ABSTRACT

In purpose to avoid subjectivity and bias selection, this study would propose to carry out the athlete selection according to their particular birth-month quartile (Q1: Jan-March; Q2: Apr-June, Q3: July-Sept and Q4: Oct-Dec). This study thus combined statistical analysis of Cluster Analysis (CA) and Discriminant Analysis (DA) to determine the relative age effect in physical attributes and motor fitness. CA assigned the children into three groups of high, moderate and low performance. The given groups were considered as predefined classes for DA. Result found that there are discriminated variables that significantly different in each of birth-month quartile. Coaches may consider the proposed parameters as a baseline in decision making of children selection in talent identification program to reduce RAEs.

Keywords: relative age effect; physical attribute; motor fitness; talent identification.

Author Correspondence, e-mail: razali896@yahoo.com
doi: http://dx.doi.org/10.4314/jfas.v9i2s.34
1. INTRODUCTION

The relative age effect (RAE) means the whole dissimilarity among age group of individual [1]. The particular variation may cause significant differences in performance of athlete. Obviously, the RAE had contributed to the majority of selected athlete were comes from early birth-month athletes [2]. Otherwise, athletes who born in the later months of the year would be drop out to be selected. It had been proved that for every two athletes in the fourth birth-month quartile, there are three or more elite athletes born in the first birth-month quartile [3]. Such ratio sufficiently confirmed there is high possibility of athletes born in the earlier month to be selected compare to the athletes born in the later month of the year. The problematic of RAE becomes crucial when goes to athlete selection process among adult players [2, 4] and youth players [2, 4-5]. Players who born early in the recruitment year were definitely have more chances to be selected as a part of elite or national teams. In contrast, players born later are seemly to be deselected. Unfortunately, they are sometimes being released from the clubs [2, 6].

The RAE could obtain the differences in growth and development among athletes. That is main cause of selecting the talented athletes were prioritized to the more mature athletes. There are evident, athletes born in the earlier months of the year are more potential to gain earlier maturation [7]. This study pointed out the improvement of vital components such as speed, strength and muscular endurance by this group which is contradicting to their younger counterparts. In addition, they also have an advantage when challenging in sport which may result in superior sport experience as a function of age [5]. The players with greater components physically can be seen more mature. As the result, most of them will be selected as talented player and commonly have chance to play more frequently in matches. Such exposures give benefits on older players to accumulate more training period, get more attention by coach and increase their self-capability and effectiveness level in contrast to the younger counterparts [3, 5]. These advantages will contribute to the improvement of performance in sports expertise as well as explain why there are older athletes in talent identification.

The RAE may lead to subjective and unfair decision making in talent detection and
identification process. The RAE obviously cannot be eliminated but struggling find solution on minimizing the effect. There are some suggestions has been published to resolve the RAE. Formerly, the common approach were grouping athlete according annual age had been carried out [8]. Such approach promoting a change in the age group cutoff date, rotating cut-off dates from year to year or shifting age grouping bandwidths. However, it was found that changing the cut-off dates only lead to a transfer of RAE [9]. After that, other recommendations had been proposed in order to reduce RAE [3]. First, the athlete selection should by grouping them based on their physical classification such as height and weight. In addition, there is more reasonable and less risk solution that is to holdup the selection process until the athletes really have reached puberty and maturity. In this case, talent identification program must be reconsidered the need for early elections, intensive training and the level of representation at a young age. Proposed option may reduce RAEs those caused athletes are at risk to be eliminated during the development of their physical or fitness. Another favourable approach could be applied by increasing awareness of the RAE among those responsible in organization of youth sport. During particular growth, coaches need to aware about physical attributes such as height and weight. Since the component likely drop out during early stages of athlete development for example 13-16 years old which is leading to selection advantages to the relatively older at a time.

The RAE among athlete in Malaysia had been established that there are significant relative age effects in both genders for particular age groups [10]. Athletes born in the first three months were significantly over represented than those born in the last quarter of the year. The weaknesses of the current practices of athlete selection procedure in talent identification program were carried out by gathering all participants from various age groups. However, there has been proven that even small age differences such as month within an annual age-group can provide generous advantage in physical and cognitive maturity [4]. For this reason, there are high possibilities to drop out of talented athlete because the overrepresentation of children born early and underrepresentation of children born later of the selection year. In purpose to avoid subjectivity and bias selection, present study would propose to carry out the evaluation according to their particular quartile groups. This study
thus combined statistical analysis of CA and DA to determine the significant parameters for each quartile. Such parameters would be considered for the purpose of identifying their level of fitness performance appropriately.

2. MATERIALS AND METHODS

2.1. Participants

The data for this study were derived from Terengganu Talent Identification program 2011. A total of 178 students with aged 12-years-old were participated in this study. The birth dates of studied group were tabulated by birth-months. These categories were compiled into quarters, which start on January and ends on 31 December following a previous study by [10]. Thus, the first quarter (Q1) included months of January, February and March. The remaining quarters were Q2 (April, May and June), Q3 (July, August and September) and Q4 (October, November and December).

2.2. Testing Procedures

All tests were performed over two consecutive days at the beginning of the competitive season. The test sessions were carried out in indoor halls with synthetic floors. Participants were asked whether they had prior experience with the tests to be carried out. Therefore, testing protocol was separately explained to each group of participants who had not been previously tested on any occasions in previous seasons for training prescription purposes. In addition, all participants were requested to have their last meal three hours before the tests and not to participate in any prolonged exercise 24 hours before the tests.

2.3. Physical Measurement

The physical attributes data included four types of measurements. There is weight, height, sitting height and arm span. The body weight measurement leads the participants to take out their heavy outer attire and shoes. The participant stands in the center of the platform so that weight distributed evenly to both feet. The weights are moved until the arrows are aligned and weight is recorded. Standing height is the measurement of maximum distance from the floor to the highest point on the head. The position of the subjects is facing directly ahead with shoes should be off, feet together and arms by the sides. Heels, buttocks and upper back
should also be in contact with the wall when the measurement is made. Sitting height gives a measure of the length of the trunk. It is a measurement of the distance from the highest point on the head to the base sitting surface. The subject sits with both feet on the floor, the lower back and shoulders against the wall, looking straight ahead. Arm span was measured from one furthermost fingertip to the other. Participants facing away from the wall, with back and buttocks touching the arms are stretched out horizontally.

2.4. Motor Fitness Assessments

After the physical measurements had been taken, each participant performed a standardized 15-minute warm-up consisting of general movements and static and dynamic stretching as per their usual training warm-up routines, followed by nine tests of motor abilities (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). The participants were divided into groups and repeated each test three times (each subject repeated the test after everyone else in the group had completed the previous repetition of the test), with an adequate recovery period between tests (up to 3 minutes of rest).

Speed was determined by a timed 10-m, 20-m and 40-m sprint. The tests were performed on an indoor synthetic pitch, and electronic timing gates were used to record completion times. They were repeated three times with a 90 second rest period and the best timing was taken to represent the sprint performances.

Power component were tested by standing broad jump and vertical jump. In standing broad jump, the participant attempts to jump as far as possible, landing on both feet without falling backwards. Three attempts are allowed and the farthest was taken into account. In vertical jump procedure, the jump height is the difference between standing height and jumping height.

To test the flexibility components, sit and reach were measured. This test involves sitting on the floor with legs out straight ahead. Feet with shoes off are placed with the soles flat against the box, shoulder-width apart. Both knees are held flat against the floor by the investigator, if required. With hands on top of each other and palms facing down, the participant reaches forward along the measuring line as far as possible. After three practice reaches, the fourth
reach is held for at least two seconds while the distance is recorded.

In maximum push up procedure, the participant starts with position arms straight with the elbows locked, body in a straight line, hands placed slightly wider than shoulder-width apart with fingers pointing forward, and both feet on the floor. The participant lowers their body until they slightly compress the foam cube, and arms are at least parallel to the floor, then pushes up again. The back must be kept straight and in each extension up, the elbows should lock.

In one minute sit up procedure, the participants begin by sitting comfortably and start with his back straight on the mat. His knees should be bent and his feet on the mat. The participant should then place his hands on both sides of his head with his elbows pointing out to the sides. From the starting position the participant should raise his upper body of the mat by tensing his core and lifting himself up towards his knees. Hands should remain on the side of head and knees should be bent with feet placed firmly on the ground.

The maximal multistage 20-m shuttle run test was used to evaluate maximal aerobic capacity from maximal aerobic speed. This test was completed after all eight other physical fitness testing had concluded. Participants were required to run between two lines 20 m apart. The pace was dictated by a cassette tape emitting tones at prescribed intervals. When runners could no longer keep up the pace by reaching the line at the time of the tone, participation was terminated and the number of laps completed was recorded [33]. Scores of the last stage number were converted to predict maximal oxygen uptake (VO\textsubscript{2} max) [34]. VO\textsubscript{2} max was expressed in ml of oxygen consumed per kilogram of body weight and per minute (ml. kg\textsuperscript{-1}. min\textsuperscript{-1}).

2.3. Data Analysis

Pre-processing data

A matrix set of 178 boys group contain 2314 matrices data (13 variables × 178 participant) were computed in this study. The total missing data in the matrices were very small (~3 %) compared to the overall data recorded. For the data analysis screening, the nearest neighbor method was applied [25]. This method examines the distance between each point and the closest point to it. The nearest neighbor method is the simplest methods, in which the
endpoint value is used to estimate missing values [26-27].

2.4. Cluster Analysis (CA)
In this study, CA was employed to investigate the grouping of the participant’s profile in a sport development program. In this study, cluster analysis (CA) was employed to identify the grouping of the relative performance pattern. CA is an appropriate method to identify and classify components or subjects (observations/population) into clusters with greater homogeneity state within the class and greater heterogeneity state among classes with regard to a predetermined selection criterion [22]. Moreover, Ward’s technique utilizing Euclidean distances as a degree of similarity in CA has shown to be a very comprehensive technique. The finding is clarified by a dendrogram, giving the clusters and their proximity [23]. However, the Euclidean distance is reported as \( \frac{D_{\text{link}}}{D_{\text{max}}} \), which represents the quotient between the linkages distances alienated by the highest distance [24]. The quotient is commonly multiplied by 100 as a way to standardizing the linkage distance signified by the y-axis [28]. Finally, the clusters given by the CA will be considered as predefined classes for DA.

2.5. Discriminant Analysis (DA)
DA was used in this study to determine their respective groups based on the mean of variables and use variables to predict group membership. It builds a discriminant capacity (DF) for every group. DFs are figured utilizing Equation (1) where \( i \) is the amount of group (V), \( c_i \) is the constant inherent to each group, \( n \) is the quantity of parameters utilized to categorize a set of data into a certain group and \( w_j \) is the mass coefficient assigned by DF analysis (DFA) to a given parameter \( (p_j) \) [29]. In this study, the DA was put into the raw data using standard, forward stepwise, and backward stepwise modes to discriminate three groups of relative performance (high performance, medium performance and low performance) that were assigned by CA [30]. These were used to construct DFs to evaluate relative performance variations in the physical attributes and motor fitness variables. The relative performances of the subjects were treated as dependent variables whereas all the assessed components were treated as independent variables. In the forward stepwise mode, variables are counted step by step starting with the utmost significant variable until no significant changes were obtained. In
the backward stepwise mode, variables are removed step by step beginning with the less significant variable until no significant changes were obtained [31-32].

\[ f(V_i) = c_i + \sum_{j=1}^{p} w_{ij}P_{ij} \]  

(1)

3. RESULTS AND DISCUSSION

3.1. Classification of Performance Level Based on Physical Attributes and Motor Fitness Variables

Following procedure by previous study [11], CA were carried out among 12-years-old students involved in this study based on their physical attributes and motor fitness data. This analysis resulted grouping the students into three groups (Fig. 1a) which is high performance group (HPG), moderate performance group (MPG) and low performance group (LPG) with a total 59, 40 and 79 students respectively. The clustering procedure generated three groups in a very convincing way, as the children in these groups have similar characteristics and natural backgrounds.

Profile plot (Fig. 1b) graphically showed HPG enclosed at highest stage on motor fitness and middle stage on physical attribute. MPG enclosed at middle stage on motor fitness and lowest stage on physical attribute. Otherwise, LPG enclosed at lowest stage on motor fitness and highest stage on physical attribute. Among 54 children in Q1, 16 (29.6%) were grouping in HPG, 26 (48.1%) in MPG and 12 (22.2%) in LPG. Among 52 children in Q2, a total 17 (32.7%), 16 (30.8%) and 19 (36.5%) children were grouping in HPG, MPG and LPG respectively. Among 41 children of Q3, assigning group of HPG consists of 3 (7.3%) students, MPG consists of 22 (53.7%) students and 16 (39.0%) students belongs to LPG. Otherwise, 31 students in Q4 were assigned a total 4 (12.9%), 15 (48.4%) and 12 (38.7%) into HPG, MPG and LPG respectively. All the results are tabulated in Table 1.
Fig. 1. The number of clusters given by CA where (a) dendrogram shows three clusters assigning different performance level of children and (b) the profile plot of groups.

Table 1 showed significant associations emerged among frequencies of children in birth-month groups of 12-year-old students. In high performance group, there are higher frequencies of children were selected among Q1 group compared to Q4 group. These results indicated that significant differences were found in the number of children in birth-month groups selected as high and low performance. This finding was in line with the previous study that suggested older year groups are overrepresented in comparison to younger ones that indicating a constituent year effect [12]. As such, it will be misleading and problematic if the children are selected on these criteria because the children who are older performed higher [1]. This is how RAE may influence the selection process. Since previous researchers already claimed that growth spurts in children occur at different period, children who are less active at
early stage are later usually bare to be more active compare to their peers who were initially more active [13].

Table 1. Number of children assigned in HPG, MPG and LPG for each birth-month quartile

<table>
<thead>
<tr>
<th>Group Assigned by CA</th>
<th>Number of Children</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1</td>
<td>Q2</td>
</tr>
<tr>
<td>HPG</td>
<td>16 (29.6%)</td>
<td>17 (32.7%)</td>
</tr>
<tr>
<td>MPG</td>
<td>26 (48.1%)</td>
<td>16 (30.8%)</td>
</tr>
<tr>
<td>LPG</td>
<td>12 (22.2%)</td>
<td>19 (36.5%)</td>
</tr>
<tr>
<td>Total</td>
<td>54</td>
<td>52</td>
</tr>
</tbody>
</table>

Regarding to the problematic of talent attrition was higher among children born in birth-month quartile of Q3 and Q4 [14], present study has revealed the effect of RAE on physical attributes and motor fitness for each birth-month quartile. So that, the selection process by grouping players in birth-month quartile group lead the process became fair and may avoid talent wastage. Current practice of evaluating player in same age group already confirmed that players who were born in the first quarter of birth is more prominent because it has advantages in terms of height and weight compared to the players who were born in the last quarter [15]. Although these children demonstrated the low ability on their physical, they should have the same opportunities in the assessment by taking into consideration the factor of relative age effect. This is because they are actually has high potential comparable to children who are born early and having a better physique. Therefore, while grouping them based on their quartile, their chances to outperform will increase.

3.2. Identification of the Discriminating Factor for RAE

DA was broadly applied in many areas such as environmental [16] and sports [17]. In this study, DA was applied on the existing data post grouping of the students aged 12-year-old into three main groups defined by CA to study the physical attributes and motor fitness at different birth-month quartile. For each quartile, groups (HPG, MPG and LPG) were treated as dependent variables. Meanwhile physical attributes and motor fitness variables were treated as independent variables. DA was run via standard, forward stepwise and backward stepwise methods.
For the older quartile Q1, the accuracy of classification using standard, forward stepwise and backward stepwise mode DA were 83.33% (12 discriminant variables), 90.74% (three discriminant variables) and 90.74% (five discriminant variables) respectively (Table 2). Using forward stepwise, weight, 20 meter run and 1 minute sit up were found to be the significant variables. This indicates that these parameters have high variation in terms of their anthropometric and physiological attributes. Backward stepwise mode on the other hand included 40 meter run and VO$_2$ max as the fourth and fifth parameter to have a high fitness variation. Otherwise, second quartile Q2 also performed high accuracy of classification using the entire mode of DA. There are a percentage of 82.69% (12 discriminant variables), 84.62% (two discriminant variables) and 90.38% (four discriminant variables) of correctness assigned by standard, forward stepwise and backward stepwise mode DA respectively. Using forward stepwise, sitting height and 40 meter run were found significantly different for every level of performance. However, backward stepwise revealed sitting height, 20 meter run, 1 minute sit up and VO$_2$ max were the most significant variables.

Continuing the DA result for third quartile Q3, the accuracy of classification using standard, forward stepwise and backward stepwise mode DA were 78.05% (11 discriminant variables), 87.80% (two discriminant variables) and 87.80% (four discriminant variables) respectively. Using forward stepwise, weight and vertical jump were found to be significant variables. Meanwhile, backward stepwise included sitting height and 40 meter run as third and fourth parameters to be high variation among them. The youngest quartile Q4, revealed percentage of 80.65% (11 discriminant variables), 96.77% (3 discriminant variables) and 96.77% (3 discriminant variables) for standard, forward stepwise and backward stepwise mode respectively. Using forward stepwise, weight, sitting height and 40 meter run were found as significant variables. On the other hand, backward stepwise replaced sitting height with height parameters to became most significant variable. The main finding of DA in this study revealed there is five discriminated variables for Q1, four discriminated variables for Q2 and Q3, on the other hand Q4 were discriminated by only three variables. It is complicated to give a full justification for the lack of evidence of significant relationship between physical components and characteristics of motor fitness and birth-month quartile in this study. However, such
finding can suggest the significant parameters are most likely as general assessment criteria that can be considered when selecting players [18]. As physical attributes and motor fitness measurements are considered very important, selection policy should take into account the two components in the selection and recruitment process. The present center is based on fair decision making. This is to prevent dropout of potential players in the future despite their weakness in physical characteristics have been identified in advance. Contrary to the present invention, there are previous studies that prove there is no clear trend in a particular component in the youngest group to the oldest group of athlete [19]. The observation in this study showed that in the group under 14 year old, elite players born late in the year and have low ability in physical size does not necessarily disadvantaged. There is also evidence that the component anthropometric variables can be negative [20]. It means that improvements in the functioning of the components and parameters vary according to the specific sport at particular competitive ages. While the evaluation and selection of players according to anthropometric capabilities and motor fitness under the age of 14 cannot be the basis for future success, children who are born early and stand will be able to strengthen their performance. This revealed that the identification and selection of talent during his early teens appear to be influenced significantly by the physical properties of the child instead of skills [5]. Besides, there was no difference in any fitness features observed among players born in different periods and also between professional and non-professional players. However, the trends identified in which the players who born in the first quarter generally over-representation those born in last quarter of the year [21].

**Table 2.** DA Classification matrix for physical attributes and motor fitness variables
<table>
<thead>
<tr>
<th>Groups assigned by CA</th>
<th>Q1 (Jan-March)</th>
<th>Q2 (Apr-June)</th>
<th>Q3 (July-Sept)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HPG</td>
<td>MPG</td>
<td>LPG</td>
</tr>
<tr>
<td>Standard</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPG</td>
<td>13</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>MPG</td>
<td>2</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>LPG</td>
<td>1</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>26</td>
<td>12</td>
</tr>
<tr>
<td>Stepwise (Forward)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPG</td>
<td>15</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>MPG</td>
<td>2</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>LPG</td>
<td>0</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>27</td>
<td>10</td>
</tr>
<tr>
<td>Stepwise (Backward)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPG</td>
<td>15</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>MPG</td>
<td>2</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>LPG</td>
<td>0</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>26</td>
<td>11</td>
</tr>
</tbody>
</table>
Therefore, the effectiveness of identification and development of athletic talent at an early age have the opportunity to change due to the growth and maturity of the child at various stages during the month of birth. Further research on the selection of player using multidisciplinary approach should be related to the run in small groups of birth month quartile rather than in large group of same age. Physical attributes and the motor fitness are significant based on the date of birth of the child in the next 12 years. It would be worthwhile to investigate the secular trend in factors related to the performance.

4. CONCLUSION
The relative age effect on physical attributes and motor fitness for each birth-month quartile are confirmed. There are different total of significant parameters and different variables to be considered while conducting player selection in small groups of birth-month quartile. There is possibility of occurrence of high, moderate and low performance of players among children at each different quartile group. The significant variables can be as guideline for player selection in talent and development program. Such approach would be able to reduce RAE that may cause bias and subjective decision making.

5. ACKNOWLEDGEMENTS
The authors thank the Research and Development Management Unit, University of Sultan Zainal Abidin (UniSZA), Kuala Terengganu for providing the research grant for the study (FRGS/1/2016/SS05/UNISZA/02/1/RR207). The authors also express their gratitude to the Terengganu State Sports Council and National Sports Institute for their support in this research.

6. REFERENCES


Sport Sciences, 1984, 9(2):64-69

**How to cite this article:**  