# Sports Talent Identification Using Cloud Computing

# Richard Meadows

# 12688825

# Research Proposal

# Research Methods and Professional Practice

# MSc Cyber Security

# University of Essex (Online)

# November 2022

Contents

[Introduction 3](#_Toc120313386)

[The Current State of Talent Identification 4](#_Toc120313387)

[The Use of Cloud Computing in Talent Identification 6](#_Toc120313388)

[Research Question 7](#_Toc120313389)

[Aims and Objectives 7](#_Toc120313390)

[Methodology 8](#_Toc120313391)

[Ethical Considerations and Risk Assessment 10](#_Toc120313392)

[Timeline of Activities 11](#_Toc120313393)

[References 13](#_Toc120313394)

## Introduction

Talent identification (TID) is important to sporting clubs, as those that excel are able to gain significant advantages over opponents. This is important as success brings a sense of wellbeing and improved self-esteem in athletes, staff and fans alike, leading to healthier environments around clubs. It can also have major financial implications, especially in elite level sport. Players picked from lower levels are significantly cheaper to acquire than those already established in higher echelons, and the development of young players is significantly cheaper than buying them in (European Club Association, 2012).

Athletes themselves also enjoy significant benefits. Those participating at lower levels of sport may miss out on state of the art training regimes, tailored dieting, performance monitoring and analysis, and earning potential that higher levels offer. Additionally, children that are not picked up by developed sports clubs miss out on quality training, discipline and habits that are essential in sport in their formative years, which may lead to them reaching a point in which they are unable to recover from. In many cases, it is possible to identify talent applicable to sport when subjects are not actively participating. Sport in this case brings several further positive aspects in to their lives including confidence, cognitive abilities, social skills, and physical and mental health (Coalter, 2005).

## The Current State of Talent Identification

In sports talent identification, evaluating individual athletes based on current performance is simplistic, especially at lower level sport. However due to its simplicity, it commonly leaves sporting clubs in direct competition with each other, and bids become unnecessarily high. In this light, emphasis can focus on recognised potential rather than current performance.

Another common approach is the one-size-fits-all approach, which discounts the late developers, and those that are uniquely talented or physically different. Understandably, height for example is often used in criteria for TID in sports such as basketball, high jump and volleyball, however long-term predictions cannot be made on physical characteristics alone, as these are subject to change over time (Abbott et al., 2005).

As well as physical characteristics, sports talent identification requires detailed analysis of technical attributes that make athletes more likely to display potential. Methods such as watching live or video analysis of participants in action, rather than trying to study a photograph for example, is beneficial, as is analysing performances in varying environmental conditions to show adaptability.

It is also important that these attributes are not solely relied upon, as the key factor in TID is not identifying talent simply on current performance, rather identifying factors which may or may not be limiting talent development, including behavioural factors such as a lack of self-confidence or mental focus. Additionally, it could be due to not being developed yet, and may be lying as unrecognised potential until later years (Abbott et al., 2005).

Athletics Training (N.D.) echo this, however concede that TID is easier for athletic events where physical prowess is the main requirement, and more difficult to identify in skill and strategy based events.

Humans also possess goal-driven mindsets on varying scales, something that is essential in every sport (Kugler et al., 1990). Although useful, these goals can often be overridden by stronger habitual behaviours and thus it may be easier to mould younger players than those that have already formed bad habits (Wood et al., 2021). On the other hand, the number of talented athletes accurately identified may increase with age as formative attributes develop with time, therefore the earlier the TID is sought, the riskier the calculation (Abbott et al., 2005).

Abbott et al. (2005) explain that individuals should not be evaluated on any singular component, instead other attributes (for example, speed or height) may be advantageously compensated for. One example is Lionel Messi who is considered by many to be the greatest footballer of all time and was diagnosed with a growth disorder as a child. He never grew taller than 5’7” (Carroll, 2013) and would hardly have been considered in TID based on physicality alone. Another example, Matthew Wolff, is renowned for his unorthodox golf swing, and it is therefore unlikely he would have been considered in TID based on technique alone, in fact it is likely that many golf clubs around the world would have encouraged him to completely change his technique. He went on to win the 3M Open in 2019 (PGA Tour, 2019).

Former football manager and Premier League champion, Arsène Wenger echoes this and describes how making a great player is analogous to the building blocks of a house, requiring technical, physical, tactical and mental aspects ([TopTekkers,](https://www.youtube.com/watch?v=t8EfQZdlf0Y) 2020).

## The Use of Cloud Computing in Talent Identification

As with most other fields, data collected by automated means is increasing in sport thanks to advancing technology (Brefeld & Zimmermann, 2017) and so projects in sport leveraging cloud technology is inevitable. Thus, although this is a relatively new field of research, Jauhiainen et al. (2019) use TID as an anomaly detection problem, and propose a support vector machine to analyse data for classification and regression analysis to support experts making TID decisions on fourteen year olds. However, this according to Arsène Wenger is too late, as he explains: “If you don't have technical ability at fourteen you can forget it” (TopTekkers, 2020).

[Khan](https://link.springer.com/article/10.1007/s41870-022-00984-z) et al. (2022) also proposed a multi-layered IoT / Fog / Cloud system to identify talent, using a list of attributes to measure against, including physical, technical and psychological aspects, and using cricket as a case study.

Finally, Louszada et al. (2016) proposed a system that can offer objective decisions on TID in association football, however only uses physical and technical data, and discounts psychological data.

## Research Question

At present, there is no research to find if cloud computing can be utilised to discover potential talent in children as young as nine years old, or in others in lower level sport or that do not currently participate (hidden gems).

This research proposal therefore seeks to find answers to the following question and sub-questions:

How can cloud computing be used to identify sporting talent?

* How can cloud computing be used to identify talent in nine year old children?
* Can the same model identify undiscovered potential talent in others?

To the best of the author’s knowledge, there is no research aimed at finding the answers to these questions.

## Aims and Objectives

Thus, this research proposes the following aim:

* To explore cloud technology’s ability to predict future high performing athletes based on expert knowledge and empirical data inputs

And the following objectives:

* Determine, based on historical data and expert opinion, which factors contribute as indicators of athletes being successful in sport
* Using cloud computing, analyse the data to produce meaningful results
* Discover if the model is transferrable for both children aged nine and hidden gems

## Methodology

As this must only include secondary data, data is gathered using further literature reviews including searches in research databases such as Google Scholar, Research Gate, the University of Essex Online library, and the public Internet. Unless particularly relevant, only secondary research from 2016 onwards is used to maintain relevance.

Primary research would require a significant number of participants for mass data gathering, including athletes and experts in the field of multiple respective sport disciplines. Physical, technical and psychological data would need to be collected through a mixture of qualitative and quantitative methods to garner required information and this additionally provides opportunity for triangulation (Saunders et al., 2019). Data collection methods include questionnaires to enable fast and wide distribution, so mass data can be gathered from many athletes. Questionnaires would not be useful for harvesting information from experts however, as an in-depth knowledge base for a cloud engine is required, and so semi-structured interviews and focus groups would need to be conducted. Questionnaires are not sufficient due to their limited ability to delve deeper into what may be a talent indicator in their respective sport. Further physical statistics can be gleaned from secondary sources such as the one from United Rugby Championship (N.D) to support the expert opinion of what makes a successful player. Finally, video analyses to capture real-time activity are particularly useful for technical aspects that would be difficult to transcribe on a written report or interview.

The research would be exploratory in design, and requires inductive reasoning as it looks for patterns (Saunders et al., 2019). Data would then need to be passed into the cloud where it could be stored in two locations: a knowledge base created from expert opinion, and a database containing athlete data, before weighting and analysis could be performed by an application to provide an estimate on the likelihood of talent and thus be used to support decision making. Athlete data would be collected from a significant number of nine year old children, and another group of random older participants. The study would include any gender, and abled and disabled participants. By leveraging the cloud in this way, data can be accessed from anywhere in the world, as well as having the possibility of handling big data, something that would be far too expensive for researchers using an on-premise model.

The study would be longitudal as results would not be known until later in the participant’s career. The qualitative data would need to be categorised and coded before it could be analysed. Finally, a normal distribution is expected and so a t-test could be performed using a p-value of 0.05 to determine confidence that results from the quantitative data sample are applicable to the population (Berenson et al., 2015).

## Ethical Considerations and Risk Assessment

There are no ethical considerations or risk assessment required as this research proposal uses only secondary research. Additionally, there is no requirement for any artefacts in the submission.

If this research were to be conducted using primary research, there would be significant ethical considerations due to minors being amongst the participants. Participants in either case would need anonymising to protect identities, and names could be easily replaced by numbers. Consent would need to be given by each participant, or by a guardian for children based on diminished autonomy (Dittrich & Kenneally, 2012). Additional approval may be required for research on children, such as a Disclosure and Barring Service check, and always having adult gatekeepers present when accessing children (UK Research and Innovation, 2021).

Full written details of the research would be given to each participant or guardian, with a signed guarantee that data would not be altered, only be used for purposes in which it is collected, and would not be published without consent. Participants or guardians would also be advised that the research is voluntary and they are free to leave at any point (Vanclay et al., 2013), and finally, the research would require approval by the University of Essex Online’s Ethics Committee.

## Timeline of Activities

Table 1 describes the timeline of activities for the research proposal. The literature search takes four days, the literature review takes three days, before another two days to analyse and evaluate, and finally three days to complete the presentation and transcript.

In order to get primary research, this would take several months to complete as time would be required to get ethical approval, design data collection questionnaires, interviews and focus group frameworks, several weeks would be required to obtain necessary equipment and investigate and construct the relevant cloud computing technologies, and several weeks to organise access to participants. It would take many more weeks to conduct the survey as it would require multiple visits to multiple locations to get the significant amount of data required. It would take several further weeks to analyse the data after uploading to the cloud, and conduct statistical testing, and finally several more weeks to produce a report based on the findings. A total of twenty-eight weeks is expected for completion of the primary research project.

Table 1: Timeline of Activities

|  |  |
| --- | --- |
| **Task** | **Duration** |
| Literature search | 4 days |
| Literature review | 3 days |
| Analyse and evaluate | 2 days |
| Complete presentation and transcript | 3 days |

This concludes the end of the research proposal and I would like to thank you for listening.

## References

Abbott, A., Button, C., Pepping, G., Collins, D. (2005) Unnatural selection: Talent identification and development in sport. *Nonlinear Dynamics Psychology and Life Sciences* 9(1): 61-88. Available from: <https://www.researchgate.net/publication/8102048_Unnatural_selection_Talent_identification_and_development_in_sport>

Athletics Training. (N.D.) Talent Identification. Available from: <http://www.athletics-training.com/articles/talent-identification.html> [Accessed 22 November 2022].

Berenson, M. L., Levine, D. M., & Szabat, K.A. (2015) Basic Business Statistics: Concepts and Applications. 3rd ed. Harlow: Pearson Education. Available from: https://web-s-ebscohost-com.uniessexlib.idm.oclc.org/ehost/ebookviewer/ebook/bmxlYmtfXzE0MTk1NDVfX0FO0?sid=511d3386-00e8-4ae9-8f70-de466e9d4c89@redis&vid=0&format=EB&rid=1 [Accessed 24 November 2022].

Brefeld, U. & Zimmermann, A. (2017) Guest editorial: Special issue on sports analytics. *Data Mining and Knowledge Discovery volume* 31(6): 1577–1579. DOI: <https://doi.org/10.1007/s10618-017-0530-1>

Carroll, W. (2013) Lionel Messi and HGH: The Truth About the Best Footballer in the World. Available from: <https://bleacherreport.com/articles/1492546-lionel-messi-and-hgh-the-truth-about-the-best-footballer-in-the-world> [Accessed 23 November 2022].

Coalter, F. (2005) *The Social Benefits of Sport.* Glasgow: SportScotland.

Dittrich, D. & Kenneally, E. (2012) The Menlo Report. *IEEE Security & Privacy* 10(2): 71-75. DOI: <https://doi.org/10.1109/MSP.2012.52>

European Club Association. (2012) *Report on Youth Academies in Europe*. Available from: <https://www.ecaeurope.com/media/2730/eca-report-on-youth-academies.pdf> [Accessed 25 November 2022].

Jauhiainen, S., Äyrämö, S., Forsman, H. & Kauppi, J.P. (2019) Talent identification in soccer using a one-class support vector machine. *International Journal of Computer Science in Sport* 18(3): 125-136. DOI: https://doi.org/10.2478/ijcss-2019-0021

Khan, N.J., Ahamad, G. & Naseem, M. (2022) An IoT/FOG based framework for sports talent identification in COVID-19 like situations. *International Journal of Information Technology* 14(5): 2513–2521. DOI: <https://doi.org/10.1007/s41870-022-00984-z>

Kugler, P. N., Shaw, R. E., Vincente, K.J. & Kinsella-Shaw, J. (1990) Inquiry into intentional systems I: Issues in ecological physics. *Psychological Research* 52(2): 98–121. DOI: <https://doi.org/10.1007/BF00877518>

Louzada, F., Maiorano, A. C. & Ara, A. (2016) iSports: A web-oriented expert system for talent identification in soccer. Expert Systems with Applications 44: 400-412. DOI: <https://doi.org/10.1016/j.eswa.2015.09.007>

PGA Tour. (2019) Wolff wins 3M Open in third professional start. Available from: <https://www.pgatour.com/daily-wrapup/2019/07/07/matthew-wolff-wins-3m-open-in-third-professional-start.html> [Accessed 23 November 2022].

Saunders, M., Lewis, P., Thornhill, A. (2019) *Research Methods for Business Students.* Harlow: Pearson Education. Available from: https://ebookcentral-proquest-com.uniessexlib.idm.oclc.org/lib/universityofessex-ebooks/reader.action?docID=5774742&ppg=205 [Accessed 23 November 2022].

[TopTekkers](https://www.youtube.com/watch?v=t8EfQZdlf0Y). (2020) ARSÈNE WENGER: If you don't have technical ability at 14 you can forget it. Available from: <https://www.youtube.com/watch?v=t8EfQZdlf0Y> [Accessed 23 November 2022].

UK Research and Innovation. (2021) Research with children and young people. Available from: <https://www.ukri.org/councils/esrc/guidance-for-applicants/research-ethics-guidance/research-with-children-and-young-people> [Accessed 23 November 2022].

United Rugby Championship (N.D.) STATS. Available from: <https://www.unitedrugby.com/statistics/players> [Accessed 25 November 2022].

Vanclay, F., Baines, J. T. & Nicholas Taylor, C. (2013) Principles for ethical research involving humans: ethical professional practice in impact assessment Part I. *Impact Assessment and Project Appraisal* 31(4): 243-253. DOI: <https://doi-org.uniessexlib.idm.oclc.org/10.1080/14615517.2013.850307>

Wood, W., Mazar, A. & Neal, D. T. (2021) Habits and Goals in Human Behavior: Separate but Interacting Systems. *Perspectives on Psychological Science* 17(2): 590-605. DOI: https://doi.org/10.1177/1745691621994226