Motor Control, **(Ahead of Print)**
https://doi.org/10.1123/mc.2021-0126

First Published Online: Apr. 28, 2022

Human Kinetics
ARTICLE

Young Swimmers' Classification Based on Performance and Biomechanical Determinants: Determining Similarities Through Cluster Analysis

Jorge E. Morais,[1,2] Tiago M. Barbosa,[1,2] Henrique P. Neiva,[2,3] Mario C. Marques,[2,3] and Daniel A. Marinho[2,3]

[1]Department of Sport Sciences, Instituto Politécnico de Bragança, Bragança, Portugal; [2]Research Centre in Sports, Health and Human Development (CIDESD), Covilhã, Portugal; [3]Department of Sport Sciences, University of Beira Interior, Covilhã, Portugal

The aim of this study was to classify and identify young swimmers' performance, and biomechanical determinant factors, and understand if both sexes can be clustered together. Thirty-eight swimmers of national level (22 boys: 15.92 ± 0.75 years and 16 girls: 14.99 ± 1.06 years) were assessed. Performance (swim speed at front crawl stroke) and a set of kinematic, efficiency, kinetic, and hydrodynamic variables were measured. Variables related to kinetics and efficiency ($p < .001$) were the ones that better discriminated the clusters. All three clusters included girls. Based on the interaction of these determinant factors, there are girls who can train together with boys. These findings indicate that not understanding the importance of the interplay between such determinants may lead to performance suppression in girls.

***Keywords**:* swimming, training, stroke mechanics, technique, swimmer similarities

Swimming performance is a multifactorial phenomenon characterized by the interactions of several determinants (Morais et al., 2013; Nikolaidis, 2012; Silva, Figueiredo, Morais, et al., 2019). In swimming, variables related to swimming technique, namely kinematics, efficiency (Silva, Figueiredo, Ribeiro, et al., 2019), and hydrodynamics (Barbosa et al., 2015), play a key role. For instance, the fastest swimmers racing freestyle sprinting events usually show longer stroke length (SL), faster stroke frequency (SF), and lower intracyclic variation of the swim speed (d*v*) compared with their worst performance counterparts (Morais et al., 2015).

Recently, it has been shown experimentally that thrust (i.e., amount of in-water force) also determines swim speed in young swimmers, in which a higher thrust is related to a higher swim speed (Dos Santos et al., 2021; Morais et al.,

Morais (morais.jorgestrela@gmail.com) is corresponding author, https://orcid.org/0000-0002-6885-0648

2021). In addition, in the front crawl, it was found that young swimmers may present a bilateral asymmetry in the thrust of the upper limbs and consequently in the swim speed achieved by each upper limb during the arm pull (Morais, Forte, et al., 2020). Therefore, assessing the thrust independently for each upper limb will reveal substantial information about contralateral limitations that young swimmers may present during the arm pull. Indeed, at least for adult swimmers, it was shown that minimizing such asymmetries will lead to improved performance (Dos Santos et al., 2013).

Genetic factors play a key role in gender differences (Ben-Zaken et al., 2022; Geladas et al., 2005). Boys are stronger and have longer limbs than girls, which will promote better performances (Nevill et al., 2020). In addition to such genetic factors, training also plays a determining role in boys' and girls' performances (Dos Santos et al., 2021). Studies found a significant sex effect on variables related to swimming performance (kinematics, efficiency, and hydrodynamics) in young swimmers from 12 to 13 years onward (Barbosa et al., 2013; Silva, Figueiredo, Ribeiro, et al., 2019). For instance, it was shown that boys tend to present larger SL, faster SF, and higher stroke index (SI) (Silva, Figueiredo, Ribeiro, et al., 2019). Conversely, girls presented a better hydrodynamical profile with lower coefficient of active drag (D_a [C_{Da}]; Barbosa et al., 2013). Mixed findings were shown for the dv. A study noted that boys present a lower dv (Silva, Figueiredo, Ribeiro, et al., 2019), and girls present a lower dv Barbosa et al. (2013). Therefore, one may argue that the performance determinants may differ between boys and girls.

However, data analysis in swimming performance, as well as in other sports, is usually carried out considering the sample or the group's main trend. Thus, each subject's profile is "masked" by the main trend of the overall sample. Cluster analysis is a feasible procedure to assess individual trends within an overall sample (Rein et al., 2010). As swimming performance is a multifactorial phenomenon, multivariate data analysis such as cluster analysis can be employed to detect patterns within data sets. Cluster analysis is a procedure that gathers in the same group subjects who share several common characteristics but are very different from others who do not belong to that cluster (Rein et al., 2010). Moreover, it also allows the identification and classification of young swimmers' determinant factors at a given moment (Figueiredo et al., 2016) or even longitudinally (Morais et al., 2015). This will allow an understanding of whether there are any determinant factors related to swim technique (kinematic, efficiency, kinetic, and hydrodynamic) that are common to boys and girls within an overall sample, allowing them to be gathered in the same training group. Well-designed training programs based on determinants related to swim technique can lead swimmers to excel themselves. Such training plans should be adequate to each swimmer based on their specific characteristics, or to a group of swimmers with a similar profile (Pyne, 2020). Thus, it can be argued that there are characteristics or determinant factors that can be common to both sexes, allowing boys and girls to train under the same training design.

Therefore, the aims of this study were to: (a) classify and identify young swimmers' performance (swim speed), and biomechanical determinant factors (kinematic, efficiency, kinetic, and hydrodynamic) gathered into subgroups and (b) understand if girls can be clustered with boys. It was hypothesized that different biomechanical variables (related to kinematics, efficiency, kinetics, and hydrodynamics) would be responsible for the cluster formation, and that girls would not be included in the cluster that gathered the fastest swimmers.

Methods

Participants

Thirty-eight young swimmers (22 boys: 15.92 ± 0.75 years, 68.93 ± 6.99 kg of body mass, 176.91 ± 5.66 cm of height, 182.82 ± 8.37 cm of arm span, Tanner Stages 4–5, and 566.77 ± 56.83 Fédération Internationale De Natation (FINA) points in the 100-m freestyle event in short course; 16 girls: 14.99 ± 1.06 years, 56.66 ± 5.94 kg of body mass, 162.63 ± 6.95 cm of height, 167.56 ± 7.11 cm of arm span, Tanner Stages 4–5, and 602.25 ± 77.36 FINA points in the 100-m freestyle event in short course) were recruited for analysis. The swimmers were recruited from a national team at the end of the second macrocycle, which corresponded to the second peak performance of the season. The inclusion criteria for the participants were: (a) national-level swimmers in their age group in the freestyle sprint events, (b) having participated in daily training sessions from the beginning of the season, and (c) without any injury since the beginning of the season. All participants had more than 5 years of competitive experience and trained six to seven swimming sessions per week, combined with at least one dryland strength, and conditioning session per week. Parents or guardians, and the swimmers themselves signed an informed consent form. All procedures were in accordance with the Declaration of Helsinki regarding human research, and the University of Beira Interior ethics board approved the research design.

Performance

The swim speed (in meters per second) was considered the performance indicator. Swimmers were invited to perform three all-out trials in front crawl with a push-off start. A mechanical apparatus (Swim speedometer; Swimsportec, Hildesheim, Germany) was attached to the swimmer's hip (Barbosa et al., 2015). The speed of the swimmer's hip was deemed appropriate as an indication of the swimmer's center of mass. Literature indicates that despite presenting a 7% error, the measurement of the swim speed based on this approach is a simple and less time-consuming process that allows evaluating several swim cycles and giving quick feedback to swimmers (Fernandes et al., 2012). An in-house built software (LabVIEW, version 2010, National Instruments) was used to acquire (f = 50 Hz) and display speed–time data over each trial on a laptop. Data were exported from the speedometer to the interface by a 12-bit resolution acquisition card (USB-6008, National Instruments), it was then imported into a signal processing software (AcqKnowledge, version 3.9.0, Biopac Systems). Signal was handled with Butterworth fourth order low-pass filter (cutoff: 5 Hz). The swim speed was retrieved from the software between the 11th and 24th meter. The best trial was used for analysis.

Kinematics and Efficiency

The kinematics and efficiency parameters were collected concurrently with the performance. The intracyclic dv was computed as the coefficient of variation:

coefficient of variation = 1 *SD*/mean (Barbosa et al., 2010). The SF was measured by calculating the number of cycles per unit of time from the time it takes to complete one full cycle ($f = 1/t$, the mean of three consecutive full stroke cycles was used for analysis), and afterward converted to Hz. The SL was computed as: SL = v/SF, in which SL (in meters), v is the swim speed (in meters per second), and SF (in hertz) (Craig & Pendergast, 1979).

The SI was computed as: SI = v·SL, in which SI (in square meters per second), v is the swim speed (in meters per second), and SL (in meters) (Costill et al., 1985). The Froude efficiency (η_F) was computed as:

$$\eta_F = \left(\frac{v \cdot 0.9}{2\pi \cdot \mathrm{SF} \cdot l}\right) \cdot \frac{2}{\pi}, \tag{1}$$

in which η_F (dimensionless), v is the swim speed (in meters per second), SF (in hertz), and l is the shoulder to hand average distance (in meters) (Zamparo et al., 2005). The latter variable was measured on land by digital photogrammetry (Morais, Marques, et al., 2020). There are more recent models to computed the η_F such as the one presented by Gonjo et al. (2018). This model (three-dimensional direct linear transformation) is based on the ratio between the body's center of mass speed, and hand speed, having the wrist as reference. However, this is a more complex and time-consuming method to be used in a translational and applied research with young swimmers.

Kinetics

The propulsive force (kinetics) was acquired concurrently to kinematics and performance testing (the same three maximum all-out trials of 25-m front crawl with a push-off). Differential pressure sensors and underwater video (Aquanex + Video, Swimming Technology Research, Tallahassee, FL) were used to measure propulsive force ($f = 100$ Hz) (Morais, Forte, et al., 2020). Such sensors were placed between the third and fourth metacarpals to measure the pressure differential between the palmar and dorsal surfaces. This location is considered a good proxy of the application point of the thrust vector in the hand (Gourgoulis et al., 2013). At the beginning of each trial, swimmers were asked to keep their hands underwater at the waistline for 10 s to calibrate the system with the hydrostatic pressure values. The video camera was placed at the side of the swimming pool recording the swimmers on the sagittal plane. The sensors were connected to an analog-to-digital converter connected to a laptop, on the pool deck with the Aquanex software (version 4.2 C1211, Aquanex; Morais, Forte, et al., 2020). Afterward, time–force series were imported into a signal processing software (AcqKnowledge, version 3.9.0, Biopac Systems). Signal was handled with Butterworth fourth order low-pass filter (cutoff: 5 Hz).

For each dominant and nondominant arm pull, the mean propulsive force ($F_{\mathrm{mean\_dominant}}$ and $F_{\mathrm{mean\_nondominant}}$ [in newton]), and the peak force ($F_{\mathrm{peak\_dominant}}$ and $F_{\mathrm{peak\_nondominant}}$ [in newton]) were analyzed. Afterward, the $F_{\mathrm{mean\_stroke\ cycle}}$ (the mean force produced in one full stroke cycle [in newton]) was calculated. The intracyclic force variation of the ($\mathrm{d}F_{\mathrm{dominant}}$, $\mathrm{d}F_{\mathrm{nondominant}}$, and $\mathrm{d}F_{\mathrm{mean\_stroke\ cycle}}$ [in percentage]) was computed as the coefficient of variation (as aforementioned).

Hydrodynamics

The C_{Da} (dimensionless) was computed based on the velocity perturbation method (Kolmogorov & Duplishcheva, 1992). Swimmers performed two all-out trials of 25-m front crawl with a push-off start. One trial was carried out towing a hydrodynamic body (i.e., a perturbation device) and the other without towing it (Kolmogorov & Duplishcheva, 1992). Between trials swimmers had 10 min for full recovery. A camera (Sony x3000, Sony Corporation) was used to record the swimmer's displacement time between the 11th and 24th meter. The swim speed was calculated as: $v = d/t$. The D_a (in newton) was computed as:

$$D_a = \frac{D_b v_b v^2}{v^3 - v_b^3}, \tag{2}$$

in which D_a is the swimmer's active drag at maximum speed (in newton), D_b is the resistance of the hydrodynamic body computed from the manufacturer's calibration of the buoy-drag characteristics and its speed (in newton), v_b, and v are the swimming velocities with and without the perturbation device (in meters per second). Afterward, the C_{Da} was computed as:

$$C_{Da} = \frac{2 \cdot D_a}{\rho \cdot S \cdot v^2}, \tag{3}$$

in which C_{Da} is the D_a coefficient (dimensionless), D_a (in newton), ρ is the density of the water (1,000 kg/m^3), the body's cross-sectional surface area (S [in square meters]), and v is the swim speed (in meters per second). For convenience of measurement, the S was computed on land by digital photogrammetry. The swimmers were photographed by a digital camera (Alpha 6000, Sony Corporation) in the transverse plane (downward view) on land simulating the streamlined position (Morais et al., 2012).

Anthropometrics

The body mass was measured with a digital scale (BC-730, Tanita), and the height with a digital stadiometer (Seca 242, Seca). The arm span was measured by digital photogrammetry (Morais, Marques, et al., 2020). For this, swimmers were photographed near a 2D calibration frame, in an orthostatic position, with both arms in lateral abduction at a 90° angle to the trunk. Both arms and fingers were fully extended. The distance between the tip of each third finger was measured. The swimmers' anthropometric characterization by cluster and sex is presented in Table 1.

Statistical Analysis

The normality and homoscedasticity assumptions were analyzed with the Kolmogorov–Smirnov and Levene tests, respectively. The mean + 1 *SD* were calculated as descriptive statistics. The modeling of clusters was performed based on the nonhierarchical *k*-means approach, which allows to previously define a number of clusters to be used. The *k*-means defines a centroid, which is the mean of a group of points/subjects, based on their similarities (Rein et al., 2010). To ensure

Table 1 Descriptive Statistics (Mean ± 1 *SD*) for the Anthropometric Variables Measured

	Mean ± 1*SD*								
	Cluster 1			Cluster 2			Cluster 3		
	Overall (*N* = 17)	Boys (*n* = 15)	Girls (*n* = 2)	Overall (*N* = 6)	Boys (*n* = 5)	Girls (*n* = 1)	Overall (*N* = 15)	Boys (*n* = 2)	Girls (*n* = 13)
Body mass (kg)	69.39 ± 7.40	69.98 ± 7.71	65.00 ± 0.42	66.50 ± 6.03	68.16 ± 4.98	58.20 ± 0.00	56.29 ± 5.80	63.00 ± 2.26	55.25 ± 5.49
Height (cm)	176.88 ± 6.17	177.40 ± 6.41	173.00 ± 0.00	174.33 ± 4.68	175.60 ± 3.91	168.00 ± 0.00	162.73 ± 7.97	176.50 ± 4.95	160.62 ± 5.97
Arm span (cm)	183.06 ± 8.55	183.93 ± 8.70	176.50 ± 3.53	178.83 ± 7.03	179.80 ± 7.40	174.00 ± 0.00	167.87 ± 8.81	182.00 ± 11.31	165.69 ± 6.45
Body's cross-sectional area (cm^2)	1,004.27 ± 125.83	1,014.76 ± 130.14	925.58 ± 46.64	991.55 ± 168.65	1,048.52 ± 105.88	706.69 ± 0.00	796.40 ± 120.86	901.66 ± 62.02	780.21 ± 120.79

a coherent comparison of data sets with different magnitudes and/or units, standardized z scores of all variables were computed.

The one-way analysis of variance was used to identify the main determinants responsible for the cluster formation in each evaluation moment ($p \leq .05$). A discriminant analysis was performed to validate the cluster formation. The total eta square (η^2) was selected as effect size index, and deemed as: (a) without effect if $0 < \eta^2 \leq .04$, (b) minimum if $.04 < \eta^2 \leq .25$, (c) moderate if $.25 < \eta^2 \leq .64$, and (d) strong if $\eta^2 > .64$ (Ferguson, 2009). Afterward, the Bonferroni post hoc test ($p \leq .05$) was used to verify the differences between clusters for each variable. Cohen's d was selected as standardized effect size, and interpreted as: (a) trivial if $d < 0.20$, (b) small if $0.20 \leq d < 0.60$, (c) moderate if $0.60 \leq d < 1.20$, (d) large if $1.20 \leq d < 2.00$, (e) very large if $2.00 \leq d < 4.00$, and (f) extremely large if $d \geq 4.00$ (Hopkins et al., 2009).

Results

The model set three different clusters. Cluster 1 gathered the fastest swimmers according to their swim speed, Cluster 2 the intermediate swimmers, and Cluster 3 the slowest swimmers. Moreover, Cluster 1 included two girls (out of 17—11.8%), Cluster 2 included one girl (out of six—16.7%), and Cluster 3 included 13 girls (out of 15—86.7%).

The determinant factors that better discriminated the cluster solutions (based on higher F ratio and standardized η^2 effect), besides the performance, were the $F_{\text{mean\_stroke cycle}}$ ($F = 37.70$, $p < .001$, $\eta^2 = .68$), the $F_{\text{mean-dominant}}$ ($F = 24.77$, $p < .001$, $\eta^2 = .57$), and the SI ($F = 23.05$, $p < .001$, $\eta^2 = .57$) (Table 2).

The determinant factors that characterized each cluster are indicated in Table 2 and summarized in Table 3. Besides a high swim speed, Cluster 1 was characterized by high SI (efficiency), $dF_{\text{nondominant}}$ (kinetics), and SF (kinematics); Cluster 2 by high $F_{\text{mean\_stroke cycle}}$ (kinetics), high $F_{\text{mean\_dominant}}$ (kinetics), and high dv (efficiency; higher values of dv indicate worst performance); and Cluster 3 by low SI (efficiency), $F_{\text{peak\_nondominant}}$ (kinetics), and $F_{\text{mean\_stroke cycle}}$ (kinetics) (Table 3).

Table 4 presents the comparison between the clusters. The $F_{\text{mean\_stroke cycle}}$ was the variable that presented the highest and most significant mean difference between Clusters 1 and 2 (mean difference = −7.032, $p < .001$, $d = 1.81$). Between Clusters 2 and 3, there was also the $F_{\text{mean\_stroke cycle}}$ (mean difference = 12.360, $p < .001$, $d = 3.12$). In addition to performance, the SI (mean difference = 0.653, $p < .001$, $d = 2.24$) was the variable that presented the highest and most significant difference between Clusters 1 and 3.

For a qualitative analysis, the discriminant analysis showed a very good compactness/separation with a correct classification of the original groups (97.4%) and a correct classification of cross-validated groups (81.6%; Figure 1).

Discussion

The aims of this study were to: (a) classify and identify young swimmers' performance and biomechanical determinant factors (kinematics, efficiency, kinetics, and hydrodynamics) gathered into subgroups and (b) understand if girls can be

Table 2 Descriptive Statistics (Mean ± 1 *SD*) for All Variables Analyzed in Each Cluster

	Cluster 1 ($N = 17$)		Cluster 2 ($N = 6$)		Cluster 3 ($N = 15$)				
	Mean ± 1 *SD*	*z*	Mean ± 1*SD*	*z*	Mean ± 1 *SD*	*z*	*F*	*p*	η^2
Performance (m/s)	1.65 ± 0.07	0.8018	1.57 ± 0.08	0.2209	1.41 ± 0.08	–0.9971	42.09	<.001	.71
SF (Hz)	0.87 ± 0.07	0.5150	0.82 ± 0.06	–0.2114	0.80 ± 0.05	–0.4991	5.23	.010	.23
SL (m)	1.91 ± 0.15	0.3552	1.93 ± 0.10	0.4630	1.77 ± 0.13	–0.5877	5.31	.010	.23
dv (%)	8.21 ± 1.88	–0.2669	13.94 ± 6.21	1.3686	8.29 ± 1.72	–0.2450	9.88	<.001	.36
η_F (%)	31.12 ± 2.60	–0.3440	33.12 ± 2.85	0.4056	32.65 ± 2.50	0.2276	1.99	.152	.10
SI (m^2/s)	3.15 ± 0.29	0.6745	3.03 ± 0.21	0.3691	2.50 ± 0.29	–0.9121	23.05	<.001	.57
C_{Da} (dimensionless)	0.51 ± 0.15	–0.4444	0.54 ± 0.17	–0.3068	0.75 ± 0.25	0.6263	6.31	.005	.27
$F_{\text{mean\_dominant}}$ (N)	38.07 ± 3.56	0.2184	45.44 ± 6.38	1.3745	31.60 ± 3.92	–0.7973	24.77	<.001	.57
$F_{\text{peak\_dominant}}$ (N)	62.30 ± 6.36	0.1978	72.58 ± 7.30	1.2364	53.22 ± 8.41	–0.7187	15.85	<.001	.48
dF_{dominant} (%)	46.43 ± 9.35	0.2719	41.56 ± 4.84	–0.3078	42.59 ± 8.14	–0.1850	1.18	.319	.06
$F_{\text{mean\_nondominant}}$ (N)	35.97 ± 4.37	0.1050	42.67 ± 6.02	1.2865	31.79 ± 3.62	–0.6336	13.54	<.001	.44
$F_{\text{peak\_nondominant}}$ (N)	63.86 ± 6.12	0.3570	71.68 ± 10.97	1.1260	51.53 ± 5.87	–0.8550	22.32	<.001	.56
d$F_{\text{nondominant}}$ (%)	53.82 ± 7.52	0.6036	47.74 ± 2.91	–0.0910	42.86 ± 8.13	–0.6477	8.97	.001	.34
$F_{\text{mean\_stroke cycle}}$ (N)	37.02 ± 2.45	0.1907	44.05 ± 4.92	1.5412	31.69 ± 2.68	–0.8326	37.70	<.001	.68
d$F_{\text{mean_stroke cycle}}$ (%)	50.12 ± 6.36	0.5255	44.65 ± 3.27	–0.2346	42.73 ± 7.37	–0.5017	5.46	.009	.24

Note. The F ratios are also presented, showing the determinant factors responsible for the cluster formation (and correspondent effect size: eta square—η^2). Performance = swimming velocity; SF = stroke frequency; SL = stroke length; dv = intracyclic swimming velocity; η_F = Froude efficiency; SI = stroke index; C_{Da} = drag coefficient; F_{mean} = mean propulsive force; F_{peak} = peak propulsive force; dF = intracyclic variation of the propulsive force; z = standardized coefficient of the final cluster centers; F = F ratios; η^2 = standardized eta square.

Table 3 Summary of Key Features Characterizing Each Cluster

Cluster 1	Cluster 2	Cluster 3
Efficiency (+)	Kinetics (+)	Efficiency (–)
Kinetics (+)	Efficiency (–)	Kinetics (–)
Kinematics (+)		

Note. (+) is the high values; (–) is the low values.

clustered with boys. The model set three different clusters. All three clusters included girls. The determinant factors that better discriminated the cluster solutions (besides the performance) were the $F_{\text{mean\_stroke cycle}}$, the $F_{\text{mean-dominant}}$, and the SI. The discriminant analysis showed that clusters presented a very good compactness/separation.

Studies in swimming showed that cluster analysis is a feasible way to identify swimming determinants by gathering swimmers with similar characteristics (Figueiredo et al., 2016; Morais et al., 2015; Silva, Figueiredo, Morais, et al., 2019). Overall, it was shown that swimming performance in youth swimming is determined by the interactions of several variables (Figueiredo et al., 2016; Morais, Forte, et al., 2020). However, at early ages (11–13 years), anthropometric features seem to be the variables that better discriminate the clusters gathering swimmers with similarities (Figueiredo et al., 2016; Morais et al., 2015). That is, swimmers with larger body dimensions are likely to perform better. Nonetheless, it was argued that the anthropometric "disadvantage" can be overcome by individual adaptations in stroke mechanics to achieve greater swim speed (Figueiredo et al., 2016).

For this reason, the study researchers chose not to include anthropometric features to test the cluster formation and identify the swimming determinants. The inclusion of such variables may have minimized the effect that other variables may have on swimming performance, especially in young swimmers and in groups of boys and girls gathered together. It should be pointed out that girls included in Cluster 1 (better performances) presented the largest anthropometric features among all female swimmers. Nevertheless, they presented less 7.66% of body mass (4.98 kg of mean difference), less 2.31% of height (4.40 cm of mean difference), and less 3.95% of arm span (7.43 cm of mean difference) in comparison to their boys counterparts included in the same cluster. Conversely, the two boys included in Cluster 3 were not the ones with shortest anthropometrics among all boys. This shows that anthropometrics "alone" are not responsible for better performance. Otherwise, the boys included in Cluster 3, which presented anthropometric features within the boys' average sample, would be pooled in a better performance cluster. These findings highlight the swimming performance as a holistic phenomenon which is highly dependent on the interaction between several determinant factors.

Our data showed that thrust (specifically $F_{\text{mean\_stroke cycle}}$, the $F_{\text{mean-dominant}}$) and SI, which is an efficiency proxy, were the variables that better determined the formation of the cluster. The SI is based on the swim speed and the SL, and it measures the swimmer's ability to complete a given distance with a particular swim

Table 4 Comparison Between Swimmers in Each Cluster

	Cluster 1 vs. Cluster 2			Cluster 2 vs. Cluster 3			Cluster 1 vs. Cluster 3		
	Mean difference [95% CI]	*p*	*d*	Mean difference [95% CI]	*p*	*d*	Mean difference [95% CI]	*p*	*d*
Performance (m/s)	0.077 [–0.011, 0.166]	.105	1.06	0.162 [0.072, 0.252]	<.001	2.00	0.239 [0.173, 0.305]	<.001	3.19
SF (Hz)	0.049 [–0.0234, 0.121]	.296	0.77	0.019 [–0.054, 0.930]	1.000	0.36	0.068 [0.014, 0.122]	.009	1.15
SL (m)	–0.016 [–0.178, 0.145]	1.000	0.16	0.158 [–0.007, 0.322]	.063	1.38	0.142 [0.021, 0.262]	.017	1.00
dv (%)	–5.730 [–9.169, –2.291]	.001	1.25	5.653 [2.155, 9.151]	.001	1.24	–0.077 [–2.642, 2.489]	1.000	0.04
η_F (%)	–1.999 [–5.102, 1.103]	.342	0.73	0.475 [–2.68, 3.63]	1.000	0.18	–1.52 [–3.84, 0.790]	.320	0.60
SI (m^2/s)	0.126 [–0.206, 0.458]	1.000	0.47	0.527 [0.190, 0.865]	.001	2.09	0.653 [0.405, 0.901]	<.001	2.24
C_{Da} (dimensionless)	–0.031 [–0.271, 0.208]	1.000	0.19	–0.213 [–0.456, 0.031]	.106	0.98	–0.244 [–0.423, –0.065]	.005	1.16
$F_{\mathrm{mean\_dominant}}$ (N)	–7.37 [–12.401, –2.333]	.002	1.43	13.841 [8.720, 18.962]	<.001	2.61	6.473 [2.718, 10.229]	<.001	1.73
$F_{\mathrm{peak\_dominant}}$ (N)	–10.283 [–19.089, –1.478]	.017	1.50	19.357 [10.400, 28.314]	<.001	2.46	9.074 [2.505, 15.642]	.004	1.22
dF_{dominant} (%)	4.868 [–5.110, 14.847]	.684	0.65	–1.032 [–11.182, 9.119]	1.000	0.15	3.836 [–3.608, 11.281]	.610	0.44
$F_{\mathrm{mean\_nodominant}}$ (N)	–6.696 [–11.920, –1.472]	.008	1.27	10.882 [5.568, 16.196]	<.001	2.19	4.186 [0.289, 8.083]	.032	1.04
$F_{\mathrm{peak\_nondominant}}$ (N)	–7.823 [–16.104, 0.457]	.069	0.88	20.155 [11.732, 28.579]	<.001	2.29	12.332 [6.155, 18.509]	<.001	2.06
d$F_{\mathrm{nondominant}}$ (%)	6.080 [–2.657, 14.817]	.267	1.07	4.873 [–4.015, 13.760]	.530	0.80	10.952 [4.435, 17.470]	<.001	1.40
$F_{\mathrm{mean\_stroke\ cycle}}$ (N)	–7.032 [–10.632, –3.433]	<.001	1.81	12.360 [8.699, 16.022]	<.001	3.12	5.328 [2.643, 8.013]	<.001	2.08
d$F_{\mathrm{mean_stroke\ cycle}}$ (%)	5.471 [–2.245, 13.187]	.250	1.08	1.923 [–5.926, 9.772]	1.000	0.34	7.394 [1.638, 13.150]	.008	1.07

Note. Performance = swimming velocity; SF = stroke frequency; SL = stroke length; dv = intracyclic swimming velocity; η_F = Froude efficiency; SI = stroke index; C_{Da} = drag coefficient; F_{mean} = mean propulsive force; F_{peak} = peak propulsive force; dF = intracyclic variation of the propulsive force; d = Cohen's d (effect size); CI = confidence interval.

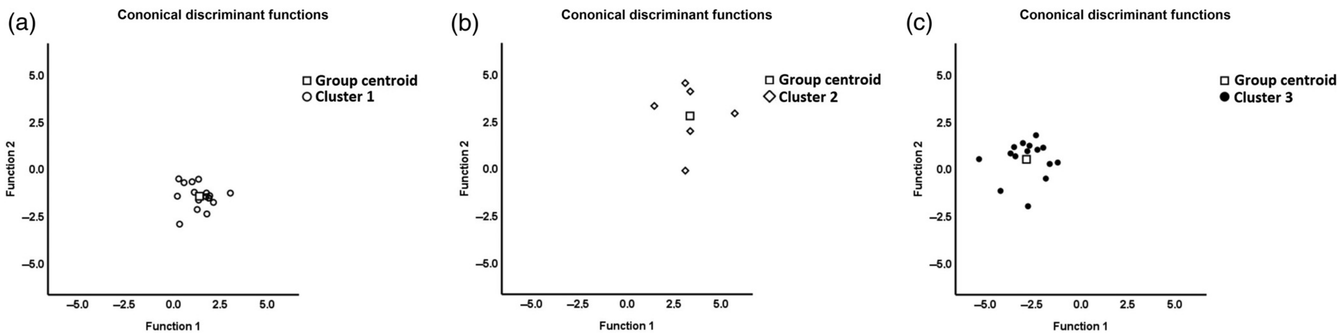

Figure 1 — Territorial map for each cluster. (a) ○—Cluster 1 membership, (b) ◊—Cluster 2 membership, and (c) •—Cluster 3 membership; □—group centroid.

speed in the fewest possible strokes (Costill et al., 1985). Thus, one can confirm that for a given speed, swimmers with longer SL are more efficient. On the other hand, there are no studies that implemented this kind of cluster analysis including thrust variables. It was shown through direct methods (with sensors placed in the swimmer's hand, which plays a key role in thrust production; Cohen et al., 2015) that thrust determines young swimmers' speed, and there is a strong relationship between these two variables, that is, whenever a decrease in thrust is observed, the swim speed also decreases (Morais, Forte, et al., 2020). Indeed, our data indicate that thrust is a key factor to differentiate swimmers with better and worst performances.

This clustering analysis revealed that girls were included in all three clusters. Sex differences in body features (namely size and composition) start developing at the onset of puberty (Sandbakk et al., 2018). Thus, it is normal that due to this "nature" phenomenon, boys present better performances than their girls' counterparts. Moreover, it was argued that the upper body revealed larger gender differences than for leg or whole-body exercise sports (Sandbakk et al., 2018). Our data showed that the $F_{\text{mean\_stroke cycle}}$ (upper limbs thrust) was the variable that better discriminated the cluster solutions. At least for the butterfly stroke in young swimmers (no data were found comparing boys and girls thrust based on pressure sensors in both upper limbs), boys presented higher $F_{\text{mean\_stroke cycle}}$ values than girls (Morais et al., 2021). The same trend was verified in front crawl but based on tethered swimming (Dos Santos et al., 2021; Oliveira et al., 2021). Indeed, in sprint events or all-out trials, the upper body has a substantial and significant effect on swimmers' performance (Bartolomeu et al., 2018; Morais, Forte, et al., 2020). It was shown that the motion of the upper limbs is responsible for nearly 90% of the swim speed in front crawl (Bartolomeu et al., 2018; Toussaint et al., 1988). Thus, and despite presenting smaller anthropometric features, girls can diminish the performance gender gap through individual adaptations namely by technical features such as thrust. Indeed, girls were included in the fastest clusters (1 and 2).

Literature reports mixed findings of differences between boys and girls in young swimmers' technical parameters. That is: (a) there are variables in which boys present higher values (Silva, Figueiredo, Ribeiro, et al., 2019); (b) other variables in which girls perform better or without significant differences in relation to boys (Wądrzyk et al., 2019); and (c) the performance determinants are different between sexes or with dissimilar effects (Morais et al., 2021). Cluster analysis by considering several variables simultaneously is a good solution to identify performance determinants considering each athlete's characteristics individually (Barbosa et al., 2016). In sports, small changes in one variable may not reflect an immediate change in another due to an existent interaction-dominant dynamic (Barbosa et al., 2016). Thus, practitioners consider that it is the sum of very small changes in several variables (through interactions) that helps the elite athlete to reach excellence. Moreover, perhaps considering longitudinal studies be employed to understand if the performance determinants change, and if girls are able to shift to the clusters that include the fastest swimmers.

Overall, beside performance, swimmers were clustered based on $F_{\text{mean\_stroke cycle}}$, the $F_{\text{mean-dominant}}$, and SI. That is, the fastest swimmers with both sexes grouped together were discriminated by higher thrust and efficiency. Despite the cluster with "worst" performances (Cluster 3) including the most girls, there were also girls included in the fastest clusters (1 and 2). Not understanding that some girls (i.e., with specific characteristics) can train as hard and strong as their boys' counterparts, can lead to retention in their performance enhancement. Indeed, it was recently argued that coaches can prescribe similar training for both girls and boys of identical training backgrounds (Reis et al., 2017). The authors observed that the most important oxygen consumption parameters were not different between sexes for the same relative intensity (Reis et al., 2017).

Moreover, swimming is characterized by interindividual variations (Bideault et al., 2013). That is, the success of a higher skilled behavior it is not dependent on a specific determinant, or from an unique path. It is rather achieved by an interaction or interplay of several determinants (Figueiredo et al., 2016; Morais et al., 2015). Indeed, literature does report that sports performance should adopt a multidisciplinary approach to better understand the athlete–environment relationship, which is based on a complex and dynamic system (Philips et al., 2010). Coaches should carefully verify each swimmer's characteristics to better understand how to master them. In the specific case of youth swimming, such characteristics or determinants are related to swim technique factors (kinematic, efficiency, kinetic, and hydrodynamic; Barbosa et al., 2015; Morais et al., 2012; Silva, Figueiredo, Ribeiro, et al., 2019). Thus, coaches must be aware that based on this dynamic and complex process, some girls can receive technical training alongside boys, whenever such characteristics, or determinants are identified. A main limitation to be considered is the fact that the rational of this study is only for postpubertal and sprinting swimmers. Therefore, it can be suggested to: (a) employ this modeling in swimmers specialized in middle and long distances; (b) add physiological variables to understand its interaction with biomechanics; and (c) employ longitudinal designs to understand if the determinant factors change over time, and specially according to the training design.

Conclusion

Cluster analysis revealed three clusters gathering girls and boys of the same age group and competitive level. Thrust ($F_{\text{mean\_stroke cycle}}$ and the $F_{\text{mean-dominant}}$) and efficiency (SI) variables were responsible for the cluster discrimination. All clusters were characterized by the interaction of several variables. Despite the most girls were included in Cluster 3 ("worst" performance), all clusters included girls, indicating that coaches must pay attention to the individual characteristic of each swimmer in order to avoid performance retention.

Acknowledgment

This work is supported by national funds (FCT—Portuguese Foundation for Science and Technology) under the project UIDB/DTP/04045/2020.

References

Barbosa, T.M., Bragada, J.A., Reis, V.M., Marinho, D.A., Carvalho, C., & Silva, A.J. (2010). Energetics and biomechanics as determining factors of swimming performance: Updating the state of the art. *Journal of Science and Medicine in Sport, 13*(2), 262–269. https://doi.org/10.1016/j.jsams.2009.01.003

Barbosa, T.M., Costa, M.J., Morais, J.E., Morouço, P., Moreira, M., Garrido, N.D., & Silva, A.J. (2013). Characterization of speed fluctuation and drag force in young swimmers: A gender comparison. *Human Movement Science, 32*(6), 1214–1225. https://doi.org/10.1016/j.humov.2012.07.009

Barbosa, T.M., Goh, W.X., Morais, J.E., Costa, M.J., & Pendergast, D. (2016). Comparison of classical kinematics, entropy, and fractal properties as measures of complexity of the motor system in swimming. *Frontiers in Psychology, 7,* 1566. https://doi.org/10.3389/fpsyg.2016.01566

Barbosa, T.M., Morais, J.E., Marques, M.C., Silva, A.J., Marinho, D.A., & Kee, Y.H. (2015). Hydrodynamic profile of young swimmers: Changes over a competitive season. *Scandinavian Journal of Medicine and Science in Sports, 25*(2), e184–e196. https://doi.org/10.1111/sms.12281

Bartolomeu, R.F., Costa, M.J., & Barbosa, T.M. (2018). Contribution of limbs' actions to the four competitive swimming strokes: A nonlinear approach. *Journal of Sports Sciences, 36*(16), 1836–1845. https://doi.org/10.1080/02640414.2018.1423608

Ben-Zaken, S., Eliakim, A., Nemet, D., Kaufman, L., & Meckel, Y. Genetic characteristics of competitive swimmers: A review. *Biology of Sport, 39*(1), 157–170. https://doi.org/10.5114/biolsport.2022.102868

Bideault, G., Herault, R., & Seifert, L. (2013). Data modelling reveals inter-individual variability of front crawl swimming. *Journal of Science and Medicine in Sport, 16*(3), 281–285. https://doi.org/10.1016/j.jsams.2012.08.001

Cohen, R.C., Cleary, P.W., Mason, B.R., & Pease, D.L. (2015). The role of the hand during freestyle swimming. *Journal of Biomechanical Engineering, 137*(11), 111007. https://doi.org/10.1115/1.4031586

Costill, D.L., Kovaleski, J., Porter, D., Kirwan, J., Fielding, R., & King, D. (1985). Energy expenditure during front crawl swimming: Predicting success in middle-distance events. *International Journal of Sports Medicine, 6*(5), 266–270. https://doi.org/10.1055/s-2008-1025849

Craig, A.B., & Pendergast, D.R. (1979). Relationships of stroke rate, distance per stroke, and velocity in competitive swimming. *Medicine and Science in Sports, 11*(3), 278–283.

Dos Santos, K.B., Pereira, G., Papoti, M., Bento, P.C.B., & Rodacki, A. (2013). Propulsive force asymmetry during tethered-swimming. *International Journal of Sports Medicine, 34*(7), 606–611. https://doi.org/10.1055/s-0032-1327575

Dos Santos, M.A., Henrique, R.S., Salvina, M., Silva, A.H.O., Junior, M.A.D.V., Queiroz, D.R., & Nevill, A.M. (2021). The influence of anthropometric variables, body composition, propulsive force and maturation on 50m freestyle swimming performance in junior swimmers: An allometric approach. *Journal of Sports Sciences, 39*(14), 1615–1620. https://doi.org/10.1080/02640414.2021.1891685

Ferguson, C.J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532–538. https://doi.org/10.1037/a0015808

Fernandes, R.J., Ribeiro, J., Figueiredo, P., Seifert, L., & Vilas-Boas, J.P. (2012). Kinematics of the hip and body center of mass in front crawl. *Journal of Human Kinetics, 33,* 15–23. https://doi.org/10.2478/v10078-012-0040-6

Figueiredo, P., Silva, A., Sampaio, A., Vilas-Boas, J.P., & Fernandes, R.J. (2016). Front crawl sprint performance: A cluster analysis of biomechanics, energetics, coordinative, and anthropometric determinants in young swimmers. *Motor Control, 20*(3), 209–221. https://doi.org/10.1123/mc.2014-0050

Geladas, N.D., Nassis, G.P., & Pavlicevic, S. (2005). Somatic and physical traits affecting sprint swimming performance in young swimmers. *International Journal of Sports Medicine, 26*(2), 139–144. https://doi.org/10.1055/s-2004-817862

Gonjo, T., McCabe, C., Sousa, A., Ribeiro, J., Fernandes, R.J., Vilas-Boas, J.P., & Sanders, R. (2018). Differences in kinematics and energy cost between front crawl and backstroke below the anaerobic threshold. *European Journal of Applied Physiology, 118*(6), 1107–1118. https://doi.org/10.1007/s00421-018-3841-z

Gourgoulis, V., Aggeloussis, N., Mavridis, G., Boli, A., Kasimatis, P., Vezos, N., & Mavrommatis, G. (2013). Acute effect of front crawl sprint resisted swimming on the propulsive forces of the hand. *Journal of Applied Biomechanics, 29*(1), 98–104. https://doi.org/10.1123/jab.29.1.98

Hopkins, W., Marshall, S., Batterham, A., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine & Science in Sports & Exercise, 41*(1), 3. https://doi.org/10.1249/MSS.0b013e31818cb278

Kolmogorov, S.V., & Duplishcheva, O.A. (1992). Active drag, useful mechanical power output and hydrodynamic force coefficient in different swimming strokes at maximal velocity. *Journal of Biomechanics, 25*(3), 311–318. https://doi.org/10.1016/0021-9290(92)90028-Y

Morais, J., Barbosa, T.M., Lopes, V.P., Marques, M.C., & Marinho, D.A. (2021). Propulsive force of upper limbs and its relationship to swim velocity in the butterfly stroke. *International Journal of Sports Medicine, 42*(12), 1105–1112. https://doi.org/10.1055/a-1386-4985

Morais, J.E., Forte, P., Nevill, A.M., Barbosa, T.M., & Marinho, D.A. (2020). Upper-limb kinematics and kinetics imbalances in the determinants of front-crawl swimming at maximal speed in young international level swimmers. *Scientific Reports, 10*(1), 1–8. https://doi.org/ 10.1038/s41598-020-68581-3

Morais, J.E., Garrido, N.D., Marques, M.C., Silva, A.J., Marinho, D.A., & Barbosa, T.M. (2013). The influence of anthropometric, kinematic and energetic variables and gender on swimming performance in youth athletes. *Journal of Human Kinetics, 39*(1), 203. https://doi.org/10.2478/hukin-2013-0083

Morais, J.E., Jesus, S., Lopes, V., Garrido, N., Silva, A., Marinho, D., & Barbosa, T.M. (2012). Linking selected kinematic, anthropometric and hydrodynamic variables to young swimmer performance. *Pediatric Exercise Science, 24*(4), 649–664. https://doi.org/10.1123/pes.24.4.649

Morais, J.E., Marques, M.C., Rodríguez-Rosell, D., Barbosa, T.M., & Marinho, D.A. (2020). Relationship between thrust, anthropometrics, and dry-land strength in a national junior swimming team. *The Physician and Sportsmedicine, 48*(3), 304–311. https://doi.org/10.1080/00913847.2019.1693240

Morais, J.E., Silva, A.J., Marinho, D.A., Seifert, L., & Barbosa, T.M. (2015). Cluster stability as a new method to assess changes in performance and its determinant factors over a season in young swimmers. *International Journal of Sports Physiology and Performance, 10*(2), 261–268. https://doi.org/10.1123/ijspp.2013-0533

Nevill, A.M., Negra, Y., Myers, T.D., Sammoud, S., & Chaabene, H. (2020). Key somatic variables associated with, and differences between the 4 swimming strokes. *Journal of Sports Sciences, 38*(7), 787–794. https://doi.org/10.1080/02640414.2020.1734311

Nikolaidis, P.T. (2012). Age- and sex-related differences in force-velocity characteristics of upper and lower limbs of competitive adolescent swimmers. *Journal of Human Kinetics, 32,* 87. https://doi.org/10.2478/v10078-012-0026-4

Oliveira, M., Henrique, R.S., Queiroz, D.R., Salvina, M., Melo, W.V., & Moura dos Santos, M.A. (2021). Anthropometric variables, propulsive force and biological maturation: A mediation analysis in young swimmers. *European Journal of Sport Science, 21*(4), 507–514. https://doi.org/10.1080/17461391.2020.1754468

Phillips, E., Davids, K., Renshaw, I., & Portus, M. (2010). Expert performance in sport and the dynamics of talent development. *Sports Medicine, 40*(4), 271–283. https://doi.org/10.2165/11319430-000000000-00000

Pyne, D. (2020). Monitoring training load in and out of the pool, optimal load and periodisation in young swimmers. In J. Dekerle (Ed.), *High performance youth swimming* (pp. 137–148). Routledge.

Rein, R., Button, C., Davids, K., & Summers, J. (2010). Cluster analysis of movement patterns in multiarticular actions: A tutorial. *Motor Control, 14*(2), 211–239. https://doi.org/10.1123/mcj.14.2.211

Reis, J.F., Millet, G.P., Bruno, P.M., Vleck, V., & Alves, F.B. (2017). Sex and exercise intensity do not influence oxygen uptake kinetics in submaximal swimming. *Frontiers in Physiology, 8,* 72. https://doi.org/10.3389/fphys.2017.00072

Sandbakk, Ø., Solli, G.S., & Holmberg, H.C. (2018). Sex differences in world-record performance: The influence of sport discipline and competition duration. *International Journal of Sports Physiology and Performance, 13*(1), 2–8. https://doi.org/10.1123/ijspp.2017-0196

Silva, A.F., Figueiredo, P., Morais, S., Vilas-Boas, J.P., Fernandes, R.J., & Seifert, L. (2019). Task constraints and coordination flexibility in young swimmers. *Motor Control, 23*(4), 535–552. https://doi.org/10.1123/mc.2018-0070

Silva, A.F., Figueiredo, P., Ribeiro, J., Alves, F., Vilas-Boas, J.P., Seifert, L., & Fernandes, R.J. (2019). Integrated analysis of young swimmers' sprint performance. *Motor Control, 23*(3), 354–364. https://doi.org/10.1123/mc.2018-0014

Toussaint, H.M., Beelen, A., Rodenburg, A., Sargeant, A.J., Groot, de G., Hollander, A.P., & van Ingen Schenau, G.J. (1988). Propelling efficiency of front-crawl swimming. *Journal of Applied Physiology, 65*(6), 2506–2512. https://doi.org/10.1152/jappl.1988.65.6.2506

Wądrzyk, Ł., Staszkiewicz, R., Kryst, Ł., & Żegleń, M. (2019). Gender effect on underwater undulatory swimming technique of young competitive swimmers. *Acta of Bioengineering and Biomechanics, 21*(4), 3–11. https://doi.org/10.37190/ABB-01422-2019-02

Zamparo, P., Pendergast, D.R., Mollendorf, J., Termin, A., & Minetti, A.E. (2005). An energy balance of front crawl. *European Journal of Applied Physiology, 94*(1), 134–144. https://doi.org/10.1007/s00421-004-1281-4