2 (2026) 100094

Contents lists available at ScienceDirect

Football Studies

journal homepage: www.sciencedirect.com/journal/football-studies

Original Investigation

Talent identification is not just a prediction problem: A sequential decision framework for football selection under uncertainty

Robbie S. Wilson [a,*], Mathew S. Crowther [b], Andrew H. Hunter [a], Nicholas M.A. Smith [a], Simon P. Lailvaux [c]

[a] School of the Environment, The University of Queensland, St Lucia, QLD, 4072, Australia
[b] School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, 2006, Australia
[c] Department of Biological Sciences, University of New Orleans, 2000 Lakeshore Drive, New Orleans, LA, 70148, USA

ARTICLE INFO

Keywords:
Talent identification
Team sport
Performance
Football
Simulations
Sequential talent
Identification

ABSTRACT

Research into football talent identification has focused heavily on developing and validating assessments that predict football performance and subsequent attainment. However, establishing the predictive validity of an assessment is not the same as determining how the information it provides should be used to make selection decisions. Football academies must make repeated selection decisions using imperfect information, which requires balancing the need to reduce candidate numbers with the resources required for further assessment. Here, we use simulations to examine how assessment information of different predictive strengths can be used within a sequential selection process. Specifically, we examine how imperfect predictors can be used to efficiently reduce candidate populations while preserving predefined target performers, how the order of assessments influences screening efficiency, and how sequential assessment can reduce resource requirements. Experiment 1 showed that an assessment explaining approximately half of the variation in performance could still reduce a candidate population by more than 50% while retaining all target performers. Experiment 2 showed that the order of assessment matters: using stronger information earlier allowed approximately 7–11 additional players per 100 candidates to be removed while retaining all target performers. Experiment 3 showed that reserving more resource-intensive assessments for the smaller groups following earlier screening stages reduced total assessment time by 70–87% but this came at the cost of retaining modestly larger final candidate groups. Together, these findings demonstrate that even assessments explaining only moderate variation in the underlying performance criterion can support substantial candidate reduction, and that their screening value depends on when they are introduced and the resources required to obtain them. The sequential framework provides a way to use existing information about predictive strength to determine which assessments should be used, when they should be introduced, and which players require further evaluation.

Introduction

For more than three decades, research into quantitative approaches in football talent identification has focused heavily on the search for better predictors of football performance and subsequent attainment [1–3]. This has led to considerable effort directed toward identifying the technical, physical, psychological, tactical and developmental characteristics associated with differences in football performance and subsequent attainment, with assessments typically evaluated by their ability to explain variation in performance outcomes, distinguish among athletes competing at different levels, or predict subsequent attainment [1–8]. This work has provided valuable insights into the attributes associated with football performance and the predictive validity of different assessment approaches [1–3,9,10]. However, establishing the predictive validity of an assessment is not the same as determining how the information it provides should be used within a selection system. Predictive validity tells us how strongly an assessment is related to a relevant performance outcome; it does not, by itself, tell us what selection decision should follow from that information. Academies must make repeated selection decisions using incomplete information, commonly integrating scouting and coach judgement with other sources of player information. These decisions may concern whether a player

* Corresponding author.
E-mail address: r.wilson@uq.edu.au (R.S. Wilson).

https://doi.org/10.1016/j.footst.2026.100094
Received 6 July 2026; Received in revised form 15 September 2026; Accepted 9 October 2026
Available online 11 October 2026
3051-2689/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

should be removed, retained, or assessed further, and the appropriate decision may depend on both the information currently available and where the player is within the selection process. We therefore consider talent identification from the broader perspective of decision-making under uncertainty: given the imperfect information currently available about a player, what decision can reasonably be made, and what information should be collected next? We use this term as a conceptual framing of the selection problem rather than to imply a complete formal decision-theoretic model. In many complex environments, effective decisions depend not only on the predictive accuracy of individual sources of information, but on how that information is used within the constraints of the decision system [11].

A sequential decision framework recognises that complex selection problems are rarely solved through a single assessment event. Instead, information can be gathered progressively, with the information available at one stage determining which candidates can reasonably be removed and which require further assessment [12,13]. Here, we examine a specific form of this problem: how far can a candidate population be reduced using assessment information that is related to, but does not perfectly predict, an underlying performance criterion, while preserving a predefined set of target performers (for example, the highest-performing 5 or 10% of the candidate population)? This allows us to distinguish the predictive strength of an assessment from its screening value. Predictive strength describes how closely an assessment is related to the underlying performance criterion, whereas screening value describes how much the candidate population can be reduced while preserving the predefined target performers. This distinction has parallels in personnel-selection research, where the practical value of an assessment depends not only on its predictive validity, but also on the selection decisions it informs and the consequences of those decisions [14-16]. This deliberately controlled simulation design allows the effects of predictive strength, information order and assessment requirements on screening outcomes to be isolated while other features of the decision process are held constant. The simulations therefore do not constitute a complete formal decision-theoretic model; utility or loss functions, risk preferences, reliability and formal uncertainty reduction are not independently modelled.

This perspective is particularly relevant to football academies because they must select relatively small numbers of players from much larger candidate pools. Early decisions may determine who warrants further observation or invitation to subsequent trials, whereas later decisions may determine academy entry, retention or release [17–21]. Staged selection itself is not unique to football, with talent detection, identification and selection representing related processes within broader sporting talent pathways [22,23]. For example, Bullock et al. [22] described a talent-identification programme in skeleton in which 26 athletes identified initially from 30-m sprint performance were reduced to 10 following a selection camp and subsequently to four selected for World Cup competition. More recently, the multi-sport YouFor2032 programme provides a contemporary example of sequential talent identification across 19 Olympic sport disciplines. The programme comprises seven phases from talent strategy and recruitment through talent detection, selection and confirmation to subsequent development and integration; following initial testing, shortlisted athletes can progress through selection into a 12-week talent-confirmation phase [23]. Such systems provide clear precedents for progressively evaluating and reducing candidate populations across successive stages, but also highlight an important distinction from the problem considered here. In many individual sports, later-stage assessment can increasingly approximate or directly measure the performance criterion that ultimately defines competitive success, such as time, speed or distance. Football provides no equivalent directly measurable individual criterion; instead, selectors must integrate multiple imperfect sources of information that differ in both their relationships with match performance and the resources required to obtain them. The contribution of the present framework is therefore not just staged selection itself, but also an explicit examination of how information of different predictive strengths and assessment requirements can be used across those stages.

Evaluating the consequences of such selection decisions using observational data is difficult. Once a player is deselected from an academy or development pathway, their subsequent trajectory is often no longer observed, creating a fundamental missing-data problem: we rarely know what would have happened had that player been retained. Similar "selective label" problems occur across decision systems in which outcomes are observed primarily for individuals who survive earlier selection stages [24]. The present simulations do not model or correct selective-label bias; instead, they deliberately bypass this observational constraint by generating the underlying performance criterion for every individual. This provides a controlled environment in which predictor–criterion relationships, assessment order, selection stringency and resource requirements can be varied while the consequences of retaining or removing every candidate remain observable. The generated criterion is used retrospectively to evaluate how far a candidate population can be reduced while preserving predefined target performers; it is not information that practitioners would possess prospectively, nor should it be interpreted as future developmental potential or eventual elite attainment. This simplification is a strength of simulation because it allows general properties of sequential screening to be isolated that may be difficult or impossible to identify from observational talent-identification data [25]. Simulation therefore provides a way to take empirically estimated predictor–criterion relationships and ask a complementary question: given information of a particular predictive strength, how can that information most effectively be used within a sequential selection system?

The value of this distinction becomes apparent because early and late stages of a selection pathway represent different decision problems. Screening 1000 young players is fundamentally different from making a final decision among 20 highly selected candidates. Early in the pathway, even an imperfect predictor may have substantial value if it can reduce the number requiring further assessment while preserving the target performers. In the present simulations, this screening capacity is evaluated under a deliberately conservative zero-false-negative constraint: the poorest candidates in a task are removed only up to the point where the next removal would exclude a predefined target performer. Because the underlying performance criterion is known, this is a retrospective measure of screening capacity rather than a prospective estimate of false-negative probability. The relevant question for an early-stage assessment is therefore not simply "how much performance variance does this measure explain?", but "how much candidate reduction can information of this predictive strength support while preserving the target performers?" Later in the pathway, when candidates are more similar and the consequences of decisions become greater, additional and more discriminating information may be required.

Assessment resources provide a further reason why information may have different value at different stages. Talent-identification systems can initially consider hundreds or thousands of players, yet only a small proportion can receive intensive assessment, coaching and monitoring [2,3]. Controlled technical assessments such as dribbling and passing can provide relatively repeatable estimates from brief, standardised observations, whereas more contextually representative assessments such as 1v1 contests, small-sided games and match observations may require repeated observations across different opponents, teammates and contexts to obtain stable estimates of individual performance. Their relevant cost is therefore not simply the duration of a single trial or game, but the assessment effort required to obtain useful individual-level information. Rapid assessments may consequently be valuable when candidate pools are large, whereas more resource-intensive assessments can be concentrated on progressively smaller groups. Importantly, reliability itself is not independently parameterised in the present simulations; differences in the observation and replication required to obtain stable estimates provide the practical

rationale for the fixed assessment-time requirements used in Experiment 3. The resulting question is not simply which assessment is the strongest predictor, but how predictive information should be balanced against the resources required to obtain it.

We therefore use simulation modelling to examine three fundamental properties of sequential talent identification. First, how does predictive strength influence the screening value of an assessment? Experiment 1 tests how much candidate reduction is supported by different levels of predictor–criterion association while preserving predefined target performers. Second, does the order in which information is introduced influence screening efficiency? Experiment 2 tests whether stronger predictors provide greater screening value when introduced earlier in the pathway. Third, how should the information provided by an assessment be balanced against the resources required to obtain it? Experiment 3 compares sequential assessment, in which increasingly resource-intensive assessments are reserved for progressively smaller candidate groups, with comprehensive assessment in which every candidate completes every assessment. Together, these experiments test how predictive strength, information order and assessment resources influence the capacity of sequential screening to reduce candidate populations while preserving predefined target performers. Prediction and decision-making are therefore not competing approaches: informative and appropriately validated predictors remain fundamental to effective selection. Rather, our central proposition is that predictive accuracy alone does not determine how information should be used within an efficient talent-identification system.

Methods

Simulation framework

Across all simulations, talent identification was modelled as a sequential decision problem rather than a single prediction task. The objective of each simulation was to evaluate how different sources of assessment information could be used to progressively reduce a candidate population while preserving predefined target performers on the generated underlying performance criterion. Because this criterion was generated by the simulation, the performance ranking of every individual was known. This allowed us to identify the predefined target performers and then determine retrospectively how far screening based on imperfect assessment information could proceed before one of those performers would be removed. Such outcomes are typically difficult to quantify in real talent-identification systems because players who are deselected are often no longer monitored.

Three simulation experiments were developed to examine different properties of sequential screening. Experiment 1 examined how the predictive strength of an assessment influences its screening capacity. Experiment 2 examined whether screening efficiency depends on the order in which information of different predictive strengths is introduced, comparing strongest-to-weakest and weakest-to-strongest sequential strategies with simultaneous evaluation using all available information. Experiment 3 examined how the information provided by an assessment can be balanced against the resources required to obtain it, testing whether assessment effort could be reduced by reserving increasingly resource-intensive assessments for progressively smaller candidate groups.

Population sizes were selected for illustrative purposes. Experiments 1 and 2 used populations of 100 players to make proportional changes in candidate retention directly interpretable, whereas Experiment 3 used 1000 players to illustrate the resource consequences of sequential assessment within a large initial candidate pool. Candidate-pool size and selection stringency were treated as distinct properties. For example, retaining the top 10 of 100 candidates represents the same proportional selection target as retaining the top 100 of 1000, whereas retaining the top 10 of 1000 represents a substantially more stringent selection problem. We therefore examined sensitivity to selection stringency directly by varying the number of target performers retained in Experiments 2 and 3.

The simulations were designed as illustrative tests of general decision principles rather than as representations of a specific academy selection system. Consequently, the parameter values define the scenarios examined rather than prescribing universal thresholds or assessment pathways for football talent identification.

Experiment 1: how does predictive strength affect the efficiency of early-stage screening?

The first simulation examined how strongly an assessment must relate to the underlying performance criterion before it becomes useful within an early-stage selection process. Rather than testing whether a measure could accurately identify the single best individual, we examined whether imperfect assessment information could substantially reduce the number of candidates requiring further evaluation while preserving all predefined target performers.

A population of 100 players was generated with an observed player characteristic sampled from a standard normal distribution (mean = 0, SD = 1). This variable represented any measurable source of information that may contribute to player evaluation, such as technical skill, physical performance, or scouting assessments.

The generated underlying performance criterion was simulated as: Performance = Predictor + ε, where ε represents variation in the criterion not explained by the measured predictor. This unexplained variation represents the combined influence of any factors affecting the performance criterion that are not captured by the assessment, including measurement error, tactical ability, psychological factors, developmental differences, environmental influences, and random variation.

The amount of unexplained variation was systematically manipulated by sampling ε from normal distributions with mean = 0 and standard deviations of 0.2, 0.6, 1.0, 1.4, 1.8, and 2.2, thereby producing progressively weaker predictor–criterion relationships.

For each predictive-strength condition, 5000 independent datasets were generated (total = 30,000 datasets; 3,000,000 simulated player observations). For each dataset, linear regression was used to quantify the predictor–criterion relationship using the coefficient of determination (R^2). Mean R^2 values were calculated across replicates.

To evaluate screening capacity, the 10 highest-ranked players on the generated underlying performance criterion were first defined as the target performers. All players were then ranked according to the measured predictor. Beginning with the lowest predictor-ranked player, candidates were removed sequentially until the next removal would eliminate one of the 10 target performers. The number of candidates remaining at this stopping point was recorded for each simulation.

Selection outcomes for each predictive-strength condition were summarised using the mean and median number of players retained, the corresponding mean proportion of the candidate population that could be removed, and the 2.5th and 97.5th percentiles of the number retained across simulations. These percentiles describe variation among individual simulated selection exercises rather than uncertainty around the estimated mean. Given the large number of simulation replicates, comparisons among predictive-strength conditions focused on the magnitude and distribution of selection outcomes rather than null-hypothesis significance tests.

This allowed us to determine whether information explaining only modest variation in the underlying performance criterion could still provide screening value by reducing the number of players requiring additional evaluation while preserving all predefined target performers. All simulations and analyses were conducted in R (version 4.3.2), with linear models fitted using the base lm() function and distributional summaries calculated across the 5000 replicates within each predictive-strength condition.

Experiment 2: does the order of information acquisition influence selection efficiency?

The second simulation examined whether the order in which information is introduced influences the efficiency of sequential selection. Populations of 100 players were generated with three correlated predictor variables (W, X and Y) and a generated underlying performance criterion (Z). Predictors were moderately intercorrelated ($r = 0.40$), while their relationships with Z differed in strength (W: $r = 0.60$; X: $r = 0.50$; Y: $r = 0.40$). These relationships were specified through a multivariate normal correlation structure. For each simulated population, target performers were defined according to their rank on Z before the assessment strategies were applied.

Three assessment strategies were compared. In the strongest-to-weakest sequential strategy, players were first ranked using W. Beginning with the lowest-ranked player, candidates were removed sequentially until the next removal would exclude a target performer. Predictor X was then introduced and the remaining players were reranked using the mean of W and X, after which the same removal procedure was repeated. Finally, Y was introduced and the remaining players were ranked using the mean of all three predictors, and the removal procedure was repeated. The weakest-to-strongest sequential strategy followed the same procedure but introduced predictors in the reverse order (Y, X, W). Predictors were equally weighted as they accumulated within both sequential strategies so that the two strategies differed only in the order in which information became available.

The composite strategy represented simultaneous evaluation using all three predictors. To avoid deriving predictor weights from the same candidate population to which they were subsequently applied, composite weights were specified from the population correlation structure used to generate the simulations. Standardised multiple-regression coefficients for predicting Z from W, X and Y were calculated from this known correlation matrix and fixed across all simulated candidate populations. The resulting relative weights were 0.533 for W, 0.333 for X and 0.133 for Y. Players were ranked according to this weighted composite score and removed sequentially from the lowest-ranked individual upward until the next removal would exclude a target performer. This strategy therefore represented a fully informed comparison in which all three sources of assessment information were available before any candidate was removed.

Selection stringency was varied by defining the target group as the top 10, top 5 or top 2 performers on Z. Each selection criterion was simulated independently 5000 times for each strategy. Selection efficiency was summarised as the mean and median number of players removed and retained, the corresponding proportion of the original candidate population removed, and the 2.5th–97.5th percentile range of outcomes across simulations. Differences between strategies were expressed directly as the additional number of candidates that could be removed while preserving all target performers and as the proportion of simulated candidate populations in which one strategy removed more players than another. Given the large number of simulation replicates, comparisons focused on the magnitude and distribution of these effects rather than null-hypothesis significance tests. All simulations were conducted in Python 3.11.

Experiment 3: should the strongest predictors of football performance always be prioritised?

The third simulation examined how the predictive information provided by an assessment can be balanced against the resources required to obtain it. Specifically, we tested whether a sequential assessment pathway could reduce assessment requirements by reserving increasingly resource-intensive sources of information for progressively smaller candidate groups while preserving predefined target performers. A population of 1000 players was generated with four correlated assessment variables representing dribbling, passing, 1v1 performance and small-sided game (SSG) performance, together with a generated underlying performance criterion.

Assessment variables were moderately intercorrelated ($r = 0.40$). Their relationships with the underlying performance criterion were parameterised to represent progressively stronger predictive information across the assessment pathway. Dribbling and passing each explained 36% of the variation in the underlying performance criterion, 1v1 performance explained 50%, and SSG performance explained 64%. These values were selected as illustrative football-relevant parameters informed by previous empirical work [9,26,27] and should therefore be interpreted as defining the simulated scenario rather than as universal properties of these assessment types. For each simulated population, target performers were defined according to their rank on the generated underlying performance criterion before either assessment strategy was applied.

Two assessment strategies were compared. Under the sequential strategy, all players first completed the dribbling assessment. Players were ranked and removed sequentially from the bottom of the ranking until the next removal would exclude a target performer. The remaining players then completed the passing assessment and were reranked according to their mean performance across dribbling and passing, after which the same removal procedure was repeated. Remaining players subsequently completed the 1v1 assessment and were reranked according to their mean performance across the three available assessments, and the removal procedure was repeated. Finally, the remaining group completed the SSG assessment and was ranked using the mean of all four assessments. Assessments were equally weighted as information accumulated across stages so that candidate reduction at each stage was based on all assessment information available at that point in the pathway.

The composite strategy represented simultaneous evaluation in which all 1000 players completed all four assessments before any candidates were removed. To avoid deriving predictor weights from the same candidate population to which they were subsequently applied, composite weights were specified from the population correlation structure used to generate the simulations. Standardised multiple-regression coefficients for predicting the underlying performance criterion from the four assessments were calculated from this known correlation matrix and fixed across all simulated candidate populations. The resulting relative weights were 0.146 for dribbling, 0.146 for passing, 0.291 for 1v1 performance and 0.417 for SSG performance. Players were ranked according to this weighted composite score and removed sequentially from the lowest-ranked individual upward until the next removal would exclude a target performer. The composite strategy therefore represented a fully informed comparison in which information from all four assessments was available before selection.

Selection stringency was varied by defining target performers as the top 20, top 10 or top five players according to the generated underlying performance criterion, corresponding to the highest-performing 2%, 1% and 0.5% of the initial candidate population, respectively. Each selection criterion was simulated independently 5000 times.

Assessment resource requirements

The resource requirements assigned to assessments in Experiment 3 represented the assessment effort required to obtain a sufficiently stable estimate of individual performance, rather than simply the duration of a single trial or game. This distinction is important because football assessments differ substantially in the amount of replication required to estimate individual performance consistently. Highly standardised physical and technical tests minimise contextual variation, such that performance is determined primarily by the focal player and relatively stable estimates can be obtained from a small number of repeated trials. As assessments become increasingly representative of football, however, observed performance becomes progressively dependent on interactions with opponents, teammates and the changing game context [28]. A

single 1v1 contest therefore reflects the performance of both players, while performance in a small-sided game is additionally influenced by multiple teammates and opponents [26]. Estimating an individual player's typical performance consequently requires repeated observations across varied opponents and team compositions so that these contextual influences can be averaged across observations. Empirical work using the assessment framework that motivated the present simulations illustrates this difference: controlled sprint and dribbling performance can be estimated from two or three trials per player [9,29] whereas representative assessments require repeated contests against multiple different opponents in 1v1 [26], and repeated small-sided games with systematically varied team compositions to obtain stable player-level estimates [30]. These empirical differences provide the practical basis for the assessment-resource requirements used in Experiment 3; reliability itself, and the relationship between additional replication and measurement precision, were not independently modelled.

Assessment costs were parameterised using illustrative estimates from these published studies based on the assessment protocols represented in the simulation. With six assessors available, dribbling required two minutes per player (total of 20 min for 60 players) and passing five minutes per player (total of 50 mins for 60 players) when players are distributed equally among assessors. The 1v1 assessment required two hours per 50 players with six assessors, while the repeated SSG assessment required three hours per 40 players using six assessors. These values therefore represent the estimated assessment effort required to obtain an individual-level performance estimate under each protocol, incorporating the practical requirements for replication, randomisation, observation and scoring.

For the sequential strategy, total assessment time was calculated from the number of players entering each stage of the pathway. Under the composite strategy, all 1000 players completed every assessment. Selection efficiency was summarised using the mean and median number of players retained, the 2.5th–97.5th percentile range of outcomes across simulations, total assessment time, hours saved relative to simultaneous assessment, and the proportional reduction in assessment time. Given the large number of simulation replicates, comparisons focused on the magnitude and distribution of differences between strategies rather than null-hypothesis significance tests.

All simulations and analyses were conducted in Python 3.11.

Results

Experiment 1: how does predictive strength affect the efficiency of early-stage screening?

Decreasing predictive strength progressively reduced the ability of a single assessment to eliminate candidates while preserving all target performers (Fig. 1). When unexplained variation was low (SD = 0.2; mean $R^2 = 0.961$), only 14.0 of the original 100 players needed to be retained on average to preserve all 10 target performers, corresponding to an 86.0% reduction in the candidate pool. At SD = 0.6 ($R^2 = 0.734$), 30.6 players were retained, representing a 69.4% reduction.

Importantly, substantial screening efficiency remained at moderate predictive strength. At SD = 1.0, the assessment explained approximately half of the variation in the underlying performance criterion ($R^2 = 0.50$), yet only 48.4 players needed to be retained on average. Thus, an assessment with substantial unexplained variation could still reduce the candidate population by 51.6% while retaining all 10 predefined target performers.

Screening efficiency declined progressively as predictive strength weakened (Fig. 1). At $R^2 = 0.340$, 61.4 players were retained (38.6% reduction), increasing to 69.7 players at $R^2 = 0.239$ (30.3% reduction) and 74.1 players at $R^2 = 0.177$ (26.0% reduction). Across the simulated

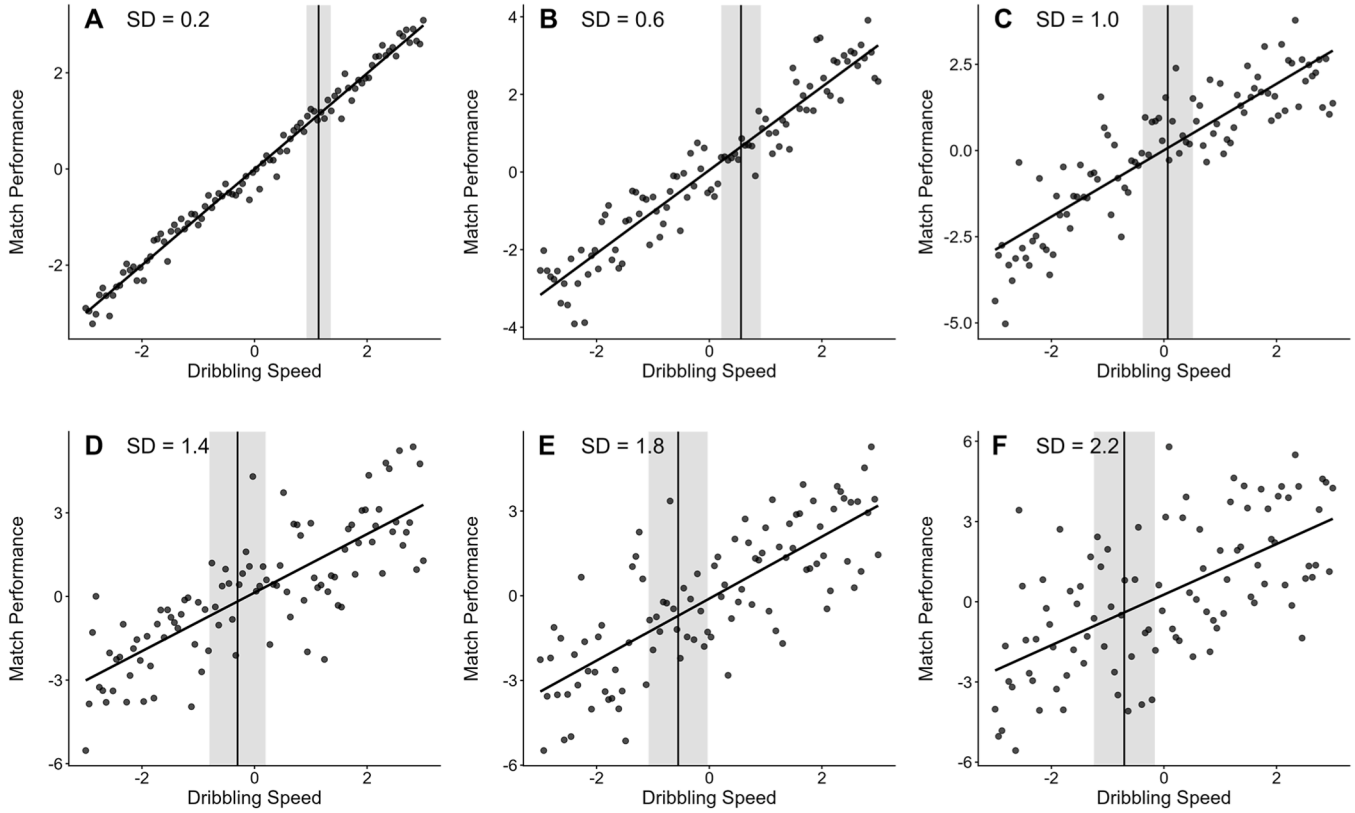


Fig. 1. Relationship between the simulated predictor and generated underlying performance criterion under six levels of added normally distributed variation (SD = 0.2, 0.6, 1.0, 1.4, 1.8 and 2.2). Vertical lines indicate the mean predictor cut-off for candidate removal while preserving all 10 predefined target performers. Shaded regions represent the standard deviation of the cut-off value.

range, therefore, reducing predictive strength from $R^2 = 0.961$–0.177 increased the average number of candidates requiring further assessment from 14.0 to 74.1.

Although these mean effects were estimated precisely across 5000 replicates, outcomes varied among individual simulated selection exercises, particularly at intermediate predictive strengths. For example, at $R^2 \approx 0.50$, the median number retained was 47 players, while 95% of simulations required between 22 and 82 players to be retained. This variation describes differences in screening outcomes among individual simulated candidate populations when an imperfect predictor is applied, rather than uncertainty around the estimated mean effect.

Experiment 2: does the order of information acquisition influence selection efficiency?

Selection efficiency differed substantially among assessment strategies and depended on the order in which information was introduced (Fig. 2). When the objective was to preserve the top 10 performers, the strongest-to-weakest sequential strategy removed an average of 57.8 of the original 100 players, compared with 49.9 players using the weakest-to-strongest strategy and 47.2 players using the composite strategy (Fig. 3). Thus, beginning with the strongest predictor allowed an additional 10.6 players per 100 candidates to be removed relative to evaluating all predictors simultaneously using the composite score.

The same pattern was evident under more stringent selection criteria (Fig. 3). When preserving the top five performers, the strongest-to-weakest strategy removed an average of 75.1 players, compared with 67.8 using the weakest-to-strongest strategy and 65.4 using the composite strategy. When preserving only the top two performers, the corresponding values were 88.7, 84.3 and 81.8 players, respectively. The advantage of strongest-to-weakest sequential assessment over the composite strategy therefore corresponded to an additional 10.6, 9.7 and 6.8 players removed per 100 candidates when preserving the top 10, five and two performers, respectively.

The advantage of beginning with the strongest predictor was also highly consistent among simulated candidate populations. The strongest-to-weakest strategy removed more players than the composite strategy in 96.5% of simulations when preserving the top 10 performers, 95.8% when preserving the top five, and 90.0% when preserving the top two. By comparison, the weakest-to-strongest strategy produced only a small average advantage over the composite strategy, removing approximately 2–3 additional players per 100 candidates across the three selection criteria.

Across all strategies, the number of candidates that could be removed increased as the target group became more selective (Fig. 3). The consistent advantage of the strongest-to-weakest sequence indicates that sequential screening efficiency depended not simply on collecting information in stages, but on when the most informative assessment entered the process. Introducing stronger information early allowed larger reductions in the candidate pool before subsequent assessments were required.

Experiment 3: should the strongest predictors of football performance always be prioritised?

Sequential assessment substantially reduced the resources required to preserve all target performers, although the fully informed composite strategy produced somewhat greater final discrimination among candidates (Fig. 4). Assessing all 1000 players using all four assessments required 134.4 h with six assessors. When preserving the top 20 performers, the sequential pathway required an average of only 39.9 h, saving 94.5 h and reducing total assessment time by 70.3% (Fig. 5).

The resource advantage increased as the selection objective became more stringent. Preserving the top 10 performers required an average of 26.7 h under sequential assessment, representing a saving of 107.7 h or 80.1% relative to assessing every player using all four assessments. When preserving the top five performers, sequential assessment required only 17.4 h, saving 117.0 h and reducing assessment time by 87.0%. Across the three selection criteria, sequential assessment therefore required an average of 28.0 h compared with 134.4 h for simultaneous assessment, corresponding to an average reduction in assessment effort of 79.2%.

This reduction in assessment effort involved a trade-off in the number of players retained at the end of the selection process (Fig. 5). When preserving the top 20 performers, the sequential strategy retained an average of 94.7 players, compared with 76.8 under the composite strategy. Corresponding values were 47.8 versus 38.3 players when preserving the top 10 performers and 22.4 versus 18.5 players when preserving the top five. Thus, collecting all four sources of information from every candidate produced somewhat greater final discrimination, whereas sequential assessment retained a larger final candidate group in exchange for substantially lower assessment requirements.

Discussion

Talent-identification research has generated a substantial body of evidence concerning the ability of different assessments to predict football performance and subsequent attainment. In this study, we asked a complementary question: how can information of different predictive strengths be used most effectively within a staged selection process? Using simulation models, we show that the screening value of assessment information depends not only on its predictive strength, but also on when that information is introduced and the resources required to obtain it. Across three experiments, we show that: (1) assessments

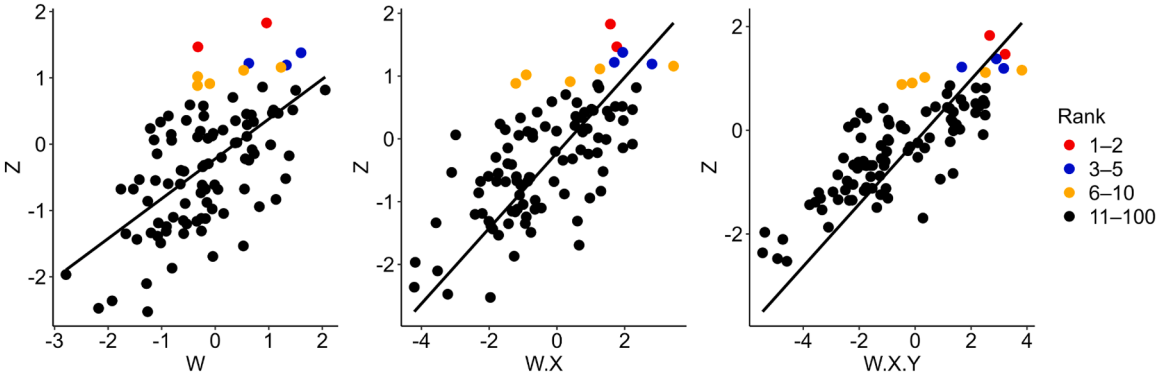


Fig. 2. Example relationships between (A) W, (B) W + X and (C) W + X + Y and the generated underlying performance criterion (Z) for a single simulated population of 100 individuals. Individuals ranked 1–2 on Z are shown in red, ranks 3–5 in blue and ranks 6–10 in orange; all remaining individuals are shown in black. Solid lines indicate linear regression lines of best fit.

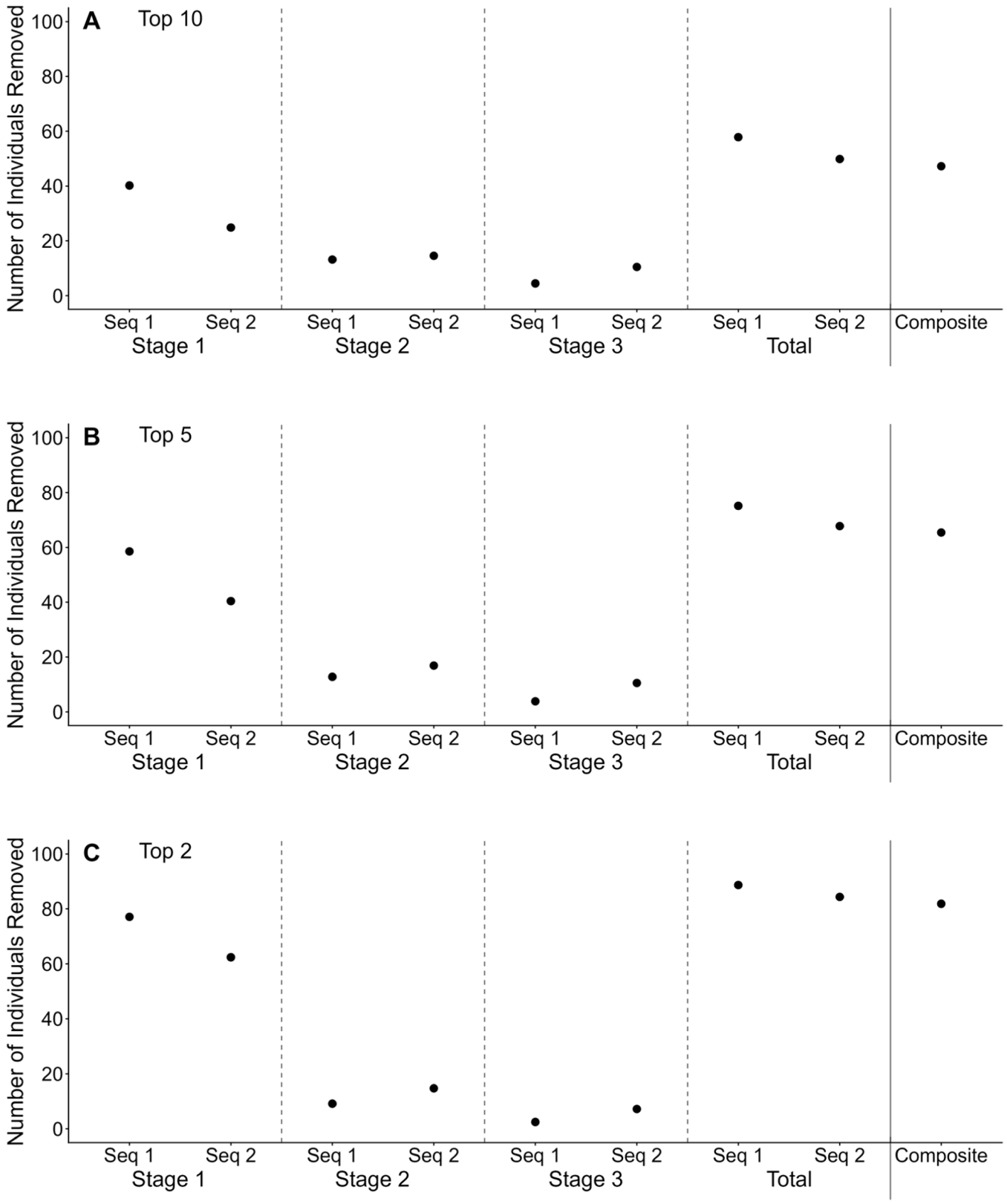


Fig. 3. Mean number of individuals removed at each stage of sequential assessment under strongest-to-weakest and weakest-to-strongest predictor ordering. Panels A–C represent selection criteria requiring preservation of the top 10, top 5 and top 2 target performers on the generated underlying performance criterion (Z), respectively. The regression-weighted composite approach is shown as a single-stage comparison because all predictor information is combined before selection. Candidate removal proceeded only while all predefined target performers were preserved.

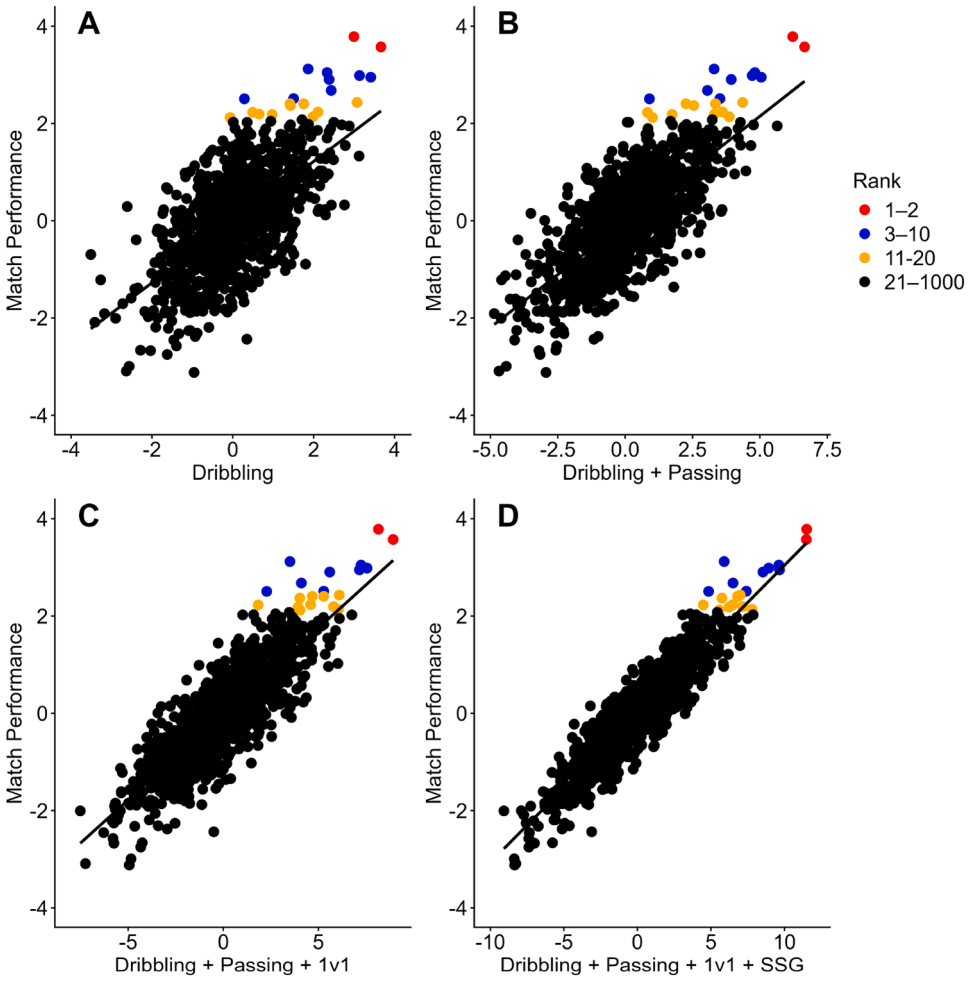


Fig. 4. Example relationships between (A) dribbling, (B) dribbling + passing, (C) dribbling + passing + 1v1 and (D) dribbling + passing + 1v1 + SSG assessment information and the generated underlying performance criterion for a single simulated population of 1000 individuals. Individuals ranked 1–2 on the underlying performance criterion are shown in red, ranks 3–5 in blue and ranks 6–10 in orange; all remaining individuals are shown in black. Solid lines indicate linear regression lines of best fit.

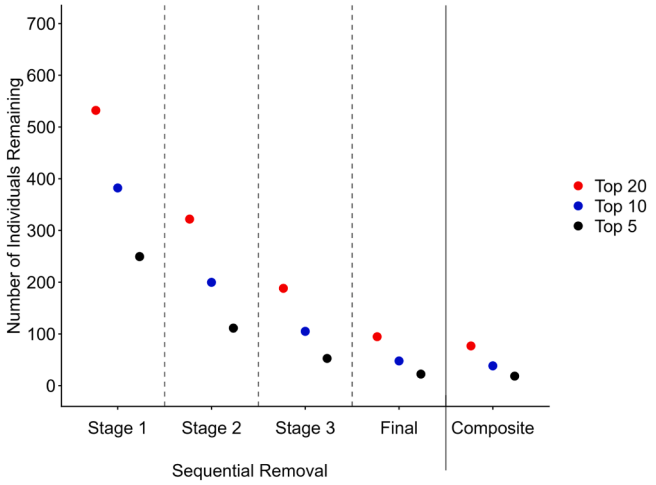


Fig. 5. Mean number of individuals remaining after each stage of sequential assessment under three selection stringencies requiring preservation of the top 20, top 10 or top 5 performers on the generated underlying performance criterion. Sequential assessment progressively incorporated dribbling, passing, 1v1 and small-sided-game (SSG) information. The regression-weighted composite approach represents the mean number retained when all four assessments were administered to all 1000 candidates before selection.

explaining only moderate variation in the underlying performance criterion can still provide substantial screening value, (2) screening efficiency depends on the order in which information is introduced, with stronger information providing greater value when introduced earlier when assessment costs are equivalent, and (3) sequential assessment can substantially reduce resource requirements by using the most resource-intensive assessments on progressively smaller candidate groups.

A major implication of this framework is that assessments should not be evaluated only by how much variation in performance they explain. Measures with modest predictor–criterion relationships may have limited ability to discriminate accurately among the highest-performing individuals but may still provide useful information for an earlier screening decision. In a sequential selection pathway, early-stage assessments can be used to reduce large player populations while preserving predefined target performers under the imposed zero-false-negative constraint. Experiment 1 demonstrated this principle by showing that even measures explaining only moderate variation in the underlying performance criterion could substantially reduce candidate numbers while retaining all target performers. A simple technical or physical assessment may therefore contribute substantial screening value even when much of the variation in the performance criterion remains unexplained. The critical question for such an assessment is not whether it predicts performance perfectly, but how much candidate reduction its predictive strength can support at that stage of the pathway.

The second major implication is that screening value depends on

when information is introduced, while its practical value also depends on the resources required to obtain it. Experiment 2 demonstrated that, when assessment costs were equivalent, introducing the strongest predictor first consistently produced greater screening efficiency than either introducing information in the reverse order or evaluating all predictors simultaneously. This advantage was substantial and highly consistent across simulated candidate populations, showing that screening value depends not only on predictive strength but also on when information enters the decision process. Experiment 3 extended this principle by incorporating differences in assessment effort. Sequential assessment reduced total assessment requirements by 70–87%, depending on selection stringency, by restricting increasingly resource-intensive assessments to progressively smaller candidate groups. Importantly, this efficiency involved a trade-off. Comprehensive assessment of every player produced somewhat greater final discrimination, whereas sequential assessment retained modestly larger final candidate groups at substantially lower resource cost. The advantage of sequential assessment therefore does not arise because it creates more information, but because it allocates assessment effort more selectively. The appropriate strategy consequently depends on the objectives and constraints of the system: when resources are limited, retaining a somewhat larger candidate group at an early stage may allow substantially greater assessment effort to be concentrated on the candidates for whom subsequent decisions are most difficult (Fig. 6).

The numerical outcomes of these simulations should not be interpreted as prescribing a universal selection pathway for football academies. The 70–87% reduction in assessment effort observed in Experiment 3 is conditional on the predictor strengths and assessment-resource requirements represented in that simulation and should not be interpreted as a universal estimate of the savings produced by sequential assessment. The optimal sequence will necessarily depend on the characteristics of a particular system, including the size of the candidate population, the predictive information provided by available assessments, the stability of the resulting performance estimates, and the resources required to obtain them. Future applications could additionally incorporate the relative consequences assigned to different selection errors. The practical contribution of the framework is therefore not a fixed ordering of assessments or set of selection thresholds, but a means of making these trade-offs explicit. Empirical estimates from a particular academy or development system could be incorporated within the same framework to identify context-specific assessment pathways and test whether sequential allocation improves decision efficiency in practice.

A limitation of the present study is that the simulations were designed to test general principles of selection rather than reproduce the full complexity of football development. Real talent pathways involve changing motivation, coaching quality, injury, psychological development, environmental influences and differences in opportunity that cannot be fully captured within a simulation model. However, this simplification is also a strength of the approach because it allows specific mechanisms to be isolated under conditions where the underlying performance criterion is known and the consequences of alternative screening decisions can be observed directly—something rarely possible using observational academy datasets. Future studies should therefore extend this framework using longitudinal player data to examine how sequential decision models perform within applied talent-development environments. The simulations also assume relatively simple linear relationships between assessment information and the underlying performance criterion. Relationships within applied talent-identification systems may instead be nonlinear, threshold-dependent or change across development, which could alter the quantitative relationship between assessment information and screening efficiency. Future extensions should therefore examine whether the same sequential decision principles hold under alternative functional relationships and developmentally changing predictor structures. In addition, predictor intercorrelations were fixed at $r = 0.40$ in Experiments 2 and 3 to provide a common moderate correlation structure. The incremental value of adding information will necessarily depend on the degree of redundancy among assessments, with highly correlated measures expected to provide less additional information than more independent measures. Future applications should therefore examine alternative correlation structures and, where possible, parameterise these relationships using empirical correlations from the assessment system being evaluated.

An important consideration for any talent-identification system is that selection decisions can reproduce or amplify existing structural inequalities in access to development opportunities. Differences in socioeconomic circumstances, geography, access to coaching and facilities, competition opportunities and other structural factors can influence both the performance information available to selectors and the opportunities subsequently provided to athletes [23,31–33]. Recent evidence from a large multi-sport talent-identification programme provides a particularly clear example: even among athletes who met predefined selection criteria, regional athletes were less likely to receive offers than metropolitan athletes and faced systematically higher empirical selection thresholds [23]. These concerns apply to

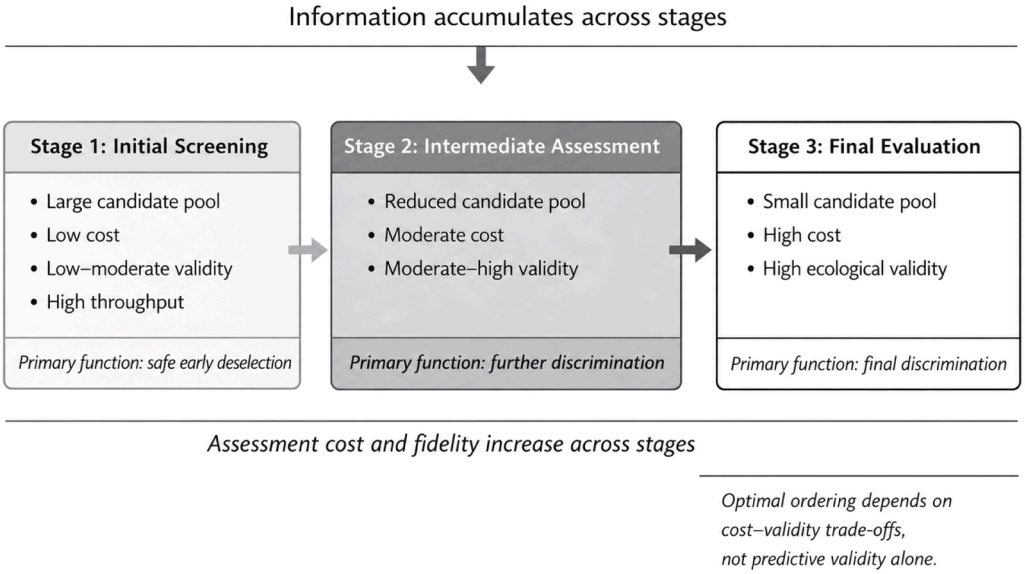


Fig. 6. Conceptual model of sequential talent identification as progressive information acquisition and candidate reduction under resource constraints.

talent-identification systems generally, but are especially important whenever early selection decisions determine access to subsequent assessment, coaching and development. Predictive validity alone should therefore not be considered sufficient justification for an early screening measure, and any staged selection system requires explicit safeguards against systematic disadvantage. These should include monitoring stage-specific retention patterns across relevant socioeconomic, geographic or opportunity groups, examining whether predictor–criterion relationships and decision thresholds operate similarly across groups, and providing additional observation or formal opportunities for re-entry where appropriate.

An important advantage of making these processes explicit within a sequential framework is that such safeguards can themselves become measurable components of the selection system. Because predictors, thresholds and deselection decisions are explicitly defined, the framework can identify where differential retention occurs, quantify its magnitude and test whether alternative thresholds, assessment weighting, additional observation or other interventions reduce those inequalities. The same outcomes can then be monitored in subsequent cohorts to determine whether changes to the selection system have actually reduced the identified disparities. Transparency does not itself ensure equitable outcomes, but it provides a mechanism through which equity can be explicitly evaluated, audited and improved rather than assumed. Equity should therefore be evaluated alongside screening efficiency when implementing the framework.

Youth football provides a clear example of another source of systematic bias because players are typically grouped into annual age cohorts despite substantial differences in chronological age, biological maturation and physical development. Relative age effects and maturation biases are well documented, with older and earlier-maturing players more likely to enter elite development pathways, despite these advantages not necessarily reflecting greater long-term potential [19, 34–37]. Although these developmental processes are not represented in the present simulations, the broader sequential framework provides a structure within which such biases could be examined in future models. From a sequential decision perspective, these biases are particularly important because errors made early in the pathway may permanently remove individuals before later-developing attributes can be expressed or recognised. For example, a physically immature player excluded during an initial screening stage may never reach later stages where technical ability, learning capacity or tactical understanding could be recognised.

The explicit structure of a sequential framework also provides opportunities to examine potential responses to these developmental biases. Rather than only determining whether age or maturation bias exists, future simulations could test how age-adjusted, maturity-adjusted or developmentally informed measures influence selection outcomes. An important challenge, however, is determining what should be adjusted. Age- and maturation-related advantages are commonly quantified for individual physical or technical characteristics, yet adjustment of these measures may not adequately represent their effects on competitive performance in team sports. A more direct approach would be to quantify how chronological age, biological maturation and body size influence performance within representative competitive tasks and use these relationships to inform developmental corrections. Estimating these effects is more difficult because competitive performance is relational: the observed performance of a focal player depends not only on their own characteristics, but also on those of their opponents and, in more complex game formats, their teammates. Developing methods that separate these interacting effects could therefore provide more appropriate developmental adjustments for sequential screening than correcting isolated physical attributes alone. Although approaches such as bio-banding attempt to reduce developmental differences, they are often applied after initial selection has already occurred. Incorporating developmentally informed corrections directly into early-stage screening decisions may therefore provide a way to reduce developmental false-negative errors and improve the retention of players whose performance develops later. More broadly, monitoring and re-entry could themselves be incorporated as additional stages within a sequential decision framework, allowing previously excluded players to re-enter a selection pathway as new information about their development becomes available. Future simulations could therefore examine how developmental corrections, monitoring frequency and re-entry opportunities alter the trade-off between screening efficiency and developmental false-negative decisions.

In conclusion, improving talent identification may require greater attention not only to developing and validating predictors, but to how predictive information is used within selection systems. The present simulations show that assessment information can be evaluated not only by its association with an underlying performance criterion, but by the candidate reduction it supports while preserving predefined target performers, when that information is introduced, and the assessment resources required to obtain it. By providing a controlled environment in which the underlying criterion is known for every individual, simulation allows these properties of sequential screening to be isolated and tested in ways that are difficult using observational talent-identification data. This framework therefore provides a bridge between the extensive empirical literature on predictive validity and the practical problem of deciding how that information should be sequenced and used within staged selection pathways.

CRediT authorship contribution statement

Robbie S. Wilson: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Mathew S. Crowther:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization. **Andrew H. Hunter:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Conceptualization. **Nicholas M. A. Smith:** Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Simon P. Lailvaux:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We sincerely thank the two anonymous reviewers for their thoughtful and detailed comments, which challenged us to clarify the scope and assumptions of our approach and substantially improved the clarity, rigour and quality of the manuscript. The authors used generative artificial intelligence (ChatGPT, OpenAI) as a coding assistance tool during development and revision of the simulation scripts, including assistance with translating author-defined simulation procedures into R and Python code, debugging and code revision, and minor manuscript editing. The research questions, simulation design, parameter choices, selection rules and interpretation of results were determined by the authors. All final simulation code was reviewed and executed by the authors, and outputs were checked against the intended simulation procedures. The complete code used for the final analyses is provided as Supplementary material.

No competing Interests.

The authors received no funding for this study.

Data and Code Availability

The complete code used to generate and analyse all three simulation experiments is provided as Supplementary material. This includes all code required to reproduce the simulated datasets, selection procedures, composite weighting, statistical summaries, and results reported in the manuscript.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.footst.2026.100094.

References

[1] Ali A. Measuring soccer skill performance: a review. Scand J Med Sci Sports 2011; 21(2):170–83. https://doi.org/10.1111/j.1600-0838.2010.01256.x.
[2] Sarmento H, Anguera MT, Pereira A, Araújo D. Talent identification and development in male football: a systematic review. Sports Med 2018;48:907–31.
[3] Williams AM, Reilly T. Talent identification and development in soccer. J Sports Sci 2000;18:657–67.
[4] Johnston K, Wattie N, Schorer J, Baker J. Talent identification in sport: a systematic review. Sports Med 2018;48:97–109.
[5] Kelly A, Wilson MR, Jackson DT, Williams CA. Technical testing and match analysis statistics as part of the talent development process in an English football academy. Int J Perform Anal Sport 2020;20:1035–51.
[6] Koopmann T, Faber IR, Baker J, Schorer J. Assessing technical skills in talented youth athletes: a systematic review. Sports Med 2020;50:1593–611.
[7] McCalman W, Crowley-McHattan ZJ, Fransen J, Bennett KJM. Skill assessments in youth soccer: a scoping review. J Sports Sci 2022;40:667–95.
[8] Vaeyens R, Malina RM, Janssens M, et al. A multidisciplinary selection model for youth soccer. Br J Sports Med 2006;40:928–34.
[9] Wilson RS, Smith NMA, Ramos SP, et al. Dribbling speed along curved paths predicts attacking performance in match-realistic one versus one soccer games. J Sports Sci 2019;37:1072–9.
[10] Wilson RS, David GK, Murphy SC, et al. Skill not athleticism predicts individual variation in match performance. Proc R Soc B 2017;284:20170953.
[11] Gigerenzer G, Gaissmaier W. Heuristic decision making. Annu Rev Psychol 2011; 62:451–82.
[12] Schmidt FL, Hunter JE. The validity and utility of selection methods in personnel psychology. Psychol Bull 1998;124:262–74.
[13] Wald A. Sequential analysis. Wiley; 1947.
[14] Schmidt FL, Hunter JE, McKenzie RC, Muldrow TW. Impact of valid selection procedures on workforce productivity. J Appl Psychol 1979;64:609–26.
[15] Boudreau JW. Utility analysis for decisions in human resource management. In: Dunnette MD, Hough LM, editors. Handbook of industrial and organizational psychology, 2. Consulting Psychologists Press; 1991. p. 621–745.
[16] Cascio WF, Aguinis H. Applied psychology in human resource management. 7th ed. Pearson; 2011.
[17] Baker J, Wattie N, Schorer J. A proposed conceptualization of talent in sport. Psychol Sport Exerc 2018;43:27–33.
[18] Phillips E, Davids K, Renshaw I, Portus M. Expert performance in sport and the dynamics of talent development. Sports Med 2010;40:271–83.
[19] Sedeaud A, Difernand A, De Larochelambert Q, et al. Talent identification: time to move forward on estimation of potentials? Sports Med 2025;55:551–68.
[20] Till K, Baker J. Challenges and misconceptions in talent identification and development. J Appl Sport Psychol 2020;32:87–98.
[21] Bergkamp TLG, Niessen ASM, den Hartigh RJR, Frencken WGP, Meijer RR. Methodological issues in soccer talent identification research. Sports Med 2019;49: 1317–35.
[22] Bullock N, Gulbin JP, Martin DT, Ross A, Holland T, Marino F. Talent identification and deliberate programming in skeleton: ice novice to Winter Olympian in 14 months. J Sports Sci 2009;27(4):397–404. https://doi.org/10.1080/ 02640410802549751.
[23] Johnston K, Wang Y, Trace J, Baker J, Richard V, Roberts A. A Bayesian perspective on geographic influences in coach decision-making for athlete identification and selection. Sci Rep 2026;16:8234. https://doi.org/10.1038/ s41598-025-30791-y.
[24] Lakkaraju H, Kleinberg J, Leskovec J, Ludwig J, Mullainathan S. The selective labels problem: evaluating algorithmic predictions in the presence of unobservables. In: Proceedings of the ACM SIGKDD conference on knowledge discovery and data mining; 2017.
[25] Epstein JM. Why model? J Artif Soc Social Simul 2008;11(4):12.
[26] Wilson RS, Sparkes G, Waller L, et al. Quantitative talent identification reimagined: sequential testing reduces decision uncertainty. Appl Sci 2025;2025:9707.
[27] Wilson RS, Hunter AH, Camata TV, et al. Dribbling and passing performances predict individual success in small-sided soccer games. Int J Sports Sci Coach 2025. https://doi.org/10.1177/17479541251346902.
[28] Travassos B, Araújo D, Davids K, O'Hara K, Leitão J, Cortinhas A. Expertise effects on decision-making in sport are constrained by requisite response behaviours. Psychol Sport Exerc 2013;14:211–9.
[29] Wilson RS, Smith NMA, de Souza NM, Moura FA. Dribbling speed predicts goal-scoring success in a soccer training game. Scand J Med Sci Sports 2020;30:2070–7.
[30] Wilson RS, Hunter AH, Camata TV, et al. Simple and reliable protocol for identifying talented junior players. Scand J Med Sci Sports 2021;31:1647–56.
[31] Kelly AL, Williams CA, Jackson DT, Turnnidge J, Reeves MJ, Dugdale JH, et al. Exploring the role of socioeconomic status and psychological characteristics on talent development in an English soccer academy. Sci Med Footb 2024;8(3):251–9. https://doi.org/10.1080/24733938.2023.2213191.
[32] Hubli TV, Collins D, Khasnis U. Achieving from deprivation in sports: a scoping review on talent identification and development in sport in low socioeconomic context. Int Rev Sport Exerc Psychol 2026. https://doi.org/10.1080/ 1750984X.2026.2705272.
[33] Mahmood H, Day M, McPherson G, Datson N, Unnithan VB. Key stakeholder perceptions of the implementation of the English Football Association's girls' Emerging Talent Centre pathway. Int J Sports Sci Coach 2026;21(3):1453–66. https://doi.org/10.1177/17479541251370592.
[34] Hunter AH, Smith NMA, Camata TV, Crowther MS, Mather A, Melo de Souza NM, et al. Age-and size-corrected kicking speed and accuracy in elite junior soccer players. Sci Med Footb 2022;6(1):29–39.
[35] Meylan C, Cronin J, Oliver J, Hughes M. Talent identification in soccer: the role of maturity status on physical, physiological and technical characteristics. Int J Sports Sci & Coach 2010;5:571–92.
[36] Figueiredo AJ, Gonçalves CE, Coelho e Silva MJ, Malina RM. Characteristics of youth soccer players aged 11–14 years according to maturity status. Ann Hum Biol 2009;36(1):60–73.
[37] Camata TV, Hunter AH, Smith NMA, Crowther MS, Santiago PRP, Wilson RS. Using age- and size-corrected measures of technical skill to better assess youth soccer players. Appl Sci 2025;15:4658.