Can Soccer Scouts Identify Talent Better Than Laypeople?

Ellen Stroemer

Master Thesis – Talent Development and Creativity

S4228790
April 2024
Department of Psychology
University of Groningen
Examiner/Daily supervisor:
Prof. Rob Meijer, Ilse Peringa, MSc

Abstract

In this study, the validity of the performance predictions of eighteen soccer scouts and 50 laypeople in soccer was assessed and compared. The application of the lens model as a theoretical framework allowed for various insights into the decision-making process of the judges. The judgment task entailed 50 real player profiles that consisted of four informational cues (e.g., tackling, interceptions, sprinting speed and game insight). The judges were asked to consider these scores to make predictions about the player's worth. The actual market value of the players served as the criterion that the predictions were compared to. The analysis showed that the scouts did not make more accurate predictions than the laypeople, which provides evidence against the expert's eye for talent. Furthermore, nearly all judges were outperformed by their own model, which administered the average weights that the judges assigned to the four cues. This finding confirmed that intuitive judgments are impacted by noise. That implies that the judges tend to weigh the cues inconsistently across cases, which reduces their validity. Therefore, applying a decision rule to combine information seems to be beneficial as it removes this noise. Moreover, the scouts were not able to capture valid insights that a model approach could not, conveying that the employment of a decision rule does not harm the validity of the judgment. Lastly, recommendations on effective and autonomy-preserving approaches to talent assessment are provided.

Keywords: expert's eye, performance prediction, decision-making, soccer

Can Soccer Scouts Identify Talents Better Than Laypeople?

According to the Economic Research Institute (2024), soccer scouts in the Netherlands earn an average yearly salary of 63.194€, which is far above the average of 34.260€ (Bureau for Economic Policy Analysis, 2023). With increasing work experience and expertise, a scout's salary typically rises about ten percent every five years, which is grounded in the belief that greater expertise correlates with higher levels of performance (Economic Research Institute, 2024; Sinclair & Agerström, 2020; Jokuschies et al., 2017).

The central task of a scout is to identify and select extraordinary players for their club or a development program (Jokuschies et al., 2017). Therefore, the ability to distinguish unique talent by recognizing relevant skill patterns and interactions is required and has been termed "expert's eye" (Christensen, 2009; Sinclair, 2023). So far, there has been little research on this eye for talent in soccer scouts, indicating that the process by which scouts identify and combine useful information about the players to predict their future performance is yet ambiguous (Jokuschies et al., 2017). Moreover, the knowledge of the scouts' judgment's reliability and validity is limited (Bergkamp et al., 2021). Accordingly, this research aims to examine the validity and reliability of the expert's eye in soccer scouts by comparing them to individuals lacking that expertise and with alternative approaches to talent assessment. Additionally, suggestions on effective methods that maintain the scout's autonomy and significance in the talent assessment process are offered. Note, that the present attempt is an extension of the research by Peringa (2023), who focused on the significance of noise in the analysis of the soccer scouts' judgments.

The Expert's Eye

The assessment and selection of talented individuals are relevant across various domains such as personnel selection, educational admission processes or athlete selection (Meijer et al., 2020). This is because the suitability and performance of the selected people determine organizational success. Therefore, organizations invest a considerable number of resources into recruiting and training selection experts, hoping that they will distinguish the most talented individuals in their field (Sinclair & Agerström, 2020). The majority of these practitioners rely on their intuition or their “gut feeling” to assess individuals (Sinclair, 2023). During the decision-making process, they typically (1) observe relevant skills, (2) predict future performance, and (3) make a selection decision (Highhouse, 2008).

To illustrate this with a soccer-specific example: When a scout is on the outlook for talented players, they usually attend a soccer game, in which they (1) observe various skills of the players such as sprinting speed and interceptions, (2) predict their future performance based on these skills, and (3) make an intuitive decision on whom they select for their club (Bergkamp et al., 2021).

In practice, intuitive assessments by experts are preferred compared to relying on a decision rule, in which the selection decision depends on a numerical formula (Bergkamp et al., 2021; Meehl, 1954). According to the naturalistic decision-making approach (NDM), the reason for that is that talent assessment is not a rational or objective process that can be reduced to mere numbers as it would be when employing a decision rule (Christensen, 2009, Highhouse & Brooks, 2023; Kahneman & Klein, 2009, p. 518). Instead, it involves complex interactions between the individual and its environment that can only be assessed holistically by another equally complex individual.

Therefore, individuals who developed the expert’s eye are highly valued in many domains to perform these assessments, as they are believed to recognize multifaceted relationships through long-term practice and reflection (Christensen, 2009; Sinclair & Agerström, 2020). They are sought to have an acquired practical sense, which is defined as a developed set of perceptual preferences and principles, often referred to as "taste" (Bourdieu, 1998). The sense entails the development of cognitive structures resulting from the internalization of external observation, as well as action plans that guide the judgments of situations and the responses to them. Consequently, this unique expertise allows experts to differentiate between talented and less talented individuals (Christensen, 2009).

For instance, when a soccer scout attends a game, they might recognize a player sprinting exceptionally fast and making good interceptions but lacking game insight. The scout might have seen this skill combination in a player previously, who became very successful after getting training to improve on the deficit. Thus, they might choose that player for a development program, as they sense their potential.

However, the theory of the expert's eye relies primarily on anecdotal evidence (Christensen, 2009; Highhouse & Brooks, 2023; Highhouse, 2008), which is highlighted by the fact that the majority of evidence on the expert's eye is conducted through the qualitative method, in which the experts elucidate on how they acquired their expertise and give subjective insights into their decision-making process (Bergkamp et al., 2021; Christensen, 2009). Thus, the NDM approach assumes that humans have precise insight into their intuition (Kahneman & Klein, 2009, p. 518). Nevertheless, testable theories in favor of the approach are lacking, which underscores the questionability of its validity (Highhouse & Books, 2023; Kahneman & Klein, 2009, p. 519).

Problems with human intuition

Countering the validity of the expert's eye and criticizing the NDM approach, various quantitative studies suggest that the quality of human intuition is overestimated (Bergkamp et al., 2021; Kahneman & Klein, 2009, p. 523; Sinclair, 2023). Instead, the competing heuristic and bias approach (HB) to assessment argues that human intuition is prone to errors and that subjective insights into decision-making processes are limited (Kahneman & Klein, 2009, p. 517; Sinclair & Agerström, 2020).

First, intuitive judgments tend to be impacted by biases, which are systematic errors that frequently occur in the assessment process (Kahneman & Klein, 2009, p. 519). An example of such is the hindsight bias, which is the tendency to perceive outcomes of a situation as predictable after they have already happened, even if the judge could, in fact, not foresee the outcome (Roese & Vohs, 2012). Consequently, they attribute this false foreseeable outcome to their ability to make accurate predictions.

Let's clarify with a soccer-specific example: When a scout attends a soccer game to identify talents and spots a player with several outstanding skills and some lacking ones, they might not select them for their club. If that player develops into a successful professional athlete in the future, the scout experiences hindsight bias, when they believe that they knew that this particular player would be successful based on the observation they made in the past, which confirms their eye for talent (Roese & Vohs, 2012).

The experts who provide anecdotes in favor of the validity of the expert's eye may be affected by systematic errors as well, therefore presenting biased evidence (Jokuschies et al., 2017). They might believe to have an accurate practical sense, although they are internalizing false interferences due to their biases. As expertise rises, these biases might further manifest as being valid, especially when the environment has unpredictable characteristics and when there is no feedback on the expert's assessment, which are two conditions for developing skilled intuition (Kahneman & Klein, 2009, p.521). Consequently, prediction accuracy might even decrease with higher subjective levels of expertise.

Second, human intuition is impacted by unsystematic error or noise. This implies that the judges do not weigh informational pieces consistently across assessed individuals (Kahneman & Klein, 2009, p.517).

In the context of soccer, a scout's assessment is affected by noise when they assign more weight to sprinting speed for one player than for another (Peringa, 2023). This behavior can be advantageous when it is correlated to valid insights about the player, which is argued to be the standard in an expert's judgment (Christensen, 2009; Kuncel, 2017). Nevertheless, inconsistent weighting is often a true error, conveying that it is not correlated to anything and thereby decreases the validity of the judgment. Even if individuals with high levels of expertise occasionally identify significant variation between assesses, this variance is likely to be outweighed by the negative impact of biases and noise (Goldberg, 1970).

Highlighting this assumption, contemporary research claims that the assessments of experts are not better than those of laypeople who lack the potential “expert’s eye”, thereby fully declining the expert’s practical sense (Sinclair & Agerström, 2020). In the context of soccer, this implies that scouts might not be able to perceive uniqueness and talent in certain players but are predicting future performance just as well as laypeople in soccer such as amateur players or fans.

Although there is some research comparing the validity of judgments between experts and laypeople in, for example, personnel selection, there are very few attempts to transfer these findings into sports (Schorer et al., 2017; Sinclair & Agerström, 2020). Schorer et al. (2017) conducted a study in handball, showing that all four groups which were national coaches, regional coaches, laypeople and novices, predicted between 72.4% (novices) and 79.3% (national coaches) players correctly as being talented or not talented. Regional coaches and laypeople both predicted 75.8% correctly. The criterion that was used to classify if the players were talented or not was their athletic achievements ten years after the predictions were made. It is important to note that the available information that the four samples used predict the future performance varied across groups, which might have impacted the findings. The national and regional coaches both watched the players on their five-day try-outs and also considered the results of the player’s motor tests, while the laypeople and novices saw excerpts of videos of handball games. These conditions might have given the coaches more insight into the player’s performance compared to the remaining groups. Moreover, the coaches who selected the players whom they perceived as most talented during the try-outs also trained them further on, which might have resulted in their training being of higher quality than those players who did not get selected and therefore could have impacted the player’s level of success. It is remarkable that, although it can be argued that the laypeople and novices' groups were disadvantaged through the reduced availability of information about the players, there was no to only moderate difference between the prediction accuracy compared to the national and regional coaches, which strengthens the evidence that the validity of the expert's eye is overestimated (Sinclair, 2023).

The present study provides more equal conditions for the prediction of the experts and laypeople by asking both groups to complete the same judgment task and eliminating the involvement of the experts in the player's training. To do so, soccer is a suitable sport to conduct this research in, as there is accurate performance data available (Peringa, 2023). This data includes the skill levels of the soccer players and their market value which acts as the criterion value that the intuitive predictions can be compared with. Furthermore, the findings are of great relevance within soccer, because selecting the best players determines the player's personal as well as the organization's success competitively and financially. Lastly, the outcome of the study and the judgment task that is administered can also be transferred to other domains as well, which emphasizes the study's relevance further. The lens model is used as a framework to explore and compare the prediction accuracy of both soccer scouts and laypeople in soccer.

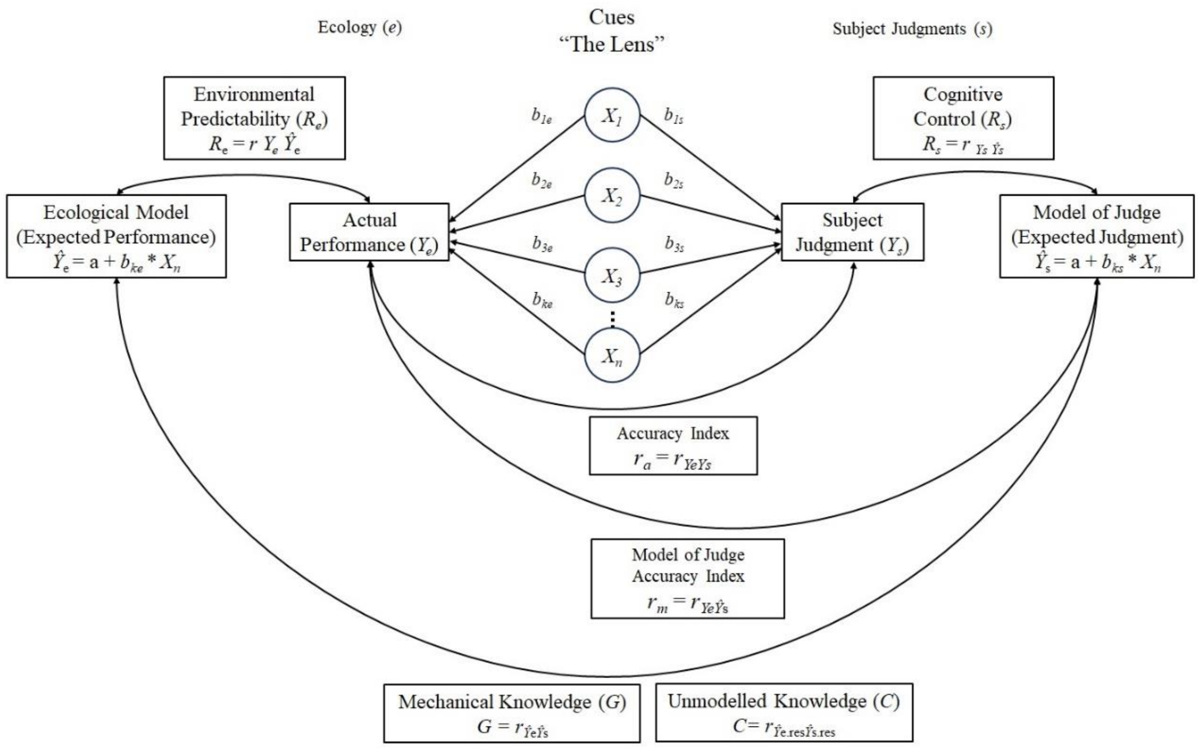
The Lens Model

The lens model is a theoretical framework that was developed to analyze judgment and decision-making processes (Brunswik, 1952; Figure 1), which can also be transferred into the context of soccer performance prediction (Peringa, 2023). The model suggests that a judgment is based on informational pieces that are weighted according to their relevance for the criterion, which is the value that can be predicted based on these informational pieces (Kuncel et al., 2013). For instance, a relevant piece of information to predict soccer performance could be sprinting speed. The model distinguishes which weighting scheme the judge assigns to these informational cues and how consistently it is applied across cases.

Thereby, it allows for the comparison between the validity of the intuitive human judgment and the validity of other approaches to information combination.

Figure 1

The Lens Model



I. Peringa (personal communication, March 18, 2024)

Criterion, Independent Variable Cues, Subject Response and Accuracy Index

The lens model consists of three central elements which are (1) the criterion value, (2) the independent variable cues and (3) the subject response, which interconnections are the basis of the decision-making process (Kuncel et al., 2013).

The (1) criterion value (*Ye*) is defined as the outcome value or the behavior that one wants to predict, for instance, future soccer performance (Kuncel et al., 2013). The (2) independent variable cues (*X1….Xk*) are the pieces of information that can be used to predict the outcome, such as the different soccer skills that are relevant for future performance. The (3) subject response (*Ys*) is the intuitive prediction of the criterion that the judge gives based on the informational cues.

To elucidate, when a soccer scout is judging a soccer player's performance, they examine different skills that they perceive as relevant such as sprinting speed and game insight. By looking through "the lens" of these cues, they judge how well the player performs in a game (Kuncel et al., 2013; Peringa, 2023). Consequently, their intuitive judgment, the subject response, can be correlated with the players' actual indicators of performance, the criterion. The value of this correlation indicates the accuracy index (*ra*) of the judge, thus how well the judge predicts the criterion.

Ecological Validity Model and Cognitive Strategy Model

Moreover, the application of the lens model provides insight into the mechanical approach to assessment which is based on multiple regression (Kuncel et al., 2013). The ecological validity model, which is represented on the left side of "the lens", applies optimal linear weights (*b1e…bke*) to the informational cues. To clarify, this weighting scheme represents the optimal distribution of relevance to the available informational cues to predict the criterion accurately. Additionally, the environmental predictability (*Re*) captures the correlation between the predictions of the ecological model ($\hat{Y}e$) and the criterion values (*Ye*). Thus, it demonstrates how well the criterion can be predicted based on cues of "the lens".

Next, the cognitive strategy model is a regression model that is represented on the right side of "the lens" (Kuncel et al., 2013). It applies the weights (*b1s…bks*), that the human judge applies to the informational cues on average by regressing the intuitive prediction on the informational cues. The resulting model is therefore termed "the model of the judge" ($\hat{Y}s$), as it entails expert-chosen weights that might be suboptimal. Nevertheless, the model qualifies as being linear because it applies the weights consistently across cases, to form an overall judgment.

Considering the model of the judge and the intuitive predictions of the judge, the judge's cognitive control (*Rs*) and the accuracy index of the model of the judge (*rm*) can be calculated (Kuncel et al., 2013). Cognitive control is the correlation between the model of the judge and the intuitive predictions of the judge, representing how consistently the judge applies the weights to the informational cues across cases. The accuracy index of the model of the judge is the correlation between the models' accuracy and the observed criterion values. It represents how accurately the model of the judge predicts the criterion.

Mechanical Knowledge and Unmodeled Knowledge

The lens model displays two more parameters below the lens which will be discussed in this section (Kuncel et al., 2013).

First, the mechanical knowledge (*G*) shows the correlation between criterion values, derived from the ecological validity model and the predictions of the model of the judge (Kuncel et al., 2013). This parameter specifies how well the weighting scheme of the judge corresponds with the optimal weighting scheme for the independent variable cues.

Second, the unmodeled knowledge (*C*) is the correlation between the residuals from the ecological validity model ($res.\hat{Y}e$) and the residuals from the model of the judge ($res.\hat{Y}s$) (Kuncel et al., 2013). It reveals if there is any variance that cannot be captured by the linear model but by the validity of the judge's intuitive prediction. This validity beyond the limits of modeling, can be termed the "expert's eye" (Christensen, 2009). For instance, recognizing a unique pattern such as an interaction effect between two informational cues would result in a positive value of unmodeled knowledge.

Alternative Approaches to Assessment

Offering an alternative to the holistic approach to assessment, in which the judge relies on intuitive information combination to come to an overall judgment, the mechanical approach entails the combination of information by a decision rule (Sinclair, 2023, Peringa, 2023). In this approach, the determined weighting scheme is applied consistently across cases, which is a crucial advantage compared to intuitive judgment, in which weights tend to be assigned somewhat inconsistently (Kahneman & Klein, 2009, p. 517; Peringa, 2023). That noise reduces the validity of intuitive decisions drastically which is why the mechanical approach is superior across various domains (Highhouse, 2008; Kuncel et al., 2013; Schorer et al., 2017; Sinclair & Agerström, 2020).

This superiority does not only hold for optimal regression models, which apply the optimal weighting scheme for the informational cues but also for improper linear models (Dawes, 1979). Improper linear models assign suboptimal weights to the predictors, such as unit weights, correlation weights, or experts-chosen weights. However, these weights are applied consistently across cases, which explains why the models maintain adequate validity. When using expert-chosen weights such as the average weights that the expert decided on in their intuitive judgment and subsequently applying these weights consistently across individuals, the "model of the judge" is created. This model reliably outperforms the holistic judgment of the expert it was created on, although it eliminates the potential "expert's eye", which assumes that inconsistent weighting captures valid variance between individuals (Goldberg, 1970). Various research in the domain of personnel selection confirmed that a consistent weighting scheme improves predictive validity more than allowing the possible expert's eye (Dawes, 1979, Sinclair, 2023). Peringa (2023) transferred these findings and applied the lens model into the context of sports, with professional soccer scouts being the experts. The present study aims to replicate her results in a sample of laypeople in soccer, which allows for the comparison between the judgments of scouts and laypeople.

Another approach to assessment is the formation of a composite, which entails averaging several independent individual judgments to attain a final judgment (Goldberg, 1970). According to the wisdom of the crowd theory (Surowiecki, 2005), a composite judgment tends to achieve higher predictive validity than an individual judgment. The reason for that is, that averaging multiple independent intuitive judgments reduces noise (Kahneman et al., 2021, Chapter 8). Another advantage of the composite assessment is that explained variance beyond the limits of modeling, thus the expert's eye is maintained in a composite assessment if all independent individual judges demonstrate that expertise in their judgment to some extent.

Nevertheless, the importance of independence between the individual judges that form the composite should be stressed (Afflerbach et al., 2020; Kahneman et al., 2021, Chapter 8). According to the principle of information diversity, the more dissimilar the judges, the better the composite judgment performs, which is more likely to be achieved when the judgments are made independently. Dissimilar judges tend to have differing characteristics, biases, and other sources of information that distinguish their judgments. Averaging them tends to result in a more accurate outcome compared to averaging a set of judgments that have very similar sources of information. When the judges even communicate with each other, thus being dependent, the validity of the composite judgment might decrease due to collective error.

The high validity of composite judgments, when gathered from independent judgments, has been supported in different contexts such as in the assessment of medical students or personnel selection (Aghdasi et al., 2015; Kuncel et al., 2013). Also, Kuncel (2023) found that a composite of three novices outperformed a single expert judgment in the context of personnel selection. Taking this a step further, Goldberg (1970) observed in his study that the composite judgment reached even higher validity than the average models of individual judges, indicating that composite judgment might be the superior method of talent assessment. Based on this knowledge, this study tries to transfer the findings into the context of soccer, hypothesizing that the validity of a composite based on three laypeople outperforms the average validity of the scouts.

The current research

Peringa (2023) was the first to apply the lens model in the context of soccer. In the judgment task, eighteen professional soccer scouts were asked to evaluate 50 player profiles in the position of a central defender. Each soccer scout was outperformed by their own model and by the ecological validity model. In her study, she primarily focused on the impact of noise on the scout's judgments. This research aims to extend the research by Peringa (2023), hypothesizing that there is no difference between the prediction accuracy of the eighteen scouts and 50 laypeople. Additionally, this study tries to replicate that each intuitive judgment is outperformed by their own models and that there is no difference between the average validity of the models, distinguishing between experts and laypeople.

The application of the lens model also allows for the comparison of the judges (a) cognitive control, (b) mechanical knowledge, and (c) unmodeled knowledge (Kuncel et al., 2013). Specifically, it can be compared (a) how consistently the experts and laypeople apply their weight distribution and (b) how accurate their weighting scheme is when comparing it to the optimal scheme, which is calculated by multiple regression. It is especially interesting to investigate (c) how much variance beyond linear modeling can be explained by the experts and laypeople, thus how valid the "expert's eye" really is and if it is exclusively present in the scouts. Additionally, the computation of a composite judgment, derived from three laypeople, allows for further comparison between methods of talent assessment.

Taken together, these analyses aim to study the validity of the expert's eye and to explore the validity of alternative methods of information combination in the context of soccer, which can be phrased as the research question: Can soccer scouts predict talent better than laypeople? Based on the previous findings of other domains and the project by Peringa (2023) in the soccer assessment context, the following hypothesis will be examined:

Hypothesis 1: There is no significant difference in the prediction accuracy between experts and laypeople.

Hypothesis 2: There is no significant difference in the prediction accuracy between the models of man, based on the experts and the laypeople.

Hypothesis 3: Both experts and laypeople are outperformed by their own model of the judge.

Hypothesis 4: There is no significant difference between the expert's and laypeople's (a) cognitive control, (b) mechanical knowledge, and (c) unmodeled knowledge.

Hypothesis 5: A composite layperson, derived from three laypeople outperforms the average validity of the scouts.

Methods

Participants

On the basis of a checklist developed by the EC-BSS (Ethics Committee of the Faculty of Behavioral and Social Sciences) at the University of Groningen, the study was exempt from full ethical review. Two groups of participants, the experts and the laypeople, were studied and compared. First, a total of eighteen professional soccer scouts formed the group of experts (Peringa, 2023). They are employed at FC Groningen, a professional soccer club in the Netherlands. The mean age of the scouts was 53 years (*SD* = 14.5) and their average scouting experience was 9.5 years (*SD* = 8.5). Their data was collected by Peringa (2023), who made use of the collaboration between the University of Groningen and FC Groningen.

Second, the sample of laypeople was invited via convenience sampling and consisted of a total of 52 participants. Primarily, individuals who were known by the researcher and players from amateur football clubs in the Netherlands were recruited. The laypeople were required to be at least eighteen years old and to have some affinity toward soccer (e.g., fans, (former) amateur players or coaches). Their mean age was 27.82 years (*SD* = 9.93), with ages ranging from 19 to 65. Two laypeople were excluded from the analysis due to missing values. Thus, 50 laypeople were included in the sample.

Measures

The measure was a self-report survey that consisted of 50 real player profiles (Figure 2). For each of them, the following variables were relevant to compute the parameters of the lens model: (a) criterion value, (b) quantified cue values, and (c) subject response.

The criterion values were the player's actual market values that were reported on the website www.sofifa.com in 2022, and are adequate representatives of the player's performances (Peringa, 2023). The cue values entailed the player's skill levels on tackling, interceptions, sprinting speed and game insight, and were derived from the FIFA video game database, published in 2022. The skills were chosen based on expert knowledge and statistical analyses, which confirmed that they predicted the criterion adequately. The skill levels ranged between 1 and 10 and were categorized through cut-off scores with < 2.4 = *far below average*; 2.5 – 4.4 = *below average*; 4.5 – 5.5 = *average*; 5.6 – 7.5 = *above average*; and > 7.6 = *far above average*. The subject responses were defined as the intuitive overall player judgments made by the participants on a 10-point Likert scale (1 = *totally not valuable*; 10 = *extremely valuable*). Note that both scouts and laypeople were asked to rate the value of each player profile for the position of a central defender. The survey was originally in Dutch but was also translated into English for laypeople who were not fluent in Dutch.

Figure 2

Player Profile and Response Scale with Example Answer Six

Skills	Score Value	Note
Tackling	5.0	Scores average on tackling
Interceptions	8.0	Score *far above* average on interceptions
Sprinting speed	3.0	Scores *below* average on sprinting
Game insight	7.0	Scores *above* average on game insight

Player 1	Totally not valuable									Extremely Valuable
How valuable is this central defender?	1	2	3	4	5	(6)	7	8	9	10

Selecting Player Profiles

The 50 player profiles were selected among the 2437 central back players, registered in the FIFA database (Peringa, 2023). To ensure that the market value of the players can be accurately estimated based on the four skill levels, the values were regressed on the cues. Through examining the residuals-vs-predicted plot, extreme outliers and profiles with Cook's distance larger than .00164 were excluded. Finally, 53 profiles remained, with three of them serving as practice profiles and 50 as the actual judgment task. Since every profile had to be judged by every participant, the total number of judgments made up the final sample, not the mere number of participants per group.

Procedure

Each participant was presented with the same judgment task administered through pencil-and-paper questionnaires. On the first page, the background of the research as well as the instructions were described. After giving informed consent, the participants were encouraged to practice the task on three practice profiles before starting to judge the actual ones. The soccer scouts came together at FC Groningen in 2023 to complete it in the presence of a researcher (Peringa, 2023). Due to the bigger sample size, the laypeople were not accompanied by a researcher but had the opportunity to take home the questionnaire for completion and hand it back to the researcher at a later time. Nevertheless, they had the option to call or email the researcher for questions. At the end of the survey, the scouts were asked for the duration of their scouting experience, while the laypeople indicated their affinity to soccer (e.g., fan, (former) amateur player or coach). Finally, all participants were asked to disclose their age. No further personal information was collected, as the data was treated anonymously.

Statistical Analysis

To calculate the parameters of the lens model, ordinary least squares regression was used (Peringa, 2023).

Ecology (e)

Starting on the left side of the lens model, the *ecological validity* and the *ecological predictability* were calculated (Peringa, 2023). The ecological model was determined by regressing the criterion (*Ye*), the market value of the players, on the informational cues (*X1….Xk*), thus tackling, sprinting speed, interceptions and game insight. This model determines and applies the optimal weight distribution across the cues (*b1e…bke*) to predict criterion values (*Ŷe*). The correlation between these predicted values and the actual criterion values represented the environmental predictability (*Re = rYe Ŷe*), therefore how well the criterion can be predicted based on the cues.

Subject judgments (s)

On the right side of the lens model, the parameters that are related to the intuitive assessments were captured (Peringa, 2023).

First, the accuracy index (*ra*) of the scouts (*raS*) and the laypeople (*raL*) were the correlations between their intuitive judgments and the criterion (*raS = rYeYsS; raL = rYeYsL*), indicating how accurate the scouts and laypeople predict the criterion. These correlations were further used to reveal the validity difference between the scouts and laypeople ($\Delta = raS - raL$), in which a positive value indicates that the scouts predict the criterion more accurately than the laypeople.

Second, using multiple regression, every subject response of the scout (*YsS*) and laypeople group (*YsL*) was regressed on the informational cues (Peringa, 2023). Thereby, the *cognitive strategy model* or the *model of the judge* was created, which utilizes the average weights that the judges assigned to the cues (*bksS; bksL*) to predict the criterion values (*ŶsS; ŶsL*). These models were also used to calculate the accuracy index of the judges' models (*rmS; rmL*), representing the correlation between the models' predicted values and the actual criterion (*rmS = rYeŶsS; rmL = rYeŶsL*).

Furthermore, the validity difference between the model of the judge and the judge from whom it was derived was calculated ($\Delta = rmS - raS$; $\Delta = rmL - raL$), in which a positive value revealed that the model had incremental validity over the judge.

The same procedure was performed with the models of the scouts and laypeople ($\Delta = rmS - rmL$), in which a positive value confirmed the superiority of the model of the scout.

Third, the cognitive control (*Rs*) of both the scouts (*RsS*) and laypeople (*RsL*) was calculated by correlating their subject responses with the predicted values of their models ($RsS = rYsS\ \hat{Y}sS$; $RsL = rYsL\ \hat{Y}sL$). Again, the difference between the scout's and the laypeople's cognitive control ($\Delta = RsS - RsL$) allowed for further comparison between the groups.

Mechanical knowledge and Unmodeled knowledge

The mechanical knowledge (*G*) was calculated by correlating the predictions of the model of the judge with the predictions of the ecological validity model ($GS = r\hat{Y}e\hat{Y}sS$; $GL = r\hat{Y}e\hat{Y}sL$). This value quantified the extent to which the model of the judge was consistent with the optimal weighting scheme, determined by the ecological model. Yet again, the outcome values of the scouts and the laypeople were compared with each other by computing the difference between their mechanical knowledge ($\Delta = GS - GL$).

The unmodeled knowledge (*C*) was calculated by correlating the residuals of the ecological validity model ($\hat{Y}e.res$) with the residuals of the model of the judge ($\hat{Y}sS.res$; $\hat{Y}sL.res$), ($CS= r\hat{Y}e.res\hat{Y}sS.res$; $CL= r\hat{Y}e.res\hat{Y}sL.res$). The value represented the relationship between the criterion and the subject response that could not be attributed to the method of modeling. The difference between the scout's and laypeople's unmodeled knowledge allowed for the comparison between the groups ($\Delta = \hat{Y}sS.res - \hat{Y}sL.res$). If the expert's eye was a phenomenon that predominantly occurred in the experts, the value was positive.

Composite Layperson

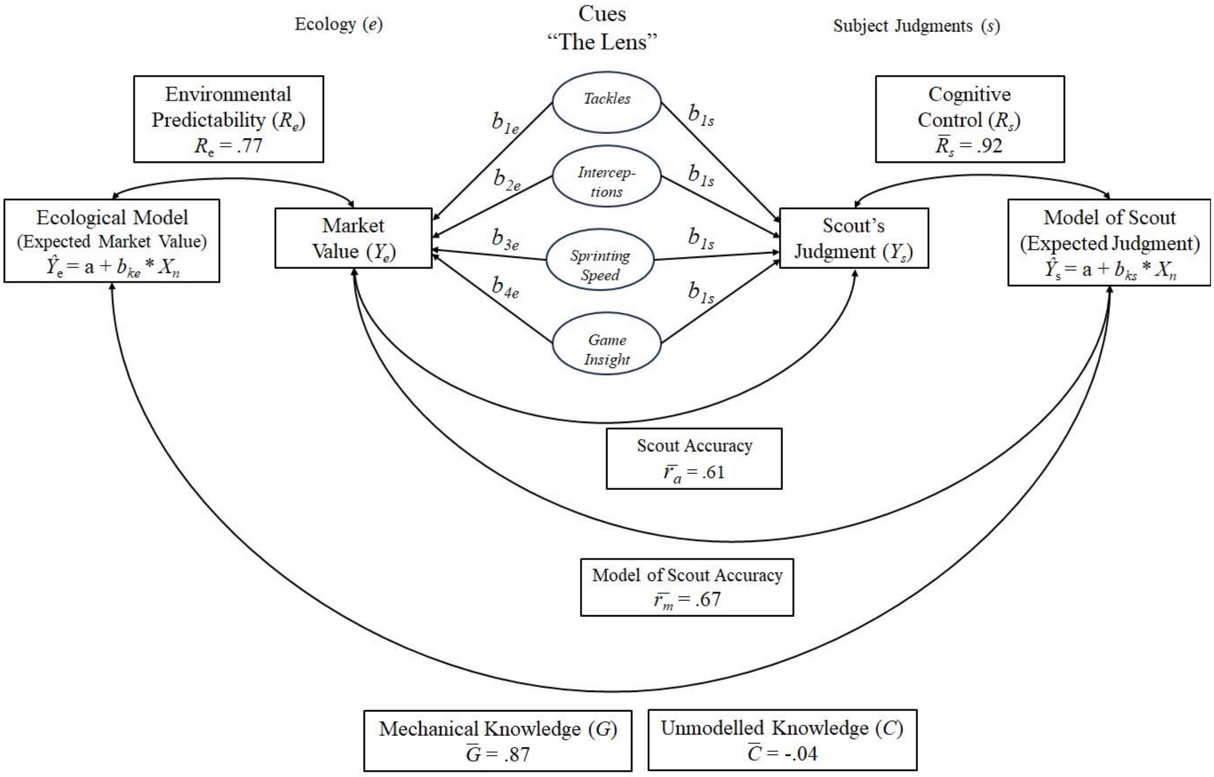
The last analysis entailed the computation of three composite judges, each derived from three laypeople (*ComL*), which were calculated by averaging their subject responses (*(YsLn1 + YsLn2 + YsLn3)/3*). The selection of the judges, which included nine laypeople in total, relied on chance to create a random judge. Subsequently, all parameters of the lens model (accuracy index, cognitive control etc.) were also determined for the composite judges. Thereby, the difference between the average validity of the scouts and the average validity of the composite judges was revealed (*raS – raComL*). A positive value indicated the superiority of the scouts.

Results

Starting on the left side of the lens model, the analysis revealed an environmental predictability (*Re*) of .77 (R^2 = .60). Note that the average values of the lens models' parameters are shown in Figure 3 for the scouts and in Figure 4 for the laypeople.

Figure 3

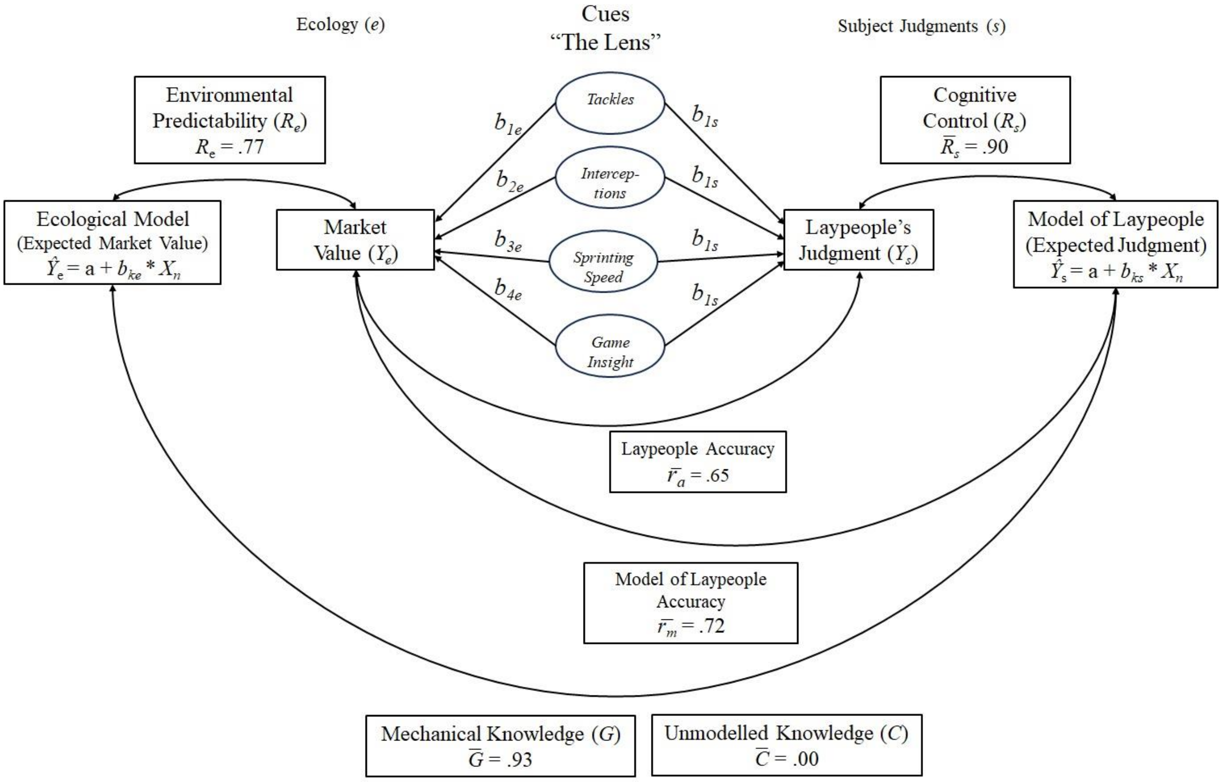
The Lens Model with the Scouts' Results



Note. The values of *ra*, *G*, *C*, and *Rs* are the average values across the scouts.

Figure 4

The Lens Model with the Laypeople's' Results



Note. The values of *ra*, *G*, *C*, and *Rs* are the average values across the laypeople.

Validity of the Judge

The average accuracy index of the scouts (*raS*) was .61, with .40 being the coefficient of the least accurate scout and .74 being the coefficient of the most accurate one. On average, the validity of the laypeople (*raL*) was .65, with .40 being the lowest and .74 being the highest coefficient, which is equal to the most accurate scout. The averages are displayed in Table 1 row 1. The findings revealed that the laypeople made slightly more accurate predictions than the scouts on average (Δ = *-.04)*. This is an interesting finding because, according to the research praising expert's intuitive skills, the scouts should have made better predictions due to their higher levels of expertise (Christensen, 2009).

However, it was unclear if the difference between the groups is significant, due to the complexity of the analysis. Therefore, a composite layperson and a composite scout were computed to test the significance of the difference between their accuracy indexes. Note, that the composite layperson was derived from 18 random laypeople from the sample to create equal conditions for both groups. Using the data of 50 laypeople would likely result in significant superiority of the composite layperson, as averaging 50 individual judgments reduces noise more than averaging the 18 judgments of the scouts does.

The composite layperson had an accuracy index of .72 and the composite scout achieved .66. Subsequently, the significance of the difference between the correlations was tested with a z-test for dependent correlations (Lee & Preacher, 2013). The result showed that the difference was significant ($z = 2.41$, $p = .02$). Still, the applied method of testing did not consider all available data, which means that the outcome might deviate when the full dataset is included.

If the difference was indeed significant, the first hypothesis would not be supported. Nevertheless, the outcome would still be in line with the direction of the research question, which is questioning the superiority of the scout’s intuitive decisions compared to the ones made by laypeople. Note, that the third column of Table 1 does not represent the values of the composites above but the average of the three composite laypeople that are analyzed in one of the subsections below.

Table 1

Mean Validity Coefficients of the Judges

	Mean of Scouts	**Mean of Laypeople**	**Mean of Compostites**
Validity of the Scouts ra	.61	.65	.68
Validity of the Model of the Scout rm	.67	.72	.70

	Mean of Scouts	Mean of Laypeople	Mean of Compostites
Δ rm – ra	.07	.07	.02
Cognitive Control Rs	.92	.90	.94
Mechanical Knowledge	.87	.93	.92
Unmodeled Knowledge	-.04	.00	.07

Validity of the Model of the Judge

The average model of the scouts achieved an accuracy index (*rmS*) of .67. The least valid model had a correlation of .45 with the criterion, while the most valid one had a correlation of .76. The average model of the laypeople (*rmL*) achieved .71. The least valid model had a correlation of .64 with the criterion, while the most valid one had a correlation of .77, which was equal to the value of the environmental predictability. The groups' averages are shown in Table 1 row 2. Overall, the average model of the laypeople achieved higher validity than the average model of the scouts ($\Delta = -.04$) suggesting that, on average, the laypeople applied more optimal weights compared to the scouts. It could be argued that this outcome should have been reversed if the scouts possessed greater expertise. Again, due to the complexity of the analysis, it was unclear if the difference was significant. A non-significant result would have supported the second hypothesis, while a significant difference would have still confirmed that scouts did not make more accurate predictions than the laypeople, who lack the scouts' level of expertise.

Furthermore, for both groups of participants, their model's incremental validity was .07 on average, as shown in Table 1 row 3. That outcome supported the third hypothesis partly, since almost every scout and layperson was outperformed by their own model. Only one layperson reached the same validity as their model, which resulted from a combination of high cognitive control (.90) and a degree of unmodeled knowledge (.13), that was not explained by the model.

Cognitive Control

The average cognitive control of the scouts (*RsS*) was .92, with .86 being the lowest value of achieved cognitive control and .95 being the highest. The average cognitive control of the laypeople (*RsL*) was slightly lower with .90, ranging from .56 to .97 (Table 1 row 4). Accordingly, the laypeople varied more in how consistently they applied a weighting scheme across the player profiles, while all scouts were quite consistent in applying weights to the four cues, which led to the average difference of $\Delta = .02$ between the average values.

Mechanical Knowledge

The average mechanical knowledge (*GS*) was .87 in the scouts group. The scout with the lowest mechanical knowledge achieved a value of .58, while the one with the highest mechanical knowledge achieved .98, which indicates a nearly perfect correlation between the predicted values from the model of the judge and the predicted market values from the ecological model. The laypeople's average mechanical knowledge (*GL*) was .92, ranging from .83 to 1.00 within the sample ($\Delta = -.05$). Note that the averages are shown in Table 1 row 5. The layperson who achieved a mechanical knowledge of 1.00 applied the optimal linear weighting scheme on average and was, therefore, the same individual whose model achieved the same validity as the ecological model (*.77*). These values revealed that the vast majority of variance which the predictions of the participants explained was based on linear modeling, which contradicts the scout's ability to capture nonlinear insight intuitively. This was also highlighted by the analysis of the unmodeled knowledge.

Unmodeled knowledge

A positive value of unmodeled knowledge captured the validity that could not be attributed to linear modeling, but to the judges' unique insights into the player's abilities (Kuncel et al., 2017). It is that part of variance that individuals suggest to be highly significant in the decisions made by experts (Christensen, 2009). In this study, the average unmodeled knowledge of the scouts (*CS*) was -.04, with -.30 being the lowest and .14 being the highest value across them. The average unmodeled knowledge of the laypeople (*CL*) was .00 and ranged between -.30 and .30 (Δ = *-.04)*. The averages of both groups are displayed in Table 1 row 6. The values indicated that both groups did not add validity beyond linear modeling, which further supports the evidence rejecting the significance of the expert's eye (Highhouse & Brooks, 2023; Sinclair & Agerström, 2020).

The results might have been in line with the fourth hypothesis, depending on the significance of the differences between the scout's and laypeople's (a) cognitive control, (b) mechanical knowledge, and (c) unmodeled knowledge. However, the differences were likely non-significant, which would have supported the hypothesis, since the mean differences between the groups are fairly small.

Composite Laypeople

Lastly, three composite judges were calculated and analyzed, each based on three random laypeople. Their average accuracy index (*raComL*) was .68 (Table 1 row 1), which confirmed the fifth hypothesis, suggesting that a composite judgment is more accurate than the average scout (*raS* = *.61*). The incremental validity was .07 but it was unclear if the difference was significant.

Furthermore, the model of the composite had an accuracy index (*rmComL*) of .70 (Table 1 row 2), which means that its' incremental validity was less than a third (Δ *rmComL* – *raComL*= *.02*) of what it was for the scouts and laypeople (Δ *rmS/L* – *raS/L* = *.07*). This was expected because the composite judgment was already less noisy due to averaging the individual judgments. Therefore, there was also less consistency to gain through modeling. The average cognitive control of the composites (*RsComL*) was .94 (Table 1 row 4), which was higher than the value of both, scouts and laypeople (*RsS* = *.92; RsL* = *.90*).

Moreover, the average mechanical knowledge of the composites (*GComL*) was .91 (Table 1 row 5), which was .02 units lower than the average of the laypeople (*GL =.93*) but .04 units higher than the average of the scouts (*GS =.87*). The average unmodeled knowledge (*CComL*) was .07 (Table 1 row 6), which was higher than the average of the scouts and laypeople (*CS =.-.04; CL =.00*).

Providing an overview, Table 2 shows the relevant judgmental variables ordered from the one with the largest predictive validity to the one with the lowest predictive validity.

Table 2

Rank Order of Validity Coefficients of Judgmental Variables

Judgmental Variable	**Validity Coefficient**
Environmental Predictability	.77
Most Accurate Model of the Laypeople	.77
Most Accurate Model of the Scouts	.76
Most Accurate Layperson	.74
Most Accurate Scout	.74
Most Accurate Model of the Composites	.72
Most Accurate Composite	.72
Mean Model of the Laypeople	.72
Mean Model of the Composites	.70
Mean Composite	.68
Least Accurate Model of the Composites	.68
Mean Model of the Scouts	.67
Least Accurate Composite	.66
Mean Laypeople	.65
Least Accurate Model of the Laypeople	.64

Judgmental Variable	Validity Coefficient
Least Accurate Model of the Scouts	.45
Least Accurate Layperson	.42
Least Accurate Scout	.40

Discussion

The present research compared how scouts and laypeople combine information about soccer players to form an overall assessment. Therefore, the lens model was used, as it allowed for the examination of various parameters that define the judge's decision-making process (Kuncel et al., 2013). These entailed the validity of the judges' assessments, the validity of their models, their cognitive control, and their mechanical and unmodeled knowledge. Furthermore, the ecological validity model was calculated as well as three composite laypeople, which were also compared with the judges. The following subsections summarize the findings.

The validity of the judge

Looking at the accuracy of the intuitive predictions, the laypeople's attempts were slightly more accurate than the scout's ones, however, both groups reached fairly high prediction accuracy on average. The slight superiority of the laypeople provides evidence beyond the existing literature in favor of the HB approach, which primarily advocates that prediction accuracy does not increase with higher levels of expertise (Highhouse, 2008; Kahneman & Klein, 2009, p. 517; Kuncel et al., 2013; Sinclair & Agerström, 2020). Moreover, the findings suggest the opposite of what the NDM approach argues (Kahneman & Klein 2009, p. 518). That approach highlights the significance of the expert's eye; however, its existence would imply that the scouts should have made better predictions than laypeople due to their expertise (Christensen, 2009; Jokuschies et al., 2017).

While the significance of the difference between the groups remains somewhat uncertain, the outcome of this analysis provides support for the main research question of this study, which suggests that scouts do not make better intuitive predictions than laypeople. It is also important to mention that the study design was quite artificial because the judges were exposed to merely four numerical cues. In real-life settings, scouts typically attend soccer games in which they can consider different sources of information and observe the behavior in action (Bergkamp et al., 2021). Thus, the conditions of this study might have influenced the scout's prediction accuracy negatively.

Nevertheless, the finding is consistent with evidence across different contexts such as personnel and athlete selection as well as admission settings, which also showed that experts and laypeople do not differ in their prediction accuracy (Highhouse, 2008; Kuncel et al., 2013; Schorer et al., 2017; Sinclair & Agerström, 2020).

The validity of the model of the judge

The scouts and laypeople reached high levels of cognitive control on average, meaning that they weighted the cues quite consistently across player profiles (Kuncel et al., 2013). It is noteworthy that the cues were strong predictors of the criterion values, which resulted in high environmental predictability and allowed for high levels of prediction accuracy based on the available information (Peringa, 2023).

However, the models still reached incremental validity over the judges. Both groups were consistently outperformed by their own model, except for one layperson, which strongly supports the second hypothesis. The finding is robust across various research comparing the mechanical approach to assessment with the holistic approach in which people rely on their intuition (Goldberg, 1970; Kuncel et al., 2013; Peringa, 2023). The main reason for the models' superiority is that human intuition tends to be impacted by noise, which is reduced by modeling (Meijer et al., 2020). Highlighting this, Peringa (2023) found that there is a clear correlation (.61) between levels of consistency across the scout's judgments and levels of prediction accuracy, with low consistency reducing prediction accuracy and high consistency increasing prediction accuracy ($p < .01$). Thus, the models that reached the highest incremental validity over the scouts were based on the ones with the least cognitive control. Conversely, the models that were based on the scouts with the highest cognitive control, reached the lowest incremental validity.

Mechanical Knowledge and Unmodeled Knowledge

The laypeople achieved slightly higher values of mechanical knowledge than the scouts, with the same being true for unmodeled knowledge, however, the difference was very small and likely non-significant. While the values of mechanical knowledge were on average high, the ones of unmodeled knowledge were close to zero. This means that the validity of the judgments of both groups of participants can be attributed to their use of an internal decision rule rather than exclusive insights that cannot be captured by modeling (Peringa, 2023). Therefore, in the context of this experiment, relying on the intuitive decisions of scouts is not beneficial when it comes to talent assessment and selection, as they do not add unique insight. Instead, applying a model that is based on regression or the judges themselves performs better, although it eliminates any potential expert's eye.

Composite

The last step of the analysis was the computation of three composite laypeople, which were based on three random laypeople from the sample. The results of the analysis are in line with the fifth hypothesis, stating that the composite outperforms the average scout. The outcome provides support for the wisdom of the crowd theory, stating that averaging the responses from several individuals improves predictive validity compared to a response from a single judge (Surowiecki, 2005). That is, because averaging the responses decreases noise, while also applying more optimal weights to the cues. Taking this a step further, calculating the model of the composite increased predictive validity even more, due to the elimination of any remaining noise. However, it can be argued that the effort of gathering and analyzing a composite judge is worth the quite small improvement of prediction accuracy compared to using the model of a single judge.

As shown in Table 2 (Environmental Predictability), the most accurate way to combine information is the application of a regression-based model. The reason for that is that human intuition is negatively impacted by noise and biases, which does not improve with higher levels of expertise across many contexts, such as sports, clinical diagnosing and personnel selection (Highhouse, 2008; Kuncel et al., 2013; Schorer et al., 2017; Sinclair & Agerström, 2020). These settings lack immediate, unambiguous feedback and are characterized by low levels of predictability of the environment, which are the conditions for developing skilled and accurate intuition according to Kahneman and Klein (2009) (p. 520). Therefore, scouts cannot reliably learn from their previous decisions to make more accurate ones in the future. Instead, they are prone to internalize false beliefs, which could even decrease their prediction accuracy with rising expertise.

Practical Implications

To tackle the errors of human intuition, a mechanical approach to combining information should be considered (Meijer et al., 2020). Peringa (2023) presented how to implement the mechanical approach in the context of soccer based on the five-step guide introduced by Kahneman (2011). However, many practitioners and individuals favoring the NDM approach are resistant to these recommendations, with one of the main reasons being the loss of the expert's autonomy and significance in the decision-making process (Highhouse, 2008; Neumann et al., 2022). Therefore, Meijer et al. (2020) suggested four ways of information combination that reach greater predictive validity than intuitive judgments, while preserving the judge's autonomy.

First, unit weighting is a method in which all relevant informational cues are weighted equally (Meijer et al., 2020; Neumann et al., 2023), for instance: Overall score = game insight x (1) + sprinting speed x (1) + tackling x (1) + interceptions x (1). The autonomy of the scouts would be limited but not completely removed, as they can still choose the cues that they perceive as being relevant to predict future performance and rate them (e.g. on a Likert scale) before the formula combines these scores to determine the overall score.

The second method, which allows for slightly more autonomy, entails a weighting scheme provided by the experts (Meijer et al., 2020; Neumann et al., 2022). For instance, the scouts would not only choose and rate the cues per player but also decide how much weight should be assigned to each cue. Referring to the cues of this study, the scouts could decide on the following scheme: Overall score = game insight x (2) + sprinting speed x (1) + tackling x (1) + interceptions x (2). These two methods above create improper linear models that maintain adequate predictive validity because the information is weighted consistently, with only the weights themselves being suboptimal (Dawes, 1979).

Third, mechanical synthesis preserves the expert's autonomy to a greater extent and also allows for the representation of the potential expert's eye (Meijer et al., 2020; Neumann et al. 2023). In this method, the scouts make an intuitive assessment which gives room to their expertise. This score is further treated as an informational cue in the decision rule such as: Overall score: game insight x (2) + sprinting speed x (1) + tackling x (1) + interceptions x (2) + intuitive assessment x (1). The predictive validity is not reduced, as long as the intuitive assessment is not weighted too heavily in the formula.

Lastly, in holistic synthesis, the cues are combined mechanically at first. Further on the experts can combine the mechanical score with other information, that they perceive as relevant (Meijer et al., 2020; Neumann et al., 2022; Neumann et al., 2023). Although the scouts are likely to feel most autonomous when using this method, it is also prone to error, as the scouts can overrule the mechanical score based on their intuition. However, it still tends to achieve more predictive validity than a complete intuitive prediction, while allowing for the expertise of the scouts.

Alternatively, the composite prediction can be implemented, in which every scout makes an intuitive prediction or makes use of one of the methods above to come to an overall score. Subsequently, the overall scores of at least two scouts can be averaged to improve predictive validity further. Thereby, the autonomy and expertise of the scouts can be partly maintained depending on the combination method chosen before the composite is formed.

Strengths and Limitations

This study had several strengths, with one being the equality of the judgment task and its statistical appropriateness. The four chosen were strong predictors of the criterion, that allowed for the computation of a valid linear model and high environmental predictability (Peringa, 2023). Furthermore, the judgment task was equal for both groups of participants, which ensured high comparability. This is an improvement compared to the design by Schorer et al. (2017), in which the conditions differed for the experts and laypeople, which might have impacted the validity of the comparison.

Next, the sample size can be considered a strength, because it involved the number of judgments made per participant instead of the number of participants. Thus, there was enough data present to ensure that the results did not happen by chance.

A limitation of the study is that the sample of laypeople might not be representative of the general layperson in soccer. That is because they were sampled by convenience and included a great number of highly educated individuals with a socio-economic status that is very likely to be higher than average. Their high levels of education might have influenced their judgments positively, masking the superiority of the scouts compared to the average layperson.

Another limitation was that this research exclusively focused on data combination. Typically, scouts are also responsible for data collection within the process of talent assessment (Bergkamp et al., 2022). Data collection is prone to bias and noise as well, which was not considered in the present study. Moreover, rating the player’s performance during a game requires abilities such as observational skills, quick reaction time, and analytical skills which might contribute to the scout's expert's eye, but are primarily relevant during data collection. Therefore, in future studies, it might be interesting to investigate the expert's eye in a design that considers both data collection and data combination.

Furthermore, some scouts employ an unstructured approach, in which base their decisions on an overall impression of the player rather than distinguishing between data collection and combination (Bergkamp et al., 2022). Future research could aim to compare the scout's and laypeople's prediction accuracy when administering an unstructured approach to extend the knowledge on the expert's eye further.

To investigate these implications, researchers could invite scouts and laypeople to a several soccer games in which they rate players using either an unstructured or structured approach that distinguishes between data collection and combination (Bergkamp et al., 2022). Next, the results could be compared with the players' current market values or their future performance achievements using a longitudinal design. However, implementing the lens model in such a design might present several challenges, as it requires identifying and quantifying the information used by judges (Peringa, 2023).

Conclusion

This study challenged the existence of the expert's eye in soccer by comparing the prediction accuracy between scouts and laypeople. Employing the lens model, the results are in line with the quantitative studies in other domains, arguing that the expert's eye is overestimated (Highhouse & Brooks, 2023; Sinclair & Agerström, 2020). Even with high levels of expertise, intuitive judgments are accompanied by significant amounts of error, which is primarily a result of inconsistent weighting of the information used within the decision-making process. Contradicting the research favoring the validity of the expert's eye, the findings show that scouts do not add more valid variance within certain players than laypeople do. In fact, on average, neither group of individuals added insights that a mechanical approach to assessment did not capture. That is the reason, why administering a method that entails a decision rule when it comes to combining information is superior to a mere intuitive judgment, as it eliminates inconsistent weighting, which is the greatest issue when it comes to solely intuitive judgments (Kahneman & Klein, 2009, p. 517).

Note

During the preparation of this work, the authors used ChatGPT (OpenAI, 2014) in order to refine the readability of the final manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

Aghdasi, N., Bly, R. A., White, L., Hannaford, B., Moe, K. S., & Lendvay, T. S. (2015). Crowd-sourced assessment of surgical skills in cricothyrotomy procedure. *Journal of Surgical Research*, *196*(2), 302–306. https://doi.org/10.1016/j.jss.2015.03.018

Afflerbach, P., Van Dun, C., Gimpel, H., Parak, D., & Seyfried, J. (2020). A Simulation-Based approach to understanding the wisdom of crowds phenomenon in aggregating expert judgment. *Business & Information Systems Engineering*, *63*(4), 329–348. https://doi.org/10.1007/s12599-020-00664-x

Bourdieu, P. (1998). *Practical reason: on the theory of action*. Polity Press.

Bergkamp, T., Frencken, W., Niessen, A. S. M., Meijer, R. R., & Hartigh, R. D. (2021). How soccer scouts identify talented players. *European Journal of Sport Science*, *22*(7), 994–1004. https://doi.org/10.1080/17461391.2021.1916081

Bergkamp, T., Meijer, R. R., Hartigh, R. D., Frencken, W., & Niessen, A. (2022). Examining the reliability and predictive validity of performance assessments by soccer coaches and scouts: The influence of structured collection and mechanical combination of information. *Psychology of Sport and Exercise*, *63*, 102257. https://doi.org/10.1016/j.psychsport.2022.102257

Brunswik, E. (1952). *The conceptual framework of psychology*. University of Chicago Press.

CPB Netherlands Bureau for Economic Policy Analysis (2023) https://www.cpb.nl/en/about-cpb

Christensen, M. K. (2009). “An Eye for Talent”: Talent identification and the “Practical Sense” of Top-Level Soccer Coaches. *Sociology of Sport Journal*, *26*(3), 365–382. https://doi.org/10.1123/ssj.26.3.365

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571–582. https://doi.org/10.1037/0003-066x.34.7.571

Economic Research Institute (2024). https://www.erieri.com/salary/job/scout-professional-sports/netherlands

Goldberg, L. R. (1970). Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin*, *73*(6), 422–432. https://doi.org/10.1037/h0029230

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology*, *1*(3), 333–342. https://doi.org/10.1111/j.1754-9434.2008.00058.x

Highhouse, S., & Brooks, M. E. (2023). Improving Workplace Judgments by Reducing Noise: Lessons Learned from a Century of Selection Research. *Annual Review of Organizational Psychology and Organizational Behavior*, *10*(1), 519–533. https://doi.org/10.1146/annurev-orgpsych-120920-050708

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

Kahneman, D. (2011). *Thinking, fast and slow*. http://ci.nii.ac.jp/ncid/BB2184891X

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). Noise: A flaw in human judgment. Little, Brown. https://doi.org/10.1111/jols.12345

Kuncel, N. R. (2017). Judgment and decision making in staffing research and practice. In *SAGE Publications Ltd eBooks* (pp. 474–487). https://doi.org/10.4135/9781473914940.n17

Kuncel, N. R. (2023, May 16). *Studium Generale: Decision Making with Algorithms* [Presentation]. Faculty of Behavioural and Social Sciences, University of Groningen. https://www.rug.nl/about-ug/latest-news/events/calendar/studium-generale/decision-making-algorithms?lang=en

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology*, *98*(6), 1060–1072. https://doi.org/10.1037/a0034156

Lee, I. A., & Preacher, K. J. (2013, September). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Available from http://quantpsy.org.

Meehl, P. E. (1954). The problem. In *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* (pp. 3–9). University of Minnesota Press. https://doi-org.proxy-ub.rug.nl/10.1037/11281-001

Meijer, R. R., Neumann, M., Hemker, B. T., & Niessen, A. S. M. (2020). A tutorial on Mechanical Decision-Making for Personnel and Educational selection. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.03002

Neumann, M., Niessen, A. S. M., Hurks, P. M., & Meijer, R. R. (2023). Holistic and mechanical combination in psychological assessment: Why algorithms are underutilized and what is needed to increase their use. International Journal of Selection and Assessment. Advance online publication. https://doi.org/10.1111/ijsa.12416

Neumann, M., Niessen, A. S. M., Tendeiro, J. N., & Meijer, R. R. (2022). The autonomy-validity dilemma in mechanical prediction procedures: The quest for a compromise.

Journal of Behavioral Decision Making, 35(4), e2270. https://doi.org/10.1002/bdm.2270

OpenAI. (2024). GPT-3.5 [Computer software]. OpenAI. https://openai.com/chatgpt

Peringa, I. (2023). *Assessment of Soccer Performance by Scouts: Does the Scout Outperform a Model of the Scout?* [Unpublished master's thesis]. University of Groningen.

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science*, *7*(5), 411–426. https://doi.org/10.1177/1745691612454303

Schorer, J., Rienhoff, R., Fischer, L., & Baker, J. (2017). Long-Term prognostic validity of talent selections: comparing national and regional coaches, laypersons and novices. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01146

Sinclair, S. (2023). Applicants’ Faith in Recruiters’ Intuition Predicts Process Favorability for the Unstructured Employment Interview. *Scandinavian Journal of Work and Organizational Psychology*, *8*(1), 10. https://doi.org/10.16993/sjwop.183

Sinclair, S., & Agerström, J. (2020). Does expertise and thinking mode matter for accuracy in judgments of job applicants' cognitive ability? *Scandinavian Journal of Psychology*, *61*(4), 484–493. https://doi.org/10.1111/sjop.12638

Surowiecki, J. (2005). *The wisdom of crowds: Why the many are smarter than the few*. https://ci.nii.ac.jp/ncid/BA8523911X