top peers

Table A13: Short passing skill evolution, and direct collaboration with top peers by age

| | Short pass _{t+6} | | | | | |
|---------------------------------|---------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | (1) | Full sample (2) | (3) | (4) | Junior (5) | (6) |
| Short pass | 0.530*** (0.042) | 0.536*** (0.040) | 0.528*** (0.041) | 0.320*** (0.030) | 0.334*** (0.030) | 0.320*** (0.030) |
| log(value / wage) | 0.178 (0.224) | 0.314 (0.228) | 0.181 (0.224) | 0.669*** (0.208) | 0.800*** (0.214) | 0.675*** (0.211) |
| log(teammates' mean wage) | 0.937** (0.426) | 1.01** (0.472) | 0.896* (0.454) | 2.32*** (0.669) | 2.14*** (0.699) | 2.32*** (0.683) |
| Teammates' avg. short pass | 0.072 (0.067) | 0.031 (0.067) | 0.065 (0.068) | 0.119 (0.096) | 0.083 (0.097) | 0.117 (0.096) |
| Pass centrality (Z) | 0.653** (0.273) | | 0.568** (0.269) | 1.19** (0.466) | | 1.09** (0.475) |
| Total minutes (Z) | 0.742*** (0.243) | | 0.597** (0.260) | 0.504 (0.444) | | 0.396 (0.475) |
| Pass count with top 5% (Z) | | 0.598*** (0.210) | 0.322* (0.183) | | 0.451 (0.392) | 0.093 (0.333) |
| Pass count with top 5-25% (Z) | | 0.631*** (0.205) | 0.140 (0.229) | | 0.822** (0.394) | 0.195 (0.391) |
| Pass count with top 25-100% (Z) | | 0.164 (0.157) | -0.074 (0.154) | | 0.400 (0.318) | 0.052 (0.324) |
| Observations | 1,644 | 1,644 | 1,644 | 653 | 653 | 653 |
| R ² | 0.635 | 0.628 | 0.636 | 0.601 | 0.590 | 0.601 |
| Within R ² | 0.425 | 0.414 | 0.427 | 0.329 | 0.310 | 0.329 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

A.3.4 Persistence

Table A14: Skill evolution of stayers vs. switchers after one and two years

| | Short pass _{t+6} (1) | Reactions _{t+6} (2) | Interceptions _{t+6} (3) | Finishing _{t+6} (4) | Sprint speed _{t+6} (5) |
|--|----------------------------------|---------------------------------|-------------------------------------|---------------------------------|------------------------------------|
| Own skill | 0.503*** (0.020) | 0.330*** (0.016) | 0.697*** (0.017) | 0.777*** (0.015) | 0.876*** (0.015) |
| Team avg. skill | 0.050 (0.039) | 0.190*** (0.043) | -0.054 (0.040) | -0.029 (0.038) | -0.056 (0.041) |
| log(value / wage) | 0.331*** (0.118) | 0.376*** (0.112) | 0.025 (0.168) | 0.322** (0.145) | -0.111 (0.151) |
| log(teammates' mean wage) | 1.72*** (0.254) | 1.53*** (0.313) | 0.968*** (0.326) | 0.344 (0.296) | 0.362 (0.314) |
| Switch in year 1 | -1.02 (1.86) | 2.78 (2.02) | -0.117 (3.51) | -3.64 (2.67) | 2.55 (3.02) |
| Switch in year 2 | 2.30 (2.00) | 7.69*** (2.04) | 5.27* (2.68) | 3.67 (3.07) | 3.23 (2.90) |
| Total minutes (Z) | 0.216 (0.156) | 1.38*** (0.159) | 0.463** (0.219) | 0.384 (0.259) | 0.112 (0.225) |
| Pass centrality (Z) | 1.11*** (0.171) | 0.274* (0.152) | 0.377* (0.203) | 0.283 (0.250) | 0.314 (0.239) |
| log(teammates' mean wage) × Switch in year 1 | -0.022 (0.186) | -0.420** (0.201) | -0.187 (0.347) | 0.283 (0.271) | -0.300 (0.308) |
| log(teammates' mean wage) × Switch in year 2 | -0.372* (0.198) | -0.921*** (0.201) | -0.704*** (0.264) | -0.413 (0.302) | -0.367 (0.281) |
| Observations | 5,227 | 5,227 | 5,227 | 5,227 | 5,227 |
| R ² | 0.637 | 0.474 | 0.873 | 0.852 | 0.702 |
| Within R ² | 0.438 | 0.298 | 0.557 | 0.600 | 0.583 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

A.4 Robustness checks

A.4.1 Results by player nationality

Table A15: Wage evolution of players from the Top-7 countries vs. the others

| | log(wage _{t+6}) | | | | | |
|---------------------------------|---------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | top 7 countries | | | Other countries | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| log(wage) | 0.447*** (0.026) | 0.342*** (0.027) | 0.321*** (0.049) | 0.459*** (0.034) | 0.371*** (0.033) | 0.267*** (0.058) |
| log(value / wage) | 0.249*** (0.023) | 0.186*** (0.022) | 0.199*** (0.043) | 0.253*** (0.032) | 0.187*** (0.030) | 0.231*** (0.052) |
| log(teammates' mean wage) | 0.353*** (0.039) | 0.455*** (0.042) | 0.343*** (0.091) | 0.302*** (0.055) | 0.400*** (0.053) | 0.522*** (0.112) |
| Total minutes (Z) | | 0.248*** (0.029) | 0.286*** (0.047) | | 0.197*** (0.036) | 0.258*** (0.060) |
| Pass centrality (Z) | | 0.013 (0.024) | -0.036 (0.046) | | 0.088** (0.035) | -0.056 (0.062) |
| Pass count with top 5% (Z) | | | 0.008 (0.039) | | | 0.072* (0.043) |
| Pass count with top 5-25% (Z) | | | -0.021 (0.041) | | | -0.0004 (0.056) |
| Pass count with top 25-100% (Z) | | | -0.051 (0.049) | | | 0.042 (0.038) |
| Observations | 3,161 | 3,161 | 939 | 2,066 | 2,066 | 705 |
| R ² | 0.503 | 0.535 | 0.491 | 0.506 | 0.544 | 0.505 |
| Within R ² | 0.208 | 0.260 | 0.231 | 0.215 | 0.276 | 0.260 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

A.4.2 Results by position

Table A16: Wage evolution of players by positions

| | log(wage _{t+6}) | | | | |
|---------------------------|---------------------------|---------------------|---------------------|---------------------|---------------------|
| | Non-GK | GK | D | M | F |
| | (1) | (2) | (3) | (4) | (5) |
| log(wage) | 0.440*** (0.021) | 0.582*** (0.068) | 0.417*** (0.040) | 0.434*** (0.036) | 0.465*** (0.040) |
| log(value / wage) | 0.241*** (0.019) | 0.149** (0.063) | 0.214*** (0.027) | 0.276*** (0.035) | 0.237*** (0.036) |
| log(teammates' mean wage) | 0.333*** (0.033) | 0.132 (0.112) | 0.395*** (0.064) | 0.267*** (0.048) | 0.328*** (0.056) |
| Observations | 5,227 | 397 | 1,861 | 1,650 | 1,716 |
| R ² | 0.495 | 0.621 | 0.513 | 0.532 | 0.478 |
| Within R ² | 0.208 | 0.267 | 0.213 | 0.205 | 0.208 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

Table A17: Wage evolution by position

| | Defender | | log(wage _{t+6}) Midfielder | | Forward | |
|-------------------------------|---------------------|---------------------|---|---------------------|---------------------|---------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| log(wage) | 0.417*** (0.040) | 0.329*** (0.042) | 0.434*** (0.036) | 0.351*** (0.037) | 0.465*** (0.040) | 0.340*** (0.035) |
| log(value / wage) | 0.214*** (0.027) | 0.161*** (0.027) | 0.276*** (0.035) | 0.214*** (0.034) | 0.237*** (0.036) | 0.157*** (0.036) |
| log(teammates' mean wage) | 0.395*** (0.064) | 0.475*** (0.098) | 0.267*** (0.048) | 0.251*** (0.074) | 0.328*** (0.056) | 0.434*** (0.084) |
| Minutes shared with top 5% | | 0.149 (0.264) | | 0.346 (0.215) | | 0.300 (0.254) |
| Minutes shared with top 5-25% | | -0.022 (0.151) | | 0.277* (0.147) | | -0.028 (0.146) |
| Total minutes (Z) | | 0.172*** (0.036) | | 0.252*** (0.052) | | 0.228*** (0.046) |
| Pass centrality (Z) | | 0.075* (0.040) | | -0.005 (0.046) | | 0.111** (0.043) |
| Observations | 1,861 | 1,861 | 1,650 | 1,650 | 1,716 | 1,716 |
| R ² | 0.513 | 0.544 | 0.532 | 0.562 | 0.478 | 0.526 |
| Within R ² | 0.213 | 0.263 | 0.205 | 0.257 | 0.208 | 0.280 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

A.4.3 Overcontrolling

Table A18: Wage evolution with overcontrolled regressions

| | (1) | (2) | (3) | log(wage _{t+6}) | | (6) | (7) |
|------------------------------|---------------------|---------------------|---------------------|---------------------------|---------------------|---------------------|---------------------|
| | | | | (4) | (5) | | |
| log(wage) | 0.440*** (0.021) | 0.422*** (0.021) | 0.478*** (0.020) | 0.343*** (0.021) | 0.325*** (0.021) | 0.374*** (0.020) | 0.356*** (0.020) |
| log(value / wage) | 0.241*** (0.019) | 0.227*** (0.019) | 0.283*** (0.020) | 0.179*** (0.018) | 0.166*** (0.018) | 0.210*** (0.020) | 0.198*** (0.020) |
| log(teammates' mean wage) | 0.333*** (0.033) | 0.225*** (0.033) | 0.299*** (0.032) | 0.435*** (0.034) | 0.332*** (0.034) | 0.407*** (0.033) | 0.299*** (0.033) |
| log(team points) | | 0.366*** (0.036) | | | 0.368*** (0.037) | | 0.376*** (0.036) |
| Dlog(value) _{t+1,t} | | | 0.299*** (0.035) | | | 0.188*** (0.035) | 0.199*** (0.034) |
| Total minutes (Z) | | | | 0.229*** (0.024) | 0.202*** (0.023) | 0.212*** (0.024) | 0.184*** (0.023) |
| Pass centrality (Z) | | | | 0.043* (0.022) | 0.072*** (0.021) | 0.041* (0.022) | 0.071*** (0.021) |
| Observations | 5,227 | 5,227 | 5,217 | 5,227 | 5,227 | 5,217 | 5,217 |
| R ² | 0.495 | 0.505 | 0.504 | 0.530 | 0.541 | 0.534 | 0.545 |
| Within R ² | 0.208 | 0.225 | 0.222 | 0.264 | 0.281 | 0.269 | 0.286 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

A.4.4 Results by teams' elite status

Table A19: Estimates by teams' elite status

| | log(wage _{t+6}) | | | | | |
|---------------------------------|---------------------------|---------------------|---------------------|---------------------|------------------------|---------------------|
| | (1) | elite teams (2) | (3) | (4) | non-elite teams (5) | (6) |
| log(wage) | 0.353*** (0.054) | 0.287*** (0.051) | 0.308*** (0.054) | 0.442*** (0.023) | 0.343*** (0.022) | 0.279*** (0.044) |
| log(value / wage) | 0.197*** (0.054) | 0.115** (0.054) | 0.125* (0.065) | 0.240*** (0.020) | 0.179*** (0.019) | 0.209*** (0.035) |
| log(teammates' mean wage) | 0.410*** (0.114) | 0.525*** (0.112) | 0.493*** (0.142) | 0.325*** (0.037) | 0.425*** (0.038) | 0.393*** (0.094) |
| Total minutes (Z) | | 0.192*** (0.064) | 0.216*** (0.075) | | 0.235*** (0.027) | 0.300*** (0.045) |
| Pass centrality (Z) | | 0.061 (0.072) | -0.022 (0.077) | | 0.042* (0.025) | -0.061 (0.046) |
| Pass count with top 5% (Z) | | | 0.191*** (0.059) | | | -0.006 (0.032) |
| Pass count with top 5-25% (Z) | | | -0.057 (0.045) | | | 0.024 (0.042) |
| Pass count with top 25-100% (Z) | | | -0.015 (0.047) | | | -0.024 (0.042) |
| Observations | 546 | 546 | 481 | 4,681 | 4,681 | 1,163 |
| R ² | 0.491 | 0.535 | 0.544 | 0.427 | 0.468 | 0.411 |
| Within R ² | 0.184 | 0.254 | 0.279 | 0.202 | 0.259 | 0.227 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

A.4.5 Results by players' wage tertiles

Table A20: Estimates by players' wage tertiles

| | log(wage _{t+6}) | | | | | |
|---------------------------|---------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | 1st tertile (1) | (2) | 2nd tertile (3) | (4) | 3rd tertile (5) | (6) |
| log(wage) | 0.385*** (0.048) | 0.278*** (0.049) | 0.406*** (0.067) | 0.299*** (0.066) | 0.566*** (0.047) | 0.443*** (0.046) |
| log(value / wage) | 0.193*** (0.028) | 0.136*** (0.027) | 0.248*** (0.033) | 0.174*** (0.032) | 0.341*** (0.041) | 0.276*** (0.041) |
| log(teammates' mean wage) | 0.328*** (0.060) | 0.437*** (0.059) | 0.317*** (0.060) | 0.462*** (0.063) | 0.306*** (0.047) | 0.391*** (0.046) |
| Total minutes (Z) | | 0.177*** (0.049) | | 0.235*** (0.039) | | 0.237*** (0.034) |
| Pass centrality (Z) | | 0.131** (0.052) | | 0.063* (0.035) | | -0.019 (0.030) |
| Observations | 1,735 | 1,735 | 1,744 | 1,744 | 1,748 | 1,748 |
| R ² | 0.357 | 0.402 | 0.316 | 0.375 | 0.461 | 0.494 |
| Within R ² | 0.103 | 0.165 | 0.092 | 0.170 | 0.195 | 0.245 |

Note: Standard errors clustered at position x league x half-season level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Fixed effects include: league x halfseason, age, position, injury, elite team, spell count, contract maturity.

A.5 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. (638,776 rows remain)
- Keep only those players who have at least 11 players with a valid valuation on their team. (418,462 rows remain)
- Keep those that are:
 - in the relevant seasons of 2012/2013-2019/2020
 - Top 7 European football leagues
 - at least 18 years of age
 - data cleaning: has valid team id, has less than 50 players to interact with, and non-negative minutes
- 42,868 observations remain, out of which:
 - for each player keep the first valid half-season observation for each spell at their team where they spent at least two half-seasons (10,256 obs remain)
 - keep those with valid minutes or passing centrality information (10,081 obs. remain)
 - keep those with valid team information: points and manager known (9,886 obs. remain)
 - keep those in the 2013–2019 seasons that have 6 half-seasons of lead market value, skill, and wage observations (5,610 obs remain)
 - Exclude goalkeepers (5,227 obs. remain)
- Final sample contains 5,227 observations.

For matching the TransferMarket–WhoScored, the EA Sports’ FIFA observations, and the Capology observations, we relied on player name, team name (also league), and season information. First we linked on the exact matches, and then as player and team names might have differed in the different sources (special characters, third names, etc.), we used the ‘jarowinkler’ function of the

[RecordLinkage](#) package to clean the names and produce string similarity indices, based on Jaro and Winkler’s algorithm [Winkler \(1990\)](#). From the rest, we matched those where player name similarities were higher than 0.95, in case of ties the one with higher similarity in the team name was selected. For the rest, we only matched when both player name’s and team name’s maximal similarity referred to the same observation in each of the datasources. Others were discarded. Table [A21](#) reports the success rates by league and seniority of the player. In the EA Sport’s FIFA dataset almost all players could be found, while in the Capology data only around 80%.

Table A21: Data matching success rate by league and seniority

| league | player_type | N | in_fifa_id | in_cap_id |
|------------------------|-------------|------|------------|-----------|
| | | | mean | mean |
| england_premier-league | junior | 1051 | 1.00 | 0.79 |
| | senior | 5766 | 1.00 | 0.86 |
| france_li | junior | 1736 | 1.00 | 0.75 |
| | senior | 4867 | 1.00 | 0.81 |
| germany_bundesliga | junior | 1600 | 1.00 | 0.84 |
| | senior | 4735 | 1.00 | 0.88 |
| italy_serie-a | junior | 1546 | 1.00 | 0.80 |
| | senior | 5942 | 1.00 | 0.78 |
| netherlands_eredivisie | junior | 1986 | 1.00 | 0.77 |
| | senior | 3121 | 1.00 | 0.76 |
| portugal_liga-nos | junior | 587 | 1.00 | 0.78 |
| | senior | 2428 | 1.00 | 0.79 |
| spain_laliga | junior | 1172 | 1.00 | 0.85 |
| | senior | 5716 | 1.00 | 0.84 |

Note: Data matching success rates based on leagues and player types.

Capology salaries and EA-FIFA wages are very similar, with a regression slope of 0.98 and R-squared of 0.653. Column (2) reports on the EA-FIFA market values and the TransferMarkt market value’s relationship, which are a bit shifted (constant being quite large), but the slope still being around 0.86.

Table A22: Capology salaries vs. FIFA wages, TM market values vs. FIFA market values

| | log(Capology salary) (1) | log(TM market value) (2) |
|-----------------------|-----------------------------|-----------------------------|
| Constant | -0.780*** (0.083) | 2.03*** (0.165) |
| log(wage in EA FIFA) | 0.976*** (0.008) | |
| log(value in EA FIFA) | | 0.855*** (0.011) |
| Observations | 29,305 | 42,214 |
| R ² | 0.653 | 0.671 |

Note: Standard errors clustered at the player level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Finally, Table A23 compares the overall sample with relevant seasons, leagues, and variables available to the final analysis sample. The final sample, due to focusing on younger, less established players, is slightly lower in market value and salary, and play with teammates on average lower value.

Two-sided t-tests between the two samples are as follows:

Table A23: Observable characteristics and selection into final sample

| | base (N=37026) | | final sample (N=5227) | | Diff. in Means | Std. Error |
|------------------------------------|----------------|-----------|-----------------------|-----------|----------------|------------|
| | Mean | Std. Dev. | Mean | Std. Dev. | | |
| TM market value (in mn EUR) | 6.8 | 12.1 | 4.5 | 7.5 | -2.4*** | 0.1 |
| log(TM market value) | 14.8 | 1.4 | 14.5 | 1.3 | -0.3*** | 0.0 |
| Yearly wage (in mn EUR, EA-FIFA) | 1.8 | 2.2 | 1.4 | 1.6 | -0.4*** | 0.0 |
| log(yearly wage) | 13.8 | 1.1 | 13.6 | 1.1 | -0.2*** | 0.0 |
| Short pass | 66.8 | 14.2 | 67.9 | 8.6 | 1.1*** | 0.1 |
| Reactions | 70.7 | 7.9 | 67.8 | 8.0 | -2.9*** | 0.1 |
| Interceptions | 54.6 | 21.9 | 53.2 | 20.1 | -1.4*** | 0.3 |
| Finishing | 51.7 | 20.0 | 54.3 | 17.4 | 2.6*** | 0.3 |
| Speed | 68.1 | 13.5 | 71.8 | 10.2 | 3.6*** | 0.2 |
| Teammates' mean value (in mn EUR) | 6.3 | 8.0 | 4.5 | 5.7 | -1.8*** | 0.1 |
| log(teammates' mean value) | 15.1 | 1.1 | 14.8 | 1.0 | -0.3*** | 0.0 |
| Teammates' mean wage (in mn EUR) | 1.7 | 1.6 | 1.6 | 1.4 | -0.1*** | 0.0 |
| log(teammates' mean wage) | 10.0 | 0.9 | 10.0 | 0.8 | 0.0** | 0.0 |
| Team HHI of market values | 0.1 | 0.0 | 0.1 | 0.0 | 0.0*** | 0.0 |
| Team HHI of wages | 0.1 | 0.0 | 0.1 | 0.0 | 0.0*** | 0.0 |
| Share of top 5% players in team | 0.1 | 0.2 | 0.1 | 0.2 | 0.0*** | 0.0 |
| Share of top 5-25% players in team | 0.3 | 0.2 | 0.3 | 0.2 | 0.0*** | 0.0 |
| Total minutes | 879.6 | 629.4 | 1076.6 | 649.0 | 197.0*** | 9.6 |
| Pass eigenvector centrality | 0.4 | 0.3 | 0.4 | 0.3 | 0.0*** | 0.0 |
| Pass count top 5% | 44.6 | 50.8 | 48.6 | 51.9 | 4.0** | 1.4 |
| Pass count top 5-25% | 31.0 | 33.0 | 37.7 | 38.7 | 6.7*** | 0.6 |
| Pass count top 25-100% | 17.1 | 19.1 | 24.0 | 23.0 | 6.9*** | 0.3 |
| Minutes shared with top 5% | 0.1 | 0.2 | 0.1 | 0.2 | 0.0*** | 0.0 |
| Minutes shared with top 5-25% | 0.4 | 0.3 | 0.3 | 0.3 | 0.0*** | 0.0 |
| Player age | 26.3 | 4.2 | 24.0 | 3.5 | -2.3*** | 0.1 |
| Elite team | 0.1 | 0.4 | 0.1 | 0.3 | 0.0*** | 0.0 |
| Injury | 0.1 | 0.3 | 0.1 | 0.2 | 0.0*** | 0.0 |
| Spell rank | 1.6 | 1.0 | 1.8 | 1.0 | 0.2*** | 0.0 |
| Contract maturity | 2.6 | 1.3 | 3.1 | 1.3 | 0.4*** | 0.0 |

Note: The table displays the differences between the initial relevant panel dataset and the final analysis sample in terms of observables.

A.6 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
- 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:

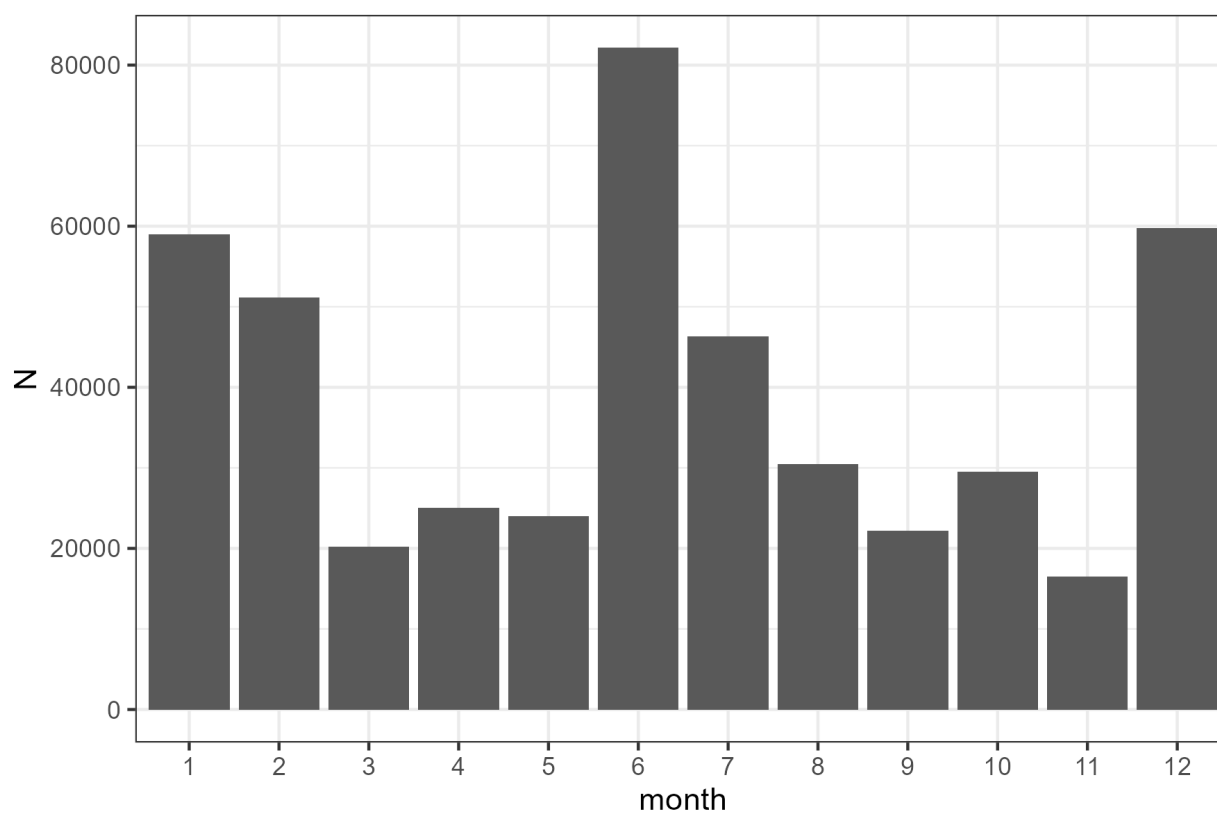
- Transfers via an option to buy/obligation to buy
- Loan fee
- Only part of transfer rights acquired
- Exit clause
- Buyback option
- Player swap deal
- Contract length
- Resale participation
- Bonus payments
- Improvement of financial balance

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

The frequency of updates in the market value is displayed below, showing that it is most prevalent in the Winter and in the Summer:

Figure A4: Frequency of market value updates



Note: The figure shows the number of market value evaluations per month in the TransferMarkt original data.