

Determining Youth Profile using Principle Component Analysis for Identifying Talent in Sports

Siti Musliha Mat-Rasid, Mohamad Razali Abdullah, Hafizan Juahir, Ahmad Bisryi Husin Musawi Maliki, Rabi Muazu Musa, Norlaila Azura Kosni, Muhammad Rabani Hashim, Vijayamurugan Eswaramoorthi, Nasree Najmi

Abstract: This study attempts to determine youth profile in sports talent identification program. Data of anthropometric and physical fitness included power, agility, speeds, flexibility, strength and endurance were obtained from 600 participants in a sports talent identification program aged 13-15 years. Data analyses were carried out using multivariate analysis of principal component analysis (PCA). PCA revealed three principal components with 71.5% total variation for this studied group. The first component consists of high factor loading in speed variables (10 meter run, 20 meter run, 40 meter run) and endurance (predicted $\dot{V}O_2$ max). The second component was constituted by anthropometric variables (weight, height, sitting height and armspan). While third component contains flexibility variable (sit and reach). These selected variables of anthropometric and fitness are, therefore, revealed as the essential attributes those must be prioritized for a talent scouting in sports. Present results had demonstrated youth profile that capable of providing an information that could help coaches in decision making during athlete selection in sports talent identification.

Index Terms: Principal Component Analysis, Profiling, Talent Identification

I. INTRODUCTION

The interest of talent identification (TI) in sports is fully related to the coaching and current sports science studies [1]. TI in sport is generally intended to identify talent at an early age in order to provide excellent development opportunities in the future [2].

Most of the TI had measured and selected young athletes based on the performance of genetically driven by a specific

component [3]. It is because talent development and identification is a multifactorial process involving many characteristics. As an example, a country that succeeds in sports such as Australia [4] decided on the basis of the physical components of talent and performance, which is believed, to be linked to sport excellence. In addition to physical factors, the athlete's body, technical characteristics, mental state, and external environment will affect the athlete's sports performance to some extent. Among the characteristics, anthropometrical and physiological factors contributed to the performances of athletes [5-7]. Evident of differences were occurred when comparing the physiological and anthropometric qualities explanatory of talent. A more adequate study had included psychology [8,9] and technical skill [10].

In line with recent studies those has explored scientific approach to be applied in athlete selection process to replace the traditional scheme [11-13], this study will apply the multidimensional analysis of PCA to provide youth profile in talent identification program. The data of anthropometric measurement and physical fitness were obtained among male players in a sport development program. Since the physical variations among adolescent are high, the identification, quantification and implementation of these anthropometric and fitness components attribute in selection decisions have a significant impact on assessing new talent in sports [14]. Proposed approach in this study can help coaches in designing appropriate training for each group that has different profile.

II. MATERIALS AND METHODS

A. Sample of Study

The anthropometric measurements and fitness test data in this study were obtained from 600 male youth (age 13.67 ± 0.57 year) in a sports identification program in Terengganu, Malaysia. The raw data was converted into a single matrix formed by 13 variables (4 anthropometric components and 9 physical tests) with 600 male youth. All the participants were informed to sign an informed consent form before the day of the program.

Revised Manuscript Received on April 25, 2019.

Siti Musliha Mat-Rasid, Faculty of Applied Social Science, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.

Mohamad Razali Abdullah, Faculty of Applied Social Science, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.

Hafizan Juahir, East Coast Environmental Research Institute, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.

Ahmad Bisryi Husin Musawi Maliki, Faculty of Applied Social Science, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.

Norlaila Azura Kosni, Faculty of Applied Social Science, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.

Rabi Muazu Musa, Centre for Foundation & Liberal Studies, Universiti Malaysia Terengganu, Terengganu, Malaysia.

Muhammad Rabani Hashim, Faculty of Applied Social Science, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.

Vijayamurugan Eswaramoorthi, Faculty of Health Science, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.

Nasree Najmi, Terengganu State Sports Council, Terengganu, Malaysia.

B. Anthropometric Measurements and Fitness Test

Anthropometric measurement involved in this study were included weight, standing height, sitting height and armspan. While fitness test of standing broad jump, 10-m run, 20-m run, 40-m run, vertical jump, sit and reach, max push up, one minute sit up and maximal multistage 20-m shuttle run test) were carried out. The participants were divided into groups with an adequate recovery period between tests (up to 3 minutes of rest).

C. Data Analysis

Preprocessing data: A matrix set of male group contains 7800 matrices data (13 variables × 600 participants) were computed in this study. For matrices that have very small amounts of data lost (~ 3%) than the overall data recorded, the nearest neighboring method can be used [15,16]. This method examines the distance between each point and its nearest point. The nearest neighboring method is the simplest method, where the endpoint of the gap is used as an estimate of all missing values [17,18].

Principal Component Analysis (PCA): PCA is a common technique for finding patterns in data of high dimension [19,7]. The idea behind of PCA is by which numbers of correlated variables are transformed into a smaller number of uncorrelated variables. The previous study applied PCA to provide indications for athlete’s classification and identified potentially important technical differences between excellent and poor skilled athletes [20-22]. The factor loadings tell about how much the variable has contributed to the factor; the larger the factor loading the more the variable has contributed to that factor [23]. Factor loadings are very similar to weights in multiple regression analysis, and they represent the strength of the correlation between the variable and the factor [24]. It can be used to compress a high dimensional dataset into a lower dimensional dataset. Recent study also revealed PCA is particularly useful when data on a number of useful variables has been gathered, and it is plausible that there is some redundancy in those variables [25-29].

III. RESULT

Table 1 shows the descriptive statistics of anthropometric measurement and physical fitness among 600 male participants. It shows minimum and maximum scores, mean and standard deviations. To reiterate, PCA was employed on the data set to determine youth profile by examining the most influence parameters of anthropometric and physical fitness. Varimax rotation (VFs) with absolute values greater than 0.6 have been standardized as selection thresholds due to the fact that these values are relatively moderate and stable, indicating a moderate and strong load on the extracted factor. It has revealed that the ten variables are satisfied with the threshold of loading 0.60 factors as shown in Table 2. Three principal components were obtained with eigenvalues larger than 1 summing almost 71.48% of the total variance in the data set. Among the factor loadings after varimax, VF1 accounted about 43.32% of the data variability. It showed

Table 1: Descriptive statistics of male participants involved in talent identification program

| Variables | Min | Max | Mean | SD |
|---|-------|-------|-------|------|
| Wg (kg) | 25.8 | 68.8 | 44.9 | 9.6 |
| Hg (cm) | 129.0 | 180.0 | 154.5 | 9.2 |
| SH (cm) | 65.0 | 93.0 | 78.0 | 5.2 |
| As (cm) | 130.6 | 187.0 | 157.5 | 10.6 |
| VJ (cm) | 25.0 | 60.0 | 42.6 | 6.5 |
| SBJ (cm) | 110.0 | 236.0 | 172.5 | 25.2 |
| 10MR (s) | 1.62 | 2.5 | 2.05 | 0.17 |
| 20MR (s) | 2.85 | 4.36 | 3.59 | 0.30 |
| 40MR (s) | 5.05 | 8.35 | 6.70 | 0.63 |
| SAR (cm) | 15.5 | 45.0 | 30.2 | 5.49 |
| MPU (no.) | 1 | 38 | 16 | 9 |
| 1MSU (no.) | 1 | 33 | 17 | 8 |
| VO ₂ max (ml.kg ⁻¹ .min ⁻¹) | 15.40 | 48.50 | 32.57 | 6.82 |

Wg: Weight, Hg: Height, Sitting Height, As: Armspan, VJ: Vertical Jump, SBJ: Standing Broad Jump, 10MR:10 Meter Run, 20MR: 20 Meter Run, 40MR: 40 Meter Run, SAR: Sit and Reach, MPU: Max Push Up, 1MSU: 1 Minute Sit Up.

strong positive loadings on physical fitness test included 10 meter run, 20 meter run, 40 meter run and predicted VO₂ max. The second component VF2 accounted 20.23% of the data variability, with a clear contribution in an anthropometric component of weight, height, sitting height and armspan. Otherwise, there are obtained VF3, explaining 7.93% of the total variance, has strong positive loadings on flexibility test (sit and reach) among this studied group. These variables are then classified as an important component which is particularly needed to identify the dominant factor for talent in sports.

IV. DISCUSSION

The result obtained from the current study implies that weight, height, sitting height, armspan, 10 meter run, 20 meter run, 40 meter run, predicted VO₂ max and sit and reach were significant variables were highly contributed in fitness youth performance. Current finding concordance to the previous study that emphasized anthropometric component and physical fitness were contributed high variation regarding the performance of sedentary youth [30,31]. The first component from PCA has revealed physical fitness of power, speed and endurance as major physical fitness must be prioritized when assessing current performance of youth. This has been supported by concrete evident when significant relationships between physical measured with both agility and explosively was noted in the Melanesian players [14]. Speed is classically defined as the shortest time required for an object to move along a fixed distance, which is the same as velocity, but without specifying the direction [32]. In practical terms, it refers to the ability to move the body as quickly as possible over a set distance.



Table 2: Factor loadings after varimax rotation.

| Variables | VF1 | VF2 | VF3 |
|--|---------------|--------------|--------------|
| Wg (kg) | -0.061 | 0.897 | -0.101 |
| Hg (cm) | 0.279 | 0.905 | 0.088 |
| SH (cm) | 0.206 | 0.907 | 0.118 |
| As (cm) | 0.239 | 0.896 | 0.115 |
| VJ (cm) | 0.554 | 0.394 | 0.413 |
| SBJ (cm) | 0.679 | 0.253 | 0.500 |
| 10MR (s) | -0.898 | -0.173 | -0.023 |
| 20MR (s) | -0.907 | -0.203 | -0.073 |
| 40MR (s) | -0.904 | -0.223 | -0.165 |
| SAR (cm) | 0.008 | 0.343 | 0.624 |
| MPU (no.) | 0.321 | -0.178 | 0.586 |
| 1MSU (no.) | 0.188 | 0.108 | 0.580 |
| VO _{2 max} (ml.kg ⁻¹ .min ⁻¹) | 0.601 | -0.239 | 0.390 |
| Eigenvalues | 5.632 | 2.630 | 1.030 |
| Variability (%) | 43.324 | 20.234 | 7.925 |
| Cumulative % | 43.324 | 63.558 | 71.483 |

Wg: Weight, Hg: Height, Sitting Height, As: Armspan, VJ: Vertical Jump, SBJ: Standing Broad Jump, 10MR: 10 Meter Run, 20MR: 20 Meter Run, 40MR: 40 Meter Run, SAR: Sit and Reach, MPU: Max Push Up, 1MSU: 1 Minute Sit Up.

First component from PCA also has revealed that the male youth are attributable to high capacity in cardiovascular endurance (VO_{2 max}). This result is similar to the previous study which indicated that cardiovascular endurance provides athletes with energy to carry out their sporting activities with vigor [33]. Moreover, cardiovascular endurance is a compulsory component for contribution in any type of sports. It is an important fundamental attribute of athletic performance since the heart controls the oxygen flow to all the working muscles [34]. Therefore, cardiovascular endurance has a strong impact on athlete's performance in various kinds of sports. This finding is related to the previous investigation which reported that cardiovascular fitness improves the ability of the heart and lungs to supply oxygen-rich blood to the working muscles which in turn accelerates the actions of the athletes to act in response to the fitness requirement of the sport he or she participates in [35].

Amongst the various factors contributing to the athlete's performance, the most important is the athlete's physique, but at the same time the athlete's body, technical characteristics, mental state, and external environment will affect the athlete's sports performance to some extent [36]. Besides, anthropometrics have the major advantage of studying the physique of different populations, athlete and non-athletes. There was true evident that measurement of anthropometric variables can create a profile of athlete within specific sports [37]. Conversely, such finding from PCA discloses that the participant varies from the other participant based on their performance on flexibility test (sit and reach). Furthermore, there is scientific evidence that demonstrates the prevalence of injury decreases on athletes who possess greater flexibility as opposed to the athletes with lesser flexibility [38]. flexibility can enhance the ability of the athletes to move unconstrained through a wider range of motion. Nevertheless, the current finding is also supported by

another study which stated that flexibility is one of the main fitness parameters imperative for high performance athlete [39]. Since different sports have different anthropometric characteristics, the findings could help coaches to select players for the professional level according to the playing position.

V. CONCLUSION

As the conclusion, multidimensional analysis using multivariate approach successfully demonstrated youth profile that could likely enables the coaches to select athlete and fit them to the most suitable sport. It could also serve an assessment tool for evaluating the success of their training program as well as placement into an appropriate training group. Similarly, the results of this study imply that the application of advanced statistical method is vital in identifying essential performance parameters in the talent identification program which can save time, energy and cost. Coaches or other experts in the field could, in the future, find it useful to follow and improve performance, through the training process which could be most appropriate for specific sport.

ACKNOWLEDGMENT

The authors would like to thank Research and Development Management Unit, University of Sultan Zainal Abidin (UniSZA) and Terengganu State Sports Council for their support in this research.

REFERENCES

1. N. Durand-Bush, J. H. Salmela. "The development of talent in sport", pp. 269-289. 2001
2. K. A. Ericsson, R. T. Krampe, C. Tesch-Römer. "The role of deliberate practice in the acquisition of expert performance." Psychological review, vol. 100, no. 3, pp. 363-, 1993
3. F. Louzada, A. C. Maiorano, and A. Ara, "iSports: A web-oriented expert system for talent identification in soccer," Expert Systems with Applications, vol. 44, pp. 400-412, 2016.
4. D. Hoare. Talent identification: A review and update. Sports Coach, Spring 1998;32±33.
5. O. Hue, O. Galy, S. Blonc, and C. Hertogh, "Anthropometrical and Physiological Determinants of Performance in French West Indian Monofin Swimmers: A First Approach," International Journal of Sports Medicine, vol. 27, no. 8, pp. 605-609, 2006.
6. C. T. Woods, A. Cripps, L. Hopper, and C. Joyce, "A comparison of the physical and anthropometric qualities explanatory of talent in the elite junior Australian football development pathway," Journal of Science and Medicine in Sport, vol. 20, no. 7, pp. 684-688, 2017.
7. M. R. Abdullah, A. B. H. M. Maliki, R. M. Musa, N. A. Kosni, H. Juahir, and M. Haque, "Multi-Hierarchical Pattern Recognition of Athlete's Relative Performance as A Criterion for Predicting Potential Athletes," Journal of Young Pharmacists, vol. 8, no. 4, pp. 463-470, 2016.
8. M. Humara. "Personnel selection in athletic programs." Athletic insight 2, no. 2, 2000.
9. M. R. Abdullah, R. M. Musa, N. A. Kosni, A. B. H. M. Maliki, and M. Haque. "Profiling and Distinction of Specific Skills Related Performance and Fitness Level between Senior and Junior Malaysian Youth Soccer Players." International Journal of Pharmaceutical Research, vol. 8, no. 3, pp. 64-71, 2016.
10. M. R. Abdullah, A. B. H. M. Maliki, R. M. Musa, N. A. Kosni, H. Juahir, and S. B. Mohamed, "Identification and Comparative Analysis of Essential Performance Indicators in Two Levels of Soccer Expertise," International Journal on Advanced Science, Engineering and Information Technology, vol. 7, no. 1, pp. 305-314, 2017.



11. A. Maszczyk, A. Gołaś, P. Pietraszewski, R. Rocznik, A. Zajac, and A. Stanula, "Application of Neural and Regression Models in Sports Results Prediction," *Procedia - Social and Behavioral Sciences*, vol. 117, pp. 482–487, 2014.
12. A. Maszczyk, A. Zajac, and I. Ryguła, "A neural network model approach to athlete selection," *Sports Engineering*, vol. 13, no. 2, pp. 83–93, 2010.
13. A. B. H. M. Maliki, M. R. Abdullah, H. Juahir, W. S. A. W. Muhamad, N. A. M. Nasir, R. M. Musa, S. M. Mat-Rasid, A. Adnan, N. A. Kosni, F. Abdullah, and N. A. S. Abdullah, "The role of anthropometric, growth and maturity index (AGaMI) influencing youth soccer relative performance," *IOP Conference Series: Materials Science and Engineering*, vol. 342, p. 012056, 2018.
14. O. Galy, P. Zongo, K. Chamari, E. Michalak, A. Dellal, C. Castagna, and O. Hue, "Anthropometric and physiological characteristics of Melanesian futsal players: a first approach to talent identification in Oceania," *Biology of Sport*, vol. 32, no. 2, pp. 135–141, 2014.
15. M. R. Abdullah, R. M. Musa, A. B. H. M. Maliki, N. A. Kosni, and P. K. Suppiah, "Development of tablet application based notational analysis system and the establishment of its reliability in soccer". *Journal of Physical Education and Sport*, vol. 16, no. 3, pp. 951-956, 2016.
16. S. M. Mat-Rasid, M. R. Abdullah, H. Juahir, R. M. Musa, A. B. H. M. Maliki, A. Adnan, N. A. Kosni, V. Eswaramoorthi, and N. Alias, "Effect of Birth Month Cut-Off Point with Fitness-Motor Performance on Malaysian Boys and Girls Relative Age". *Man In India*, vol. 97, no. 22, pp. 397-408, 2011.
17. M. H. Abdullah, H. Juahir, F. Lananan, M. K. A. Kamarudin, A. Ghazali, A. Azemin, N. A. Wahab, S. M. Mat-Rasid, M. H. M. Saad, "Pattern recognition of Melaleuca cajuputi powell essential oils fingerprint in Terengganu, Malaysia using chemometric approach". *International Journal of Engineering and Technology (UAE)*, vol. 7, no. 14, pp. 132-138, 2018.
18. A. Norzaida, S. M. Mat-Rasid, M. D. Zalina, A. H. Syafrina, "Probability distributions comparative analysis in assessing rainfall process in time and space". *International Journal of Civil Engineering and Technology*, vol. 8, no. 10, pp. 1679-1688, 2017.
19. J. Sarkar, S. Saha, and S. Agrawal, "An Efficient Use of Principal Component Analysis in Workload Characterization-A Study," *AASRI Procedia*, vol. 8, pp. 68–74, 2014.
20. R. M. Musa, M. R. Abdullah, H. Juahir, V. Eswaramoorthi, N. Alias, M. R. Hashim, and A. S. F. Alnamat, "An Exploratory Study of Personality Traits and Psychological Coping Skills on Archery Performance," *Indian Journal of Public Health Research & Development*, vol. 10, no. 3, pp. 630-635, 2019.
21. G. Donà, E. Preatoni, C. Cobelli, R. Rodano, and A. J. Harrison, "Application of functional principal component analysis in race walking: An emerging methodology," *Sports Biomechanics*, vol. 8, no. 4, pp. 284–301, 2009.
22. M. S. N. A. Hassan, A. Z. Azhar, N. Kamaruddin, and A. Mohamad, "Perspective of Malaysian Youths towards Homosexuals," *Indian Journal of Public Health Research & Development*, vol. 10, no. 4, pp. 747-751, 2019.
23. H. H. Harman, *Modern factor analysis*. Chicago: The University Press, 1976.
24. P. Kline, *An easy guide to factor analysis*. London: Routledge, 2008.
25. V. Eswaramoorthi, M. R. Abdullah, R. M. Musa, A. B. H. M. Maliki, N. A. Kosni, N. B. Raj, N. Alias, H. Azahari, S. M. Mat-Rashid, and H. Juahir, "A multivariate analysis of cardiopulmonary parameters in archery performance," *Human Movement*, vol. 19, no. 4, pp. 35–41, 2018.
26. A. B. H. M. Maliki, M. R. Abdullah, H. Juahir, F. Abdullah, N. A. S. Abdullah, R. M. Musa, S. M. Mat-Rasid, A. Adnan, N. A. Kosni, W. S. A. W. Muhamad, and N. A. M. Nasir, "A multilateral modelling of Youth Soccer Performance Index (YSPI)," *IOP Conference Series: Materials Science and Engineering*, vol. 342, p. 012057, 2018.
27. H. M. Zolkipli, H. Juahir, G. Adiana, N. Zainuddin, A. B. H. M. Maliki, M. E. Toriman, Mazlin Mokhtar, and R. Elfriti, "Spatial assessment of Selangor, Malaysia water treatment plant performance using chemometric technique". *International Journal of Engineering and Technology (UAE)*, vol. 7, no. 3.14 Special Issue 14, pp. 139-146, 2018.
28. H. M. Zolkipli, H. Juahir, G. Adiana, N. Zainuddin, A. Ismail, A. B. H. M. Maliki, N. I. Hussain, M. K. A. Kamarudin, M. E. Toriman, and Mazlin Mokhtar, "Spatial assessment and the most significant parameters for drinking water quality using chemometric technique: A case study at Malaysia water treatment plants". *International Journal of Engineering and Technology (UAE)*, vol. 7, no. 3.14 Special Issue 14, pp. 115-122, 2018.
29. M. R. Abdullah, N. Alias, A. B. H. M. Maliki, R. M. Musa, and L. A. Kosni, H. Juahir, "Unsupervised pattern recognition of physical fitness related performance parameters among Terengganu youth female field hockey players." *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, no. 1, pp. 100-105, 2017.
30. M. A. G. @ Charles, M. R. Abdullah, R. M. Musa, N. A. Kosni, and A. B. H. M. Maliki, "The Effectiveness of Traditional Games intervention program in the Improvement of Form One School-Age Children's Motor Skills Related Performance Components," *Movement, Health & Exercise*, vol. 6, no. 2, pp. 925-930, 2017.
31. N. I. Yusoff, M. R. Abdullah, H. Juahir, J. L. F. Lee, S. M. Mat-Rasid, N. A. Kosni, and M. K. Zawi, "The Effect of Residence Area on Motor Skill Development among Children," *Indian Journal of Public Health Research & Development*, vol. 10, no. 3, pp. 614-618, 2019.
32. E. Harman, J. Garhammer, and C. Pandorf, "Administration, scoring, and interpretation of selected tests". *Essentials of strength training and conditioning*, vol. 13, pp. 287-317, 2008.
33. C. Stangier, T. Abel, C. Hesse, S. Claes, J. Mierau, W. Hollmann, and H. K. Strüder, "Effects of Cycling vs. Running Training on Endurance Performance in Preparation for Inline Speed Skating," *Journal of Strength and Conditioning Research*, vol. 30, no. 6, pp. 1597–1606, 2016.
34. T. Miyamoto, H. Kamada, A. Tamaki, and T. Moritani, "Low-intensity electrical muscle stimulation induces significant increases in muscle strength and cardiorespiratory fitness," *European Journal of Sport Science*, vol. 16, no. 8, pp. 1104–1110, 2016.
35. S. A. M. Fenton, J. L. Duda, and T. Barrett, "Inter-participant variability in daily physical activity and sedentary time among male youth sport footballers: independent associations with indicators of adiposity and cardiorespiratory fitness," *Journal of Sports Sciences*, vol. 34, no. 3, pp. 239–251, 2015.
36. Z. Taha, R. M. Musa, A. P. P. A. Majeed, M. R. Abdullah, M. A. Zakaria, M. M. Alim, J. A. M. Jizat, and M. F. Ibrahim, "The identification of high potential archers based on relative psychological coping skills variables: A Support Vector Machine approach," *IOP Conference Series: Materials Science and Engineering*, vol. 319, p. 012027, 2018.
37. I.-M. Popovici, L. Popescu, and L.-E. Radu, "Evaluation of some physical fitness characteristics at age 11 to 13," *Timisoara Physical Education and Rehabilitation Journal*, vol. 9, no. 17, pp. 24–28, 2016.
38. Z. Taha, M. Haque, R. M. Musa, M. R. Abdullah, A. B. H. M. Maliki, N. Alias, and N. A. Kosni, "Intelligent Prediction of Suitable Physical Characteristics Toward Archery Performance Using Multivariate Techniques." *Journal of Global Pharma Technology*, vol. 9, no. 7, pp. 44-52, 2009.
39. L. Brunkhorst, and H. Kielstein, "Comparison of anthropometric characteristics between professional triathletes and cyclists." *Biology of sport*, vol. 30, no. 4, pp. 269-273, 2013.