SCIENTIFIC JOURNAL OF SPORT AND PERFORMANCE

ORIGINAL ARTICLE

Decomposing and classifying Top3 performance across age categories in individual rhythmic gymnastics

Jonatha Flávio Souza Lemos . *Group of Studies in Physical Education and Sports. Sports Analysis Laboratory (GEPEFE/LAE/UFMT). Federal University of Mato Grosso (UFMT). Brazil.*

Andrey Aragão de Souza. *Group of Studies in Physical Education and Sports. Sports Analysis Laboratory (GEPEFE/LAE/UFMT). Federal University of Mato Grosso (UFMT). Brazil.*

Cláudia Marlise Balbinotti Andrade. *Post-Graduate Program in Health Sciences (PPGCS). Federal University of Mato Grosso (UFMT). Brazil*

Henrique de Oliveira Castro. *Group of Studies in Physical Education and Sports. Sports Analysis Laboratory (GEPEFE/LAE/UFMT). Federal University of Mato Grosso (UFMT). Brazil.*

ABSTRACT

This study examined how judging components differentiated Top 3 from non-Top 3 performance across age categories in individual rhythmic gymnastics. Publicly available results from the Brazilian Rhythmic Gymnastics Championships (2023–2026) were analyzed, comprising 890 performances by 288 gymnasts across 109 apparatus finals. Standardized Top 3 advantages were compared using generalized estimating equations, while total-score differences were decomposed by judging component. Separate Random Forest classifiers were developed for each category and interpreted using SHapley Additive exPlanations, with performance assessed across 20 repeated stratified group cross-validation partitions. K-means clustering and principal component analysis were used as complementary exploratory analyses. Top 3 performances obtained higher scores across all judging components. Difficulty of Body accounted for the largest absolute contribution in Junior and Senior. Mean relative SHAP importance identified Difficulty of Body as the leading discriminative component in Pre-Juvenile (33.79%), Execution in Youth (36.05%) and Junior (35.41%), and Artistry in Senior (35.21%). Four overlapping technical-competitive configurations were identified among podium performances, with uniformly high scores across all components strongly associated with first place. Overall, competitive success was multidimensional and category-specific, while absolute score contribution and model-based discriminative importance provided complementary perspectives on Top 3 performance.

Keywords: performance analysis; motor performance; competitive ranking; routine composition; interpretable artificial intelligence.

Cite this article as:
Lemos, J. F. S., de Souza, A. A., Andrade, C. M. B., & Castro, H. de O. (2027). Decomposing and classifying Top3 performance across age categories in individual rhythmic gymnastics. *Scientific Journal of Sport and Performance*, *6*(1), 80-100. https://doi.org/10.55860/IKEC7185

Corresponding author. *Group of Studies in Physical Education and Sports. Sports Analysis Laboratory (GEPEFE/LAE/UFMT). Federal University of Mato Grosso (UFMT). Brazil.*
E-mail: jonatha.lemos@ufmt.br
Submitted for publication August 18, 2026.
Accepted for publication September 25, 2026.
Published November 25, 2026.
Scientific Journal of Sport and Performance. ISSN 2794-0586.
©Asociación Española de Análisis del Rendimiento Deportivo. Alicante. Spain.
doi: https://doi.org/10.55860/IKEC7185

© 2027 ARD Asociación Española

INTRODUCTION

Rhythmic gymnastics is a judged sport in which competitive performance depends on the integration of technical difficulty, apparatus handling, movement quality, musical interpretation, and choreographic composition (Fédération Internationale de Gymnastique [FIG], 2024). Unlike sports in which success is determined by an objective outcome such as time, distance, or goals scored, the final ranking in rhythmic gymnastics results from the evaluation of several interrelated dimensions of the routine. Gymnasts must therefore present technically demanding content while maintaining precision, continuity, expression, and coordination between body movements, apparatus work, and music (FIG, 2024). This multidimensional structure makes it difficult to determine whether competitive success is primarily differentiated by difficulty, execution quality, artistic composition, or the combined contribution of all components.

The final score in individual rhythmic gymnastics is composed of Difficulty of Body (DB), Difficulty of Apparatus (DA), Artistry (A), and Execution (E), with penalties deducted when applicable (FIG, 2024; Freire-Maceiras et al., 2023). DB includes recognized body difficulties such as jumps and leaps, balances, and rotations, whereas DA represents difficulties created through complex interactions between the gymnast and the apparatus. Artistry evaluates the compositional and expressive characteristics of the routine, including musicality, movement connections, variety, and the integration of technical content, while Execution assesses technical faults in body and apparatus work (FIG, 2024). Although each component contributes numerically to the final score, its capacity to distinguish medallists may depend on the competitive context and the extent to which scores vary among finalists.

The relative contribution of these components may differ across age categories because gymnasts in different competitive categories also differ in physical capacities, accumulated training experience, technical repertoire, and routine-construction demands. Flexibility, explosive strength, balance, coordination, and other physical fitness characteristics have been associated with technical execution, competitive level, and performance in rhythmic gymnastics (Donti et al., 2016; Douda et al., 2008; Gaspari et al., 2024). Routine structure may also be influenced by age-specific competitive programs and by the Code of Points applied in each scoring cycle. Changes in the Code can modify the number, type, and value of technical elements and, consequently, the relative contribution of difficulty, artistry, and execution to overall performance (Freire-Maceiras et al., 2025; Sierra-Palmeiro et al., 2019). Therefore, the components that distinguish successful routines in one competitive context may not have the same relevance in another.

Previous research has investigated how judging components contribute to performance in individual rhythmic gymnastics. In a sample of 2,378 routines across different age categories, competitive levels, and apparatuses, DA showed the greatest contribution to the final score, followed by DB and technical execution, whereas the artistic component presented a comparatively smaller contribution (Freire-Maceiras et al., 2023). However, the contribution of each component is shaped by the competitive and regulatory context in which routines are constructed and evaluated. Changes in the Code of Points can shift the relative influence of DB, DA, Artistry, and Execution (Freire-Maceiras et al., 2025), while their importance may also vary according to apparatus, ranking position, and competitive level (Batista et al., 2019b; Leandro et al., 2017; Örs, 2020). Consequently, competitive success cannot be explained by a single universally dominant judging component.

Despite these advances, the ways in which judging components distinguish successful performances across age categories remain unclear. Previous studies have primarily examined associations with final scores or rankings, whereas absolute score contribution, relative competitive advantage, multivariable discriminative importance, and the existence of different technical-competitive configurations represent complementary

aspects of competitive success. Examining these dimensions together may provide a more complete understanding of whether the same judging components differentiate Top 3 performance across age categories and whether podium-ranked routines exhibit similar or distinct combinations of scoring characteristics.

Therefore, the present study aimed to analyse how DB, DA, Artistry, and Execution differentiated Top 3 from non-Top 3 performance across Pre-Juvenile, Youth, Junior, and Senior categories in individual rhythmic gymnastics. Specifically, the study compared the relative Top 3 advantage across judging components, decomposed the total-score difference into its component contributions, evaluated the category-specific discriminative importance of each component using Random Forest classifiers interpreted with SHAP, and explored technical-competitive configurations among podium-ranked performances. It was hypothesized that the relative contribution of the judging components would vary across age categories, without assuming that these cross-sectional differences represent within-athlete developmental progression. Based on previous evidence and the structure of age-specific competitive programs, difficulty and execution were expected to show greater discriminative relevance in younger and intermediate categories, whereas artistic and compositional dimensions were expected to show greater relevance in Senior. It was also expected that first-place performances would be associated with a more balanced and globally elevated scoring configuration.

MATERIALS AND METHODS

Participants

This retrospective observational study analysed publicly available results from the Brazilian Rhythmic Gymnastics Championships held between 2023 and 2026. The primary unit of analysis was a competitive performance, defined as a gymnast's result in a single apparatus final, because the same gymnast could compete in multiple apparatuses, seasons, competitive levels, and age categories.

The final dataset comprised 890 competitive performances produced by 288 gymnasts across 109 apparatus finals. Pre-Juvenile included 90 performances from 45 gymnasts across nine finals; Youth included 192 performances from 61 gymnasts across 24 finals; Junior included 224 performances from 96 gymnasts across 28 finals; and Senior included 384 performances from 130 gymnasts across 48 finals. Category-specific gymnast counts should not be summed because some gymnasts progressed to a higher category during the study period.

The age categories followed the classifications adopted in the analysed competitions: Pre-Juvenile included gymnasts aged 9–10 years, Youth 11–12 years, Junior 13–15 years, and Senior ≥16 years. Competitive levels also differed across categories. Pre-Juvenile included Level Unique (U); Youth included Levels I, II, III, and IV; Junior included Levels I, II, and III; and Senior included Levels I, II, and Elite. Pre-Juvenile gymnasts competed with Hoop, Ball, and Rope, whereas Youth, Junior, and Senior gymnasts competed with Hoop, Ball, Clubs, and Ribbon.

Measures

Competitive success was defined according to the final ranking within each apparatus final. Performances ranked first, second, or third were classified as Top 3, whereas all remaining performances were classified as non-Top 3. This binary classification was used as the outcome in the comparative and machine-learning analyses. When identical total scores occurred across the third-to-fourth-place boundary, Top 3 status followed the official ranking published in the competition results.

 © 2027 ARD Asociación Española

The judging components analysed were Difficulty of Body (DB), Difficulty of Apparatus (DA), Artistry (A), and Execution (E). Penalties (P) represented deductions applied to the routine. Total difficulty corresponded to the sum of DB and DA, while the final score corresponded to the sum of DB, DA, A, and E minus penalties. Competition year, age category, competitive level, apparatus, and final ranking were retained as contextual variables.

Procedures

Competition results were obtained from the official Sporttech results platform for the Brazilian Rhythmic Gymnastics Championships organized by the Brazilian Gymnastics Confederation. Pre-Juvenile, Youth, and Junior results were included from the 2023, 2024, and 2025 editions, while Senior results were included from 2023 to 2026.

All available individual apparatus-final performances from the selected categories and competitive levels were screened. Each apparatus final was identified by the combination of competition year, age category, competitive level, and apparatus. All 109 retained finals contained complete judging results and exactly three performances officially classified within the Top 3.

Repeated participation was identified using a normalized athlete identifier based on each gymnast's full name and year of birth. Accents, punctuation, formatting differences, and common name abbreviations were standardized to prevent the same gymnast from being counted as different individuals. This identifier was also used to quantify repeated participation across seasons and categories and in sensitivity analyses addressing dependence arising from repeated observations of the same gymnast.

The study relied exclusively on unrestricted, publicly accessible competition records and involved no intervention, direct contact with participants, or collection of non-public information. Under Resolution No. 674/2022 of the Brazilian National Health Council, research conducted exclusively with publicly accessible information is exempt from review by the Brazilian CEP/Conep research ethics system. Accordingly, the study was not submitted for ethics committee review, and no formal exemption determination was obtained. Athlete names and years of birth were processed only to identify repeated participation across competitions and were not retained in the analytical outputs. All analyses and reporting were conducted at the group level.

Analysis

All analyses were conducted in Python 3.14.4 using pandas, NumPy, SciPy, statsmodels, scikit-learn, SHAP, and Matplotlib. Statistical significance was set at $p < .05$, and Holm's procedure was used to adjust p-values in multiple post-hoc comparisons (Holm, 1979). Where applicable, effect estimates were accompanied by robust standard errors and 95% confidence intervals.

To account for differences in scoring scales across competitive contexts, DB, DA, A, and E were standardized within each apparatus final. Mean standardized scores were calculated separately for Top 3 and non-Top 3 performances. The relative Top 3 advantage was defined as the mean standardized score of the Top 3 minus that of the non-Top 3 group, with positive values indicating higher relative scores among podium-ranked performances.

Standardized differences in DB, DA, A, and E were analysed using generalized estimating equations with a Gaussian distribution, an exchangeable working-correlation structure, and robust covariance estimators (Liang & Zeger, 1986). Age category, score component, and their interaction were included as fixed factors, while final identifier was used as the clustering variable to account for the four correlated component-specific

values obtained from each final. The primary model therefore contained 436 final-by-component observations (109 finals × four judging components), with Pre-Juvenile and DB used as the reference categories. Global effects were evaluated using Wald tests. Significant interactions were followed by comparisons between components within each category and between categories within each component. Post-hoc contrasts were reported as estimated differences with robust standard errors, 95% confidence intervals, and Holm-adjusted *p*-values.

The consistency of the Top 3 advantage was described as the proportion of finals in which Top 3 performances obtained higher scores than non-Top 3 performances in each component. Penalties were analysed separately because they represent deductions rather than positive judging components. The probability of receiving any penalty was examined using a binomial generalized estimating equation with an exchangeable working-correlation structure and robust covariance estimation. Age category, Top 3 status, and their interaction were included as fixed factors, with apparatus final as the clustering variable. Category-specific Top 3 versus non-Top 3 contrasts were expressed as odds ratios with 95% confidence intervals and Holm-adjusted *p*-values. Penalty magnitudes were additionally summarized descriptively.

The absolute score advantage was decomposed to estimate the number and percentage of points contributed by DB, DA, A, E, and penalties to the total-score difference between Top 3 and non-Top 3 performances. Component contributions were compared across categories using generalized estimating equations, while penalty contributions were reported descriptively. Post-hoc comparisons of component contributions were expressed as estimated point differences with robust standard errors, 95% confidence intervals, and Holm-adjusted *p*-values.

A sensitivity analysis was conducted using only Hoop and Ball because these were the two apparatuses contested in all four categories. Standardized differences from the two apparatuses were averaged within each combination of season, category, and competitive level and analysed to determine whether the category-specific patterns remained after controlling for differences in apparatus programs.

Temporal differences were examined, rather than interpreted as evidence of equivalence or stability, using separate generalized estimating equation models in Senior from 2023 to 2026 and in Junior Levels I and II from 2023 to 2025. Competition year, score component, and their interaction were included as fixed factors. An additional regulatory-context analysis compared performances conducted under the 2022–2024 and 2025–2028 scoring cycles and tested the scoring-cycle × component interaction.

Separate Random Forest classifiers were developed for each age category using DB, DA, A, and E as predictors and Top 3 classification as the outcome (Breiman, 2001). Each model included 1,000 trees, square-root feature selection at each split, balanced class weights, and a fixed random seed. The classifiers were prespecified rather than optimized: the Gini splitting criterion, unrestricted tree depth, a minimum of two observations for splitting, and a minimum leaf size of one were retained, and no hyperparameter tuning was performed. The models were intended to quantify retrospective discrimination within the scoring system rather than prospective prediction of future competitive outcomes.

Model performance was assessed using 20 repetitions of stratified group cross-validation, ensuring that all performances from the same apparatus final remained within the same training or testing fold. Three folds were used for Pre-Juvenile and five folds for Youth, Junior, and Senior. Accuracy, balanced accuracy, sensitivity, specificity, precision, F1-score, and area under the receiver operating characteristic curve were calculated from the combined out-of-fold predictions for each repetition and summarized as mean ± standard

 © 2027 ARD Asociación Española

deviation across the 20 repetitions. Balanced accuracy was considered the primary classification metric. Because the same gymnast could contribute performances to different finals and seasons, a complementary sensitivity analysis repeated cross-validation with gymnast identity, rather than apparatus final, as the grouping variable. Simultaneous separation by both gymnast and final was not feasible because the athlete–final dependency structure formed a single connected component within each age category.

A further sensitivity model added penalties as a fifth predictor (DB, DA, A, E, and P) to determine whether their direct contribution to the official total score materially changed classification performance.

The discriminative contribution of each judging component was interpreted using SHapley Additive exPlanations (Lundberg & Lee, 2017). SHAP values were calculated only for observations in the held-out testing folds. This procedure was repeated across all 20 cross-validation repetitions, and out-of-fold SHAP values were averaged for each performance across repetitions. Global importance was calculated from the mean absolute SHAP value of each component and expressed as a percentage of the total importance across DB, DA, A, and E. These percentages were interpreted as relative importance within the fitted classifier and not as causal or uniquely attributable contributions to competitive success.

The Random Forest and SHAP analyses were repeated using scores standardized within each apparatus final. This sensitivity analysis assessed whether the hierarchy of the judging components remained consistent after accounting for differences in scoring scales. Models based on raw scores were considered the primary classification analyses, whereas standardized-score models were interpreted as measures of contextual discrimination within each completed final. Because within-final standardization uses the score distribution of the complete competitive field, these models were explicitly interpreted as retrospective contextual analyses and not as deployable precompetition prediction models. Within-final percentile ranks were additionally examined as a scaling sensitivity analysis.

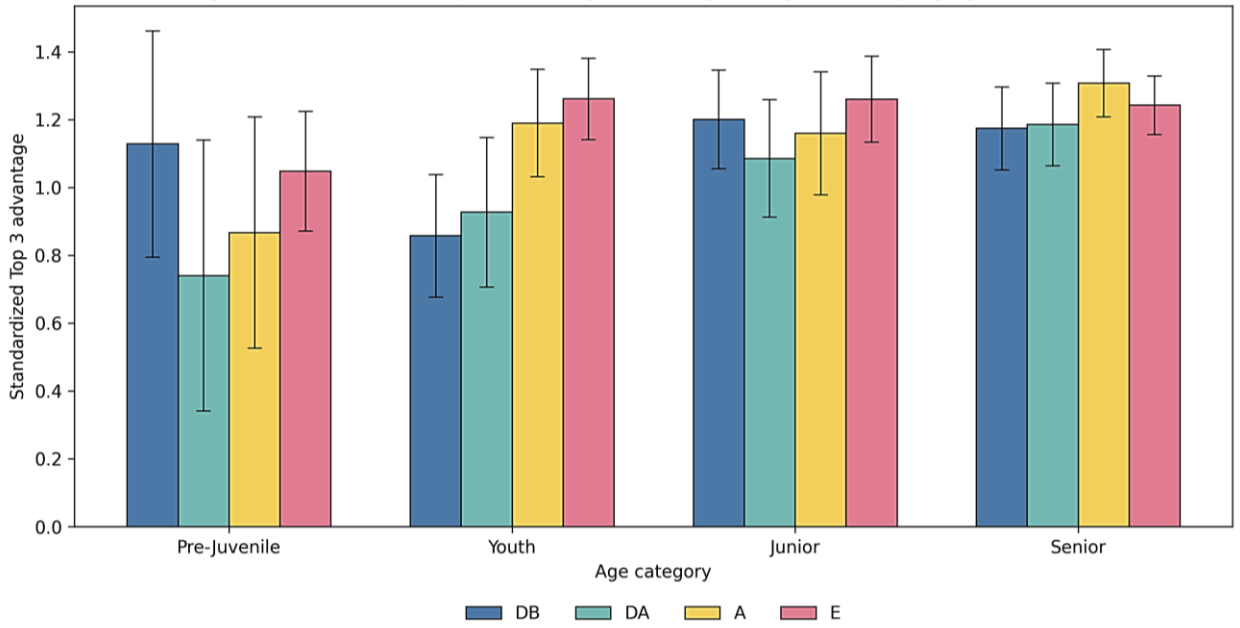
To assess whether ties at the podium boundary affected the findings, the principal GEE and Random Forest analyses were repeated after excluding the five apparatus finals in which third- and fourth-place performances had identical total scores.

As a complementary exploratory analysis, only Top 3 performances were retained to identify overlapping technical-competitive configurations among podium-ranked routines. Standardized DB, DA, A, and E scores were entered into K-means clustering (MacQueen, 1967). Candidate solutions with two to six clusters were compared using the silhouette coefficient, Calinski–Harabasz index, Davies–Bouldin index, and bootstrap stability assessed through the adjusted Rand index. The four-cluster solution was retained based on the combined evidence from these indices and bootstrap stability rather than assuming the existence of discrete natural performance types. The selected cluster solution was visualized using principal component analysis (Jolliffe & Cadima, 2016). Associations between cluster membership, age category, and podium position were examined using chi-square tests and Cramér's V. Adjusted residuals were used to identify relevant category-profile and position-profile combinations, with Holm correction for multiple comparisons.

RESULTS

Standardized Top 3 advantage across categories

Standardized differences between Top 3 and non-Top 3 performances were positive for all score components across the four age categories, indicating higher relative DB, DA, Artistry, and Execution scores among podium-ranked performances (Figure 1).



Note. Positive values indicate higher standardized scores among Top 3 performances. DB = Difficulty of Body; DA = Difficulty of Apparatus; A = Artistry; E = Execution.

Figure 1. Standardized differences between Top 3 and non-Top 3 performances across age categories and judging components.

The generalized estimating equation analysis revealed a significant effect of age category, $\chi^2(3) = 10.4$, $p = .015$, and a significant category × component interaction, $\chi^2(9) = 18.15$, $p = .034$. The overall effect of score component was not significant, $\chi^2(3) = 7.02$, $p = .071$.

No significant differences between components were observed in the Pre-Juvenile, Junior, or Senior categories after Holm adjustment. In Youth, Execution showed a significantly greater standardized advantage than DB (estimated difference = 0.432, 95% CI [0.200, 0.663], Holm-adjusted $p = .002$). Between-category comparisons showed that the DB advantage was lower in Youth than in Junior and Senior (both Holm-adjusted $p = .019$), while the Artistry advantage was lower in Pre-Juvenile than in Senior (Holm-adjusted $p = .045$).

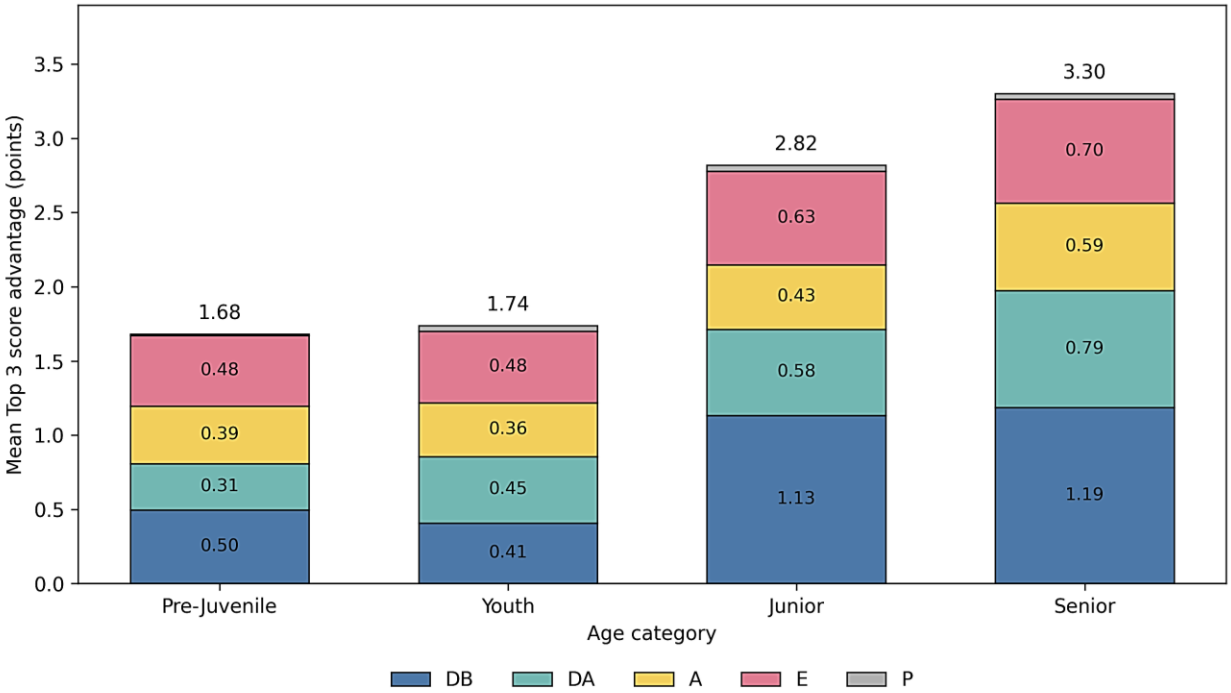
Execution showed the most consistent separation between Top 3 and non-Top 3 performances, with a positive advantage in all analysed finals. Positive DB and Artistry advantages were also observed in most finals, whereas DA showed greater variability in the younger categories. Penalty occurrence was analysed separately using a binomial GEE. There was a significant overall effect of age category, $\chi^2(3) = 8.50$, $p = .037$, but neither Top 3 status, $\chi^2(1) = 0.16$, $p = .685$, nor the category × Top 3 interaction, $\chi^2(3) = 4.04$, $p = .257$, was significant. Although unadjusted contrasts suggested lower odds of receiving a penalty among Top 3 performances in Youth (OR = 0.35, 95% CI [0.14, 0.87]) and Senior (OR = 0.55, 95% CI [0.32, 0.95]), neither comparison remained significant after Holm adjustment ($p = .092$ and $p = .100$, respectively).

Decomposition of the Top 3 score advantage

The mean total-score advantage of Top 3 performances increased from 1.68 points in Pre-Juvenile and 1.74 points in Youth to 2.82 points in Junior and 3.30 points in Senior. In the younger categories, the advantage was distributed relatively evenly across the score components. In contrast, DB became the largest contributor

 © 2027 ARD Asociación Española

in Junior and Senior, accounting for 40.24% and 35.93% of the total difference, respectively. Penalties accounted for only 0.53%–2.34% of the advantage across categories (Figure 2).



Note. Each stacked segment represents the mean contribution of Difficulty of Body, Difficulty of Apparatus, Artistry, Execution, and penalties to the total-score difference between Top 3 and non-Top 3 performances.

Figure 2. Decomposition of the mean Top 3 score advantage across competitive age categories.

The generalized estimating equation analysis showed significant effects of age category, $\chi^2(3) = 67.17$, $p < .001$, and score component, $\chi^2(3) = 9.70$, $p = .021$, as well as a significant category × component interaction, $\chi^2(9) = 37.75$, $p < .001$. No significant differences among component contributions were observed in Pre-Juvenile or Youth.

In Junior and Senior, DB contributed significantly more points than DA, Artistry, and Execution after Holm adjustment. Execution also contributed more than Artistry in both categories, while DA contributed more than Artistry only in Senior. Between-category comparisons confirmed that the DB contribution was greater in Junior and Senior than in the younger categories. These results indicate that the component contributing the largest number of points to the Top 3 score gap was not necessarily the component carrying the greatest discriminative information in the multivariable classifiers described below.

Random Forest classification performance

The primary raw-score Random Forest models showed moderate retrospective discrimination between Top 3 and non-Top 3 performances (Table 1). The reported balanced accuracy was 0.786 in Pre-Juvenile, 0.653 in Youth, 0.686 in Junior, and 0.635 in Senior, while the corresponding ROC-AUC values were 0.892, 0.752, 0.745, and 0.710. Specificity exceeded sensitivity in every category, indicating that the models classified non-Top 3 performances more accurately than podium performances.

Table 1. Repeated out-of-fold classification performance of the Random Forest models across age categories.

Category	Accuracy	Balanced accuracy	Sensitivity	Specificity	Precision	F1-score	ROC-AUC
Pre-Juvenile	0.833	0.786	0.667	0.905	0.750	0.706	0.892
Youth	0.688	0.653	0.514	0.792	0.597	0.552	0.752
Junior	0.723	0.686	0.536	0.836	0.662	0.592	0.745
Senior	0.669	0.635	0.500	0.771	0.567	0.531	0.710

Note. Values are summary performance estimates from 20 repeated stratified group cross-validation partitions; the corresponding between-repetition standard deviations should be reported for every metric. Performances from the same apparatus final were retained within the same training or testing fold. Raw-score models were the primary analyses. Within-final standardized models represent retrospective contextual discrimination because standardization requires knowledge of the complete score distribution within each final. Balanced accuracy was the primary classification metric.

Within-final standardized models produced substantially greater discrimination than the raw-score models, particularly in Youth, Junior, and Senior (Table 1). This improvement indicates that a routine's relative position within its specific competitive field contained considerably more classification information than raw scores pooled across heterogeneous finals. Because these standardized predictors require the complete final-score distribution, however, they represent retrospective contextual classification rather than prospective prediction of an isolated future routine.

When cross-validation was repeated using gymnast identity as the grouping variable, balanced accuracy remained similar to the final-grouped models: 0.723 ± 0.052 in Pre-Juvenile, 0.722 ± 0.017 in Youth, 0.686 ± 0.015 in Junior, and 0.637 ± 0.015 in Senior. ROC-AUC values were 0.881 ± 0.031, 0.805 ± 0.016, 0.771 ± 0.017, and 0.723 ± 0.014, respectively. No gymnast appeared simultaneously in training and testing partitions, indicating that repeated participation by the same athlete did not account for the principal classification findings.

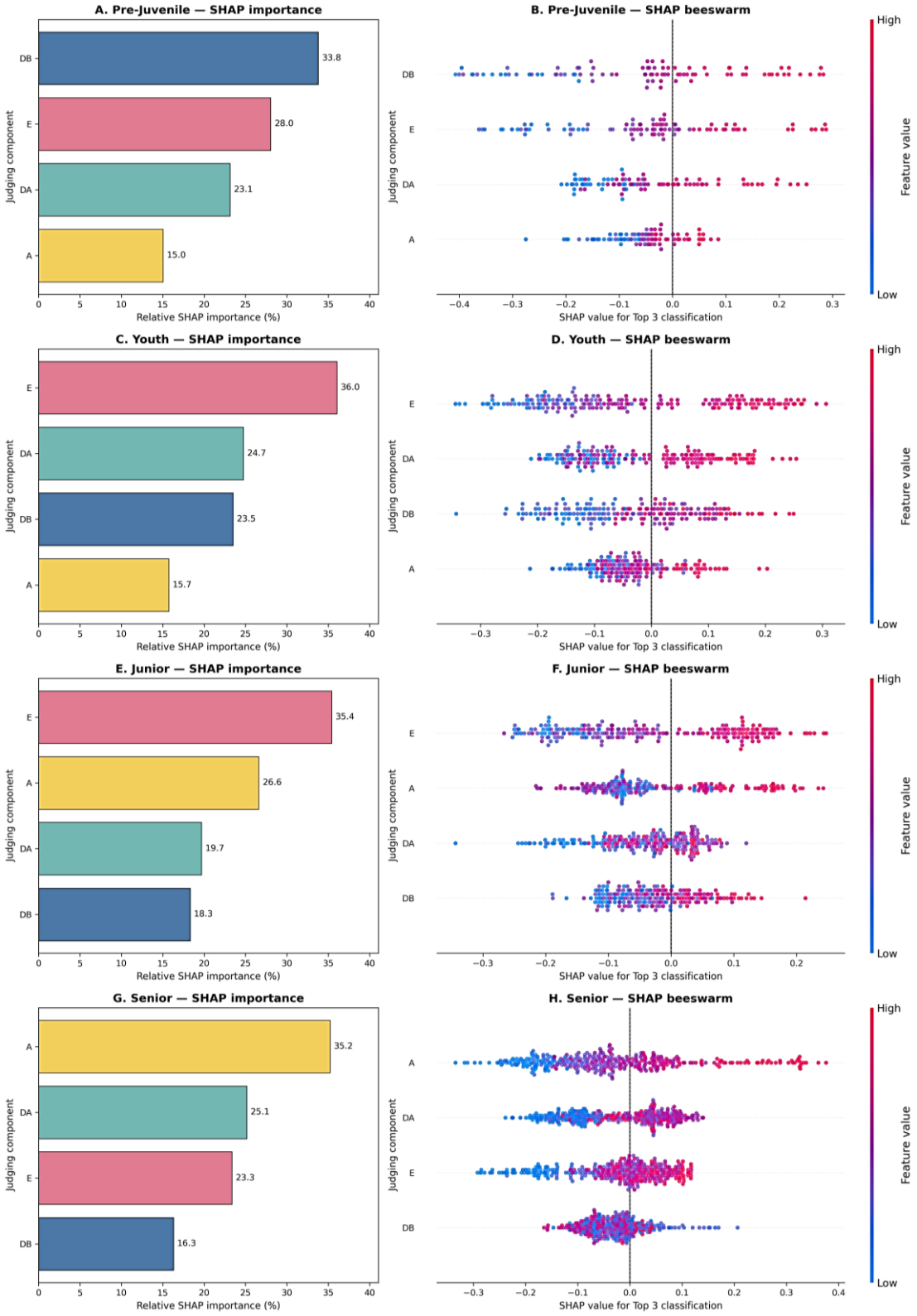
Including penalties as a fifth predictor produced only small changes in classification performance. The change in balanced accuracy ranged from +0.002 to +0.022 across categories, and changes in ROC-AUC ranged from −0.013 to +0.015, indicating that omission of penalties from the primary four-component models did not materially account for their discriminative performance.

SHAP-Based discriminative importance

The relative model-based importance of the judging components differed across age categories (Figure 3). In Pre-Juvenile, DB showed the greatest relative SHAP importance (33.79%), followed by Execution (28.04%), DA (23.13%), and Artistry (15.05%). In Youth, Execution was the leading component (36.05%), followed by DA (24.73%), DB (23.49%), and Artistry (15.73%). Execution also showed the greatest importance in Junior (35.41%), followed by Artistry (26.60%), DA (19.68%), and DB (18.32%). In Senior, Artistry was the leading component (35.21%), followed by DA (25.15%), Execution (23.35%), and DB (16.30%). The leading raw-score component was identical in all 20 repetitions within each category, indicating that the category-specific hierarchy was not dependent on a single cross-validation partition.

SHAP percentages represent relative importance within the fitted multivariable classifier and should not be interpreted as the proportion of competitive success uniquely or causally explained by each judging component. Because DB, DA, Artistry, and Execution arise from the same routine and are correlated dimensions of overall performance quality, their SHAP importance reflects how the fitted model distributed discriminative information among the available predictors.

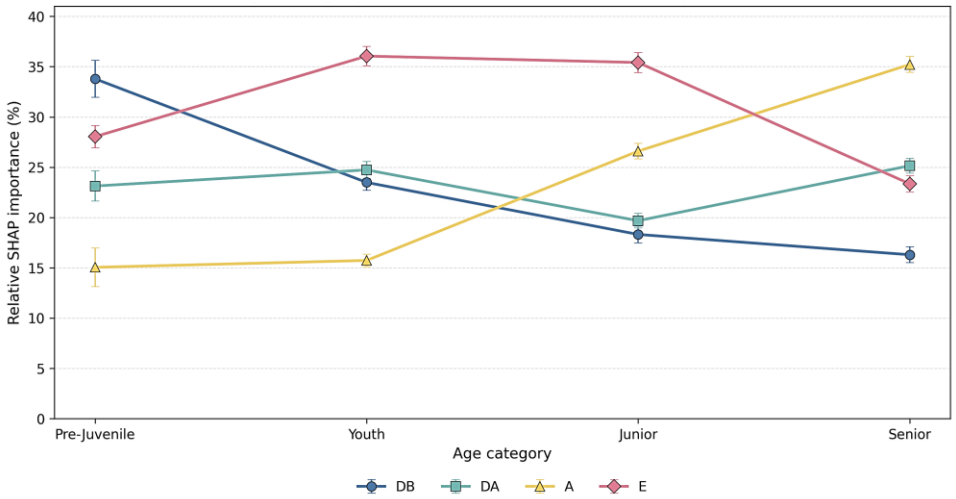
© 2027 ARD Asociación Española



Note. Horizontal bars represent relative mean absolute out-of-fold SHAP importance, while beeswarm plots show the magnitude and direction of SHAP values averaged for each performance across the repeated cross-validation partitions. Positive SHAP values increase the model output toward Top 3 classification, whereas negative values decrease it. Point colour represents the corresponding feature value from low to high.

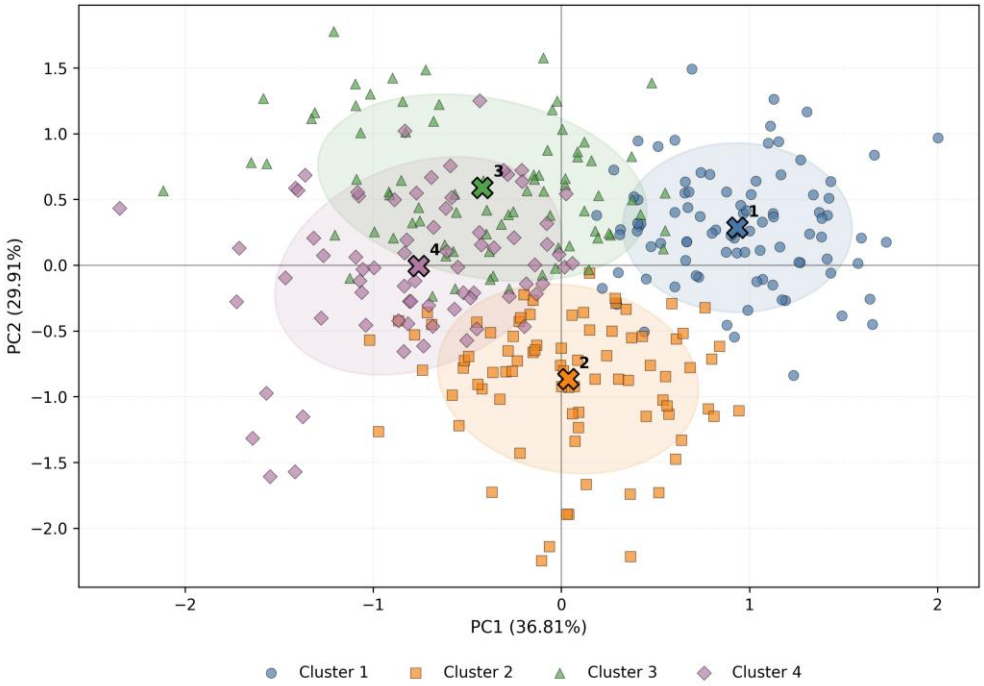
Figure 3. SHAP-based interpretation of the Random Forest classifiers across age categories.

Across categories, the pattern was therefore cross-sectional rather than developmental: DB carried the greatest relative discriminative importance in Pre-Juvenile, Execution in Youth and Junior, and Artistry in Senior (Figure 4). This category-specific hierarchy contrasts with the absolute score decomposition, particularly in Junior and Senior, where DB contributed the largest number of points to the Top 3 advantage despite Execution and Artistry, respectively, carrying greater model-based discriminative importance.



Note. Values represent the percentage contribution of each component to total mean absolute out-of-fold SHAP importance within each category. Error bars represent variability across the 20 repeated cross-validation partitions.

Figure 4. Age-category differences in the relative SHAP importance of the judging components.



Note. PC1 and PC2 explained 36.81% and 29.91% of the variance, respectively (66.72% combined). Ellipses illustrate within-cluster dispersion, and X markers indicate cluster centroids.

Figure 5. Principal component analysis representation of the four exploratory Top 3 performance configurations.

© 2027 ARD Asociación Española

Complementary exploratory analysis of Top 3 performance configurations

Among the 327 Top 3 performances, the four-cluster solution showed the highest silhouette coefficient among the evaluated solutions (k = 2–6) and the greatest bootstrap stability, although overall cluster separation remained modest (silhouette coefficient = 0.247; Calinski–Harabasz index = 109.20; Davies–Bouldin index = 1.332; mean adjusted Rand index = 0.785 ± 0.093). The resulting groups were therefore treated as overlapping exploratory configurations rather than discrete natural performance types. The component-specific centroids and sizes of the four clusters are presented in Table 2.

Cluster 1 showed uniformly elevated standardized scores across DB, DA, Artistry, and Execution. Cluster 2 was characterized by comparatively low DB alongside higher DA, Artistry, and Execution; Cluster 3 combined elevated DB with comparatively low DA; and Cluster 4 showed positive DB and DA but comparatively lower Artistry and Execution. These descriptions refer to centroid patterns and were not interpreted as fixed athlete or routine types.

Table 2. Standardized judging-component characteristics of the four exploratory Top 3 performance configurations.

Profile	n	DB	DA	Artistry	Execution
High overall performance	91	1.314	1.130	1.307	1.252
Difficulty-dominant	65	0.689	0.956	0.086	0.196
Artistic-executive with low DB	81	−0.108	0.851	1.089	0.965
DB-dominant with low DA	82	0.952	−0.103	0.516	0.696

Note. Values for DB, DA, Artistry, and Execution are cluster centroids expressed as within-final standardized scores. The n column reports the number of Top 3 performances assigned to each cluster; the four cluster counts sum to 327.

Cluster membership was significantly associated with age category, $\chi^2(9) = 18.50$, $p = .030$, although the association was small (Cramér's V = 0.137). No individual category × cluster combination remained significant after Holm correction, indicating that the global association should not be interpreted as evidence of sharply category-specific profiles.

A substantially stronger association was observed between cluster membership and podium position, $\chi^2(6) = 141.89$, $p < .001$, Cramér's V = 0.466. Cluster 1 contained 73 first-place, 18 second-place, and no third-place performances; thus, 80.2% of Cluster 1 performances were winners. First place was significantly overrepresented in Cluster 1, whereas second- and third-place performances were underrepresented. Cluster 2 showed an overrepresentation of second-place performances, while Cluster 4 showed a strong overrepresentation of third-place performances. These associations describe structure within the same scoring system used to determine the official ranking and should not be interpreted as causal winning strategies.

Sensitivity, temporal, and regulatory-context analyses

The analysis restricted to Hoop and Ball indicated that the relationship between age category and score component was not explained solely by differences in apparatus programs. Significant effects were observed for age category, $\chi^2(3) = 21.66$, $p < .001$, and the category × component interaction, $\chi^2(9) = 31.60$, $p < .001$, whereas the overall effect of score component was not significant, $\chi^2(3) = 3.48$, $p = .323$. In Youth, Artistry and Execution produced greater standardized advantages than DB and DA, while Artistry exceeded DA in Senior. Between-category comparisons also showed lower DB advantages in Youth than in Junior and Senior.

Using within-final percentile ranks instead of z-scores yielded similarly high contextual discrimination. Balanced accuracy was 0.798 in Pre-Juvenile, 0.837 in Youth, 0.865 in Junior, and 0.898 in Senior, compared with 0.790, 0.873, 0.880, and 0.911 for the corresponding z-score models. ROC-AUC values showed the same pattern, ranging from 0.919 to 0.967 for percentile-rank models and from 0.914 to 0.973 for z-score models. These findings indicate that the improvement observed after within-final scaling was not specific to z-score standardization.

No statistically detectable year effect was observed in Senior from 2023 to 2026, $\chi^2(3) = 1.06$, $p = .786$, nor was there evidence of a year × component interaction, $\chi^2(9) = 7.27$, $p = .609$. Similarly, Junior Levels I and II showed no statistically detectable year effect from 2023 to 2025, $\chi^2(2) = 0.03$, $p = .986$, or year × component interaction, $\chi^2(6) = 4.17$, $p = .654$. These nonsignificant tests were interpreted as an absence of statistically detectable temporal differences rather than evidence of equivalence or temporal stability.

The regulatory-context analysis likewise showed no statistically detectable scoring-cycle effect in Senior, $\chi^2(1) = 0.04$, $p = .844$, and no scoring-cycle × component interaction, $\chi^2(3) = 2.06$, $p = .560$. In Junior Levels I and II, neither the scoring-cycle effect, $\chi^2(1) < 0.01$, $p = .961$, nor the scoring-cycle × component interaction, $\chi^2(3) = 2.85$, $p = .415$, was significant. These findings do not establish equivalence between scoring cycles but indicate that the available data did not provide evidence of a systematic cycle-related change in the category-specific component patterns.

Five apparatus finals contained identical total scores across the third-to-fourth-place boundary. When these finals were excluded, 848 performances from 104 finals remained. The category × component interaction persisted for both the standardized Top 3 advantage, $\chi^2(9) = 24.40$, $p = .004$, and the absolute score decomposition, $\chi^2(9) = 42.64$, $p < .001$. Random Forest balanced accuracy changed by no more than 0.008 and ROC-AUC by no more than 0.022 in any category, indicating that boundary ties did not account for the principal findings.

DISCUSSION

The present study investigated how judging components differentiated Top 3 from non-Top 3 performance across age categories in individual rhythmic gymnastics. Podium-ranked routines achieved higher DB, DA, Artistry, and Execution scores than the remaining finalists, although the relative contribution of these components changed across categories. The total-score advantage was relatively balanced among components in Pre-Juvenile and Youth, whereas DB accounted for the largest absolute share of the difference in Junior and Senior. Model-based discriminative importance showed a different category-specific pattern, with DB emerging as the leading component in Pre-Juvenile, Execution in Youth and Junior, and Artistry in Senior. The complementary cluster analysis further indicated that podium performance could be achieved through different overlapping technical-competitive configurations, although first place was strongly associated with high and balanced scores across all components. These findings reinforce the multidimensional nature of competitive success in rhythmic gymnastics and indicate that the features distinguishing Top 3 from non-Top 3 performances differed across age categories and competitive contexts (Freire-Maceiras et al., 2023; Lemos et al., 2026).

The positive standardized gaps observed for all score components indicate that podium performance was characterized by broad superiority rather than by an advantage restricted to one judging domain. Execution was particularly consistent, presenting a positive advantage in every analysed final. Technical precision, body and apparatus control, and the ability to minimize visible faults may therefore represent important

 © 2027 ARD Asociación Española

characteristics associated with competitive success across age categories (Ciorăsteanu et al., 2026; Donti et al., 2016; Zhou et al., 2025). Because execution deductions directly reduce the score obtained from the content presented, high difficulty may provide limited competitive benefit when accompanied by technical errors (FIG, 2024, 2026). DA showed greater variability in the younger categories, suggesting that apparatus-difficulty content may not consistently separate medallists from the remaining finalists in these categories (Freire-Maceiras et al., 2023).

The significant category × component interaction indicates that the relative capacity of each component to differentiate Top 3 performance changed across age groups. In Youth, the Execution advantage was greater than the DB advantage, whereas the DB gap was greater in Junior and Senior than in Youth. One possible explanation lies in the degree of freedom allowed in routine construction. The category- and level-specific programs adopted in the analysed national competitions can restrict the number, value, or type of difficulty content available to younger gymnasts, thereby reducing between-gymnast variability in D scores (Confederação Brasileira de Ginástica [CBG], 2026). When gymnasts perform similar required elements or operate within comparable difficulty limits, competitive separation may depend more strongly on how accurately and consistently that content is executed. This interpretation is consistent with evidence that the relative contribution of the scoring components changes according to category, level, apparatus, and the regulatory framework under which routines are constructed (Freire-Maceiras et al., 2023, 2025).

Differences between age categories may also reflect differences in accumulated training exposure and in the physical and technical capacities typically observed across competitive levels. Flexibility, strength, power, balance, coordination, and sport-specific motor control have been consistently associated with rhythmic-gymnastics performance and competitive level (Donti et al., 2016; Douda et al., 2008; Gaspari et al., 2024). Recent evidence also indicates age-related differences in postural stability among young rhythmic gymnasts and shows that complex body difficulties depend on precise segmental coordination and rotational control (Aleksić Veljković et al., 2025; Liu et al., 2025). These factors may contribute to the larger DB advantage observed in Junior and Senior than in Youth, although the present cross-sectional design cannot establish a developmental progression. Artistry also produced a greater standardized advantage in Senior than in Pre-Juvenile, which is consistent with the possibility that compositional quality, musical interpretation, movement variety, and the integration of expressive and technical content play a greater differentiating role in advanced categories (Ciorăsteanu & Balint, 2025; Toledo & Antualpa, 2016).

The decomposition of the total-score difference provided a complementary perspective. In Pre-Juvenile and Youth, the advantage was distributed relatively evenly across DB, DA, Artistry, and Execution, suggesting that the observed Top 3 advantage in these categories reflected the accumulation of smaller gains across the complete routine. In Junior and Senior, however, DB became the largest absolute contributor to the score advantage. Higher-ranked gymnasts have previously been shown to differ in the diversity and complexity of the body difficulties included in their routines (Batista et al., 2019a, 2019b). Moreover, recent biomechanical research confirms that high-value body elements require specific kinetic and kinematic characteristics that may help explain larger DB differences among advanced competitors (Liu et al., 2025; Nogueras et al., 2026). Penalties contributed minimally across all categories, indicating that competitive separation was mainly generated by the scored components rather than by formal deductions. Consistent with this interpretation, Top 3 status was not significantly associated with penalty occurrence after adjustment for multiple comparisons, and adding penalties as a fifth Random Forest predictor produced only minor changes in classification performance.

The larger absolute contribution of DB in Junior and Senior does not contradict the finding that Execution and Artistry were the leading SHAP components in these categories. Score decomposition and SHAP importance answer different questions. Decomposition identifies where the numerical difference in the final score was accumulated, whereas SHAP estimates how much relative discriminative information the fitted classifier attributed to each component when all predictors were considered simultaneously (Lundberg & Lee, 2017). DB may therefore generate a large difference in points but provide less distinctive classification information when it covaries with Execution, Artistry, or DA. This distinction is particularly relevant in rhythmic gymnastics, where difficulty content, compositional construction, and performance quality are interdependent within the same routine (Ciorăşteanu & Balint, 2025; Freire-Maceiras et al., 2023, 2025). Accordingly, SHAP percentages should not be interpreted as causal contributions or as the proportion of competitive success uniquely explained by each component.

The Random Forest models showed moderate retrospective classification performance using raw scores, with the highest balanced accuracy and ROC-AUC observed in Pre-Juvenile. This result may reflect clearer score separation between medallists and the remaining finalists at this stage. However, it should be interpreted cautiously because the Pre-Juvenile model was based on only nine finals and may therefore be more sensitive to sample-specific characteristics. Substantial variation in model performance across age categories has also been observed in group rhythmic gymnastics, suggesting that classification performance may depend on category-specific score dispersion, technical homogeneity, and sample size (Lemos et al., 2026). In all categories, specificity was higher than sensitivity, indicating that the models identified non-Top 3 performances more accurately than they identified all podium-ranked routines. The markedly higher performance of the within-final standardized models, particularly in Youth, Junior, and Senior, further indicates that relative standing within a specific competitive field carries more discriminative information than raw scores pooled across heterogeneous finals. Part of this gain is also structurally expected because both the transformed predictors and Top 3 classification are defined relative to the same competitive field. These standardized models are nevertheless retrospective because the complete score distribution of the final is required to calculate the transformed predictors.

The SHAP analysis revealed a category-specific pattern in the leading discriminative components of Top 3 classification. In Pre-Juvenile, DB was the most important component, closely followed by Execution. In the Pre-Juvenile category, differences in the ability to perform recognized body difficulties may be particularly visible because younger gymnasts are typically still acquiring the physical and technical capacities required for leaps, balances, rotations, and flexibility-based elements (Douda et al., 2008; Gaspari et al., 2024). Age-related differences in postural control and the biomechanical complexity of high-value body elements support the view that the realization of DB depends on capacities that differ substantially among young gymnasts (Aleksić Veljković et al., 2025; Liu et al., 2025). Nevertheless, the similar contribution of Execution indicates that difficulty alone was insufficient to characterize the model's discrimination and that successful routines also depended on performing the proposed content with adequate control.

Execution showed the greatest SHAP importance in Youth and Junior. These categories may represent competitive contexts characterized by greater routine complexity than younger competitive programs, while gymnasts differ in their ability to perform that content without substantial technical faults. Physical fitness contributes to technical execution in young rhythmic gymnasts, while the successful performance of apparatus actions depends on perceptual-motor control that changes according to task difficulty and apparatus type (Donti et al., 2016; Zhou et al., 2025). If category regulations establish comparable required content or difficulty restrictions, variability in D scores may be reduced, making execution quality more discriminative within the scoring system (CBG, 2026). The predominance of Execution in these categories is

 © 2027 ARD Asociación Española

also consistent with recent findings in group rhythmic gymnastics, in which Execution was the leading classification component in the younger categories (Lemos et al., 2026). From a practical perspective, these findings are compatible with emphasizing the quality of execution when increasing routine complexity, although the present observational data cannot determine the effect of specific training strategies.

In Senior, Artistry emerged as the leading discriminative component, while DA, Execution, and DB retained relevant contributions. At advanced levels, difficulty and execution may be more consistently developed among finalists, reducing their ability to distinguish between technically accomplished gymnasts. Artistry may then provide additional separation through musical interpretation, compositional coherence, movement variety, expression, and the integration of body and apparatus work. The artistic dimension has been progressively redefined across Codes of Points and is closely connected to physical, technical, and expressive capacities (Kritikou et al., 2017; Toledo & Antualpa, 2016). Recent interventions also indicate that artistic preparation involving dance, ballet, theatre, motor control, and body aesthetics can improve or favourably influence routine performance, although individual responses remain heterogeneous (Ciorășteanu et al., 2026; Ciorășteanu & Balint, 2025). Thus, one possible interpretation is that Artistry provides additional discrimination among Senior finalists whose routines already contain substantial levels of difficulty and execution quality; however, this explanation remains inferential and cannot be established from the present design.

Group and individual rhythmic gymnastics appear to present partially different category-specific patterns. In group routines, Execution predominated in younger categories, while DB and DA became more influential in Junior and Senior; in the present individual analysis, Artistry emerged as the leading discriminative component in Senior (Lemos et al., 2026). The difference may reflect the specific structural demands of the two events. Group routines require collective synchronization, exchanges, collaborations, and coordinated apparatus handling, which may increase the competitive importance of difficulty and collective execution. Individual routines, in contrast, may permit greater differentiation through gymnast-specific composition, interpretation, and expressive organization. These findings reinforce that performance models should not be generalized indiscriminately between age categories or event formats (Freire-Maceiras et al., 2023; Lemos et al., 2026).

The sensitivity analyses supported the consistency of several main findings across alternative analytical specifications. The leading raw-score SHAP component was identical across all 20 repeated cross-validation partitions in each category. When within-final z-scores were used, the mean leading component remained DB in Pre-Juvenile, Execution in Youth and Junior, and Artistry in Senior, although DB led 15 of 20 standardized Pre-Juvenile repetitions and Execution led the remaining five. The analysis restricted to Hoop and Ball also indicated that the results were not solely explained by differences in apparatus programs. Within-final percentile ranks produced similarly high contextual discrimination to z-score standardization, indicating that the improvement associated with within-final scaling was not dependent on a single transformation. Cross-validation grouped by gymnast also yielded classification performance similar to the primary final-grouped analysis, reducing concern that repeated observations of the same athlete accounted for the main Random Forest results. Excluding the five finals with ties across the third-to-fourth-place boundary likewise produced only minor changes in classification performance and preserved the principal GEE interactions.

The temporal analyses found no statistically detectable year effects or year × component interactions in Senior or in Junior Levels I and II. Similarly, no statistically detectable scoring-cycle effects or scoring-cycle × component interactions were observed when the 2022–2024 and 2025–2028 regulatory contexts were

compared. These findings should not be interpreted as evidence of temporal or regulatory equivalence. The contribution of judging components is known to vary according to age category, competitive level, apparatus, and scoring context (Freire-Maceiras et al., 2023). Changes in the Code of Points have been shown to reshape the performance model in individual rhythmic gymnastics, shifting it from a marked influence of DA toward a more balanced contribution of the scoring components (Freire-Maceiras et al., 2025). Accordingly, the present results indicate only that systematic temporal or scoring-cycle differences were not detected in the available national-level sample. The analyses were not designed or powered as formal equivalence tests and cannot establish that regulatory changes had no influence on routine construction or scoring.

The complementary cluster analysis showed that podium-ranked routines were distributed across four overlapping technical-competitive configurations. This finding is compatible with the presence of multiple combinations of judging-component scores among podium routines, including globally elevated scores and configurations characterized by comparatively stronger or weaker components. Such heterogeneity is consistent with the multidimensional structure of rhythmic gymnastics and with evidence that physical, technical, motor-control, and artistic dimensions may contribute differently to competitive performance (Ciorăşteanu et al., 2026; Freire-Maceiras et al., 2023). However, the modest silhouette coefficient indicates partial overlap between the clusters. The clusters should therefore be interpreted as exploratory configurations within this dataset rather than as clearly separated, natural, or permanent types of gymnasts.

Cluster membership showed only a small association with age category, and no individual category × cluster combination remained significant after adjustment. Consequently, the observed distribution of clusters across categories should not be treated as evidence of fixed developmental trajectories. The association with podium position was considerably stronger. Cluster 1, characterized by uniformly high standardized scores across all four components, was strongly overrepresented among first-place performances, whereas Cluster 2 was overrepresented among second-place performances and Cluster 4 among third-place performances. This pattern indicates an association between broadly elevated component scores and winning within the same scoring system, but it should not be interpreted as evidence that a particular technical configuration causally produces victory. Because the final score integrates these complementary dimensions, weakness in one domain may limit the competitive benefit provided by excellence in another (FIG, 2024, 2026; Freire-Maceiras et al., 2023).

These findings may inform hypotheses for category-specific training and routine construction. In Pre-Juvenile, the joint importance of DB and Execution suggests that increases in body-difficulty content should be considered alongside technical control rather than in isolation. In Youth and Junior, the prominence of Execution is consistent with emphasizing movement precision, postural control, apparatus handling, and the quality of progressively complex content (Aleksić Veljković et al., 2025; Gaspari et al., 2024; Zhou et al., 2025). In Senior, the greater relative SHAP importance of Artistry highlights the potential relevance of integrating technical preparation with artistic and compositional development, including musical interpretation, movement variety, expression, and coherence between body and apparatus work (Ciorăşteanu et al., 2026; Ciorăşteanu & Balint, 2025). Across all categories, the strong association between Cluster 1 and first place is compatible with the practical value of balanced performance across scoring domains. However, because the study was observational and retrospective, these findings should be viewed as descriptive guidance rather than evidence that modifying a specific component will cause an improvement in competitive ranking.

The findings should also be interpreted in light of several limitations. The study analysed Brazilian national-level competitions, and the results may not be directly generalizable to international events, other competitive

 © 2027 ARD Asociación Española

systems, or gymnasts from different countries. The number of finals differed across categories, particularly in Pre-Juvenile, which may have affected model stability. Competitive levels and apparatus programs also varied across categories and seasons, although the sensitivity and temporal analyses addressed these sources of heterogeneity to some extent. In addition, the same gymnast could contribute multiple performances across apparatuses, seasons, levels, and categories. The primary cross-validation grouped performances by apparatus final, whereas a sensitivity analysis grouped observations by gymnast and prevented the same athlete from appearing simultaneously in training and testing sets. The similar results obtained under this alternative grouping reduce, but do not eliminate, concerns related to dependence from repeated participation. The primary inferential and classification interpretations therefore remain at the performance level rather than at the athlete level.

A further limitation is that Top 3 classification was derived from a final score composed of the same judging components used as predictors. The Random Forest models consequently evaluated retrospective discrimination within the scoring system rather than prospective success based on variables independent of the final ranking. The SHAP results should therefore be interpreted as relative model-based discriminative importance rather than causal effects. The substantially greater performance of within-final standardized models also reflects the incorporation of information about each routine's position relative to the complete competitive field; these analyses are consequently contextual and retrospective rather than deployable prediction models. Similarly, the cluster analysis was exploratory and showed only modest separation between configurations. The identified clusters describe patterns among podium performances but should not be interpreted as stable athlete characteristics. Finally, the cross-sectional comparison of age categories cannot establish developmental transitions. Longitudinal studies following the same gymnasts across competitive categories would be required to determine whether the category-specific patterns observed here reflect within-athlete changes over time.

CONCLUSION

Top 3 performance in individual rhythmic gymnastics was associated with higher scores across all judging components, but the components that most strongly distinguished competitive success varied across age categories. The total-score advantage was relatively balanced in Pre-Juvenile and Youth, whereas DB accounted for the largest absolute contribution in Junior and Senior. Model-based discriminative importance followed a different pattern, with DB leading in Pre-Juvenile, Execution in Youth and Junior, and Artistry in Senior. The exploratory configuration analysis further indicated that different technical-competitive configurations were observed among podium performances, although first place was strongly associated with high and balanced performance across all components. Overall, the findings show that the technical-competitive characteristics associated with Top 3 performance differ across age categories and that absolute score contribution and multivariable discriminative importance provide complementary perspectives on competitive performance.

AUTHOR CONTRIBUTIONS

All authors meet the criteria for authorship in accordance with established ethical guidelines. Contributions are specified according to the CRediT (Contributor Roles Taxonomy) as follows:

Conceptualisation: Jonatha Flávio Souza Lemos. Methodology: Jonatha Flávio Souza Lemos. Formal analysis: Jonatha Flávio Souza Lemos. Investigation: Jonatha Flávio Souza Lemos and Andrey Aragão de Souza. Data curation: Andrey Aragão de Souza. Writing – original draft: Jonatha Flávio Souza Lemos. Writing

– review & editing: Cláudia Marlise Balbinotti Andrade and Henrique de Oliveira Castro. Supervision: Cláudia Marlise Balbinotti Andrade and Henrique de Oliveira Castro.

All authors have critically reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

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

AI USE DISCLOSURE

During the preparation of this work, the authors used Artificial Intelligence (AI) tools exclusively for the following purposes: (1) data processing and statistical code generation in Python 3.14.4 using scikit-learn, scipy, pandas, and shap libraries; and (2) language translation from Portuguese to British English. After using these AI tools, the authors thoroughly reviewed, edited, and validated all content, statistical outputs, interpretations, and conclusions. The authors retain full responsibility for the content of the manuscript and confirm its originality, integrity, and accuracy.

REFERENCES

Aleksić Veljković, A. Z., Marković, A. S., Vukadinović Jurišić, M. D., Obradović, J. M., & Katanić, B. D. (2025). Age-related differences in postural stability and asymmetries among female rhythmic gymnasts: Implications for injury prevention and performance enhancement. *The Journal of Sports Medicine and Physical Fitness, 65*(4), 546–553. https://doi.org/10.23736/S0022-4707.24.16136-1

Batista, A., Garganta, R., & Ávila-Carvalho, L. (2019a). Biological maturation of Portuguese rhythmic gymnasts in different competition levels of performance. *International Journal of Physical Education, Fitness and Sports, 8*(1), 38–48. https://doi.org/10.26524/ijpefs1915

Batista, A., Garganta, R., & Ávila-Carvalho, L. (2019b). Body difficulties in rhythmic gymnastics routines. *Science of Gymnastics Journal, 11*(1), 37–55. https://doi.org/10.52165/sgj.11.1.37-55

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Ciorăşteanu, R. M., & Balint, L. (2025). Body aesthetics and the artistic component in rhythmic gymnastics: The impact of a specific training program. *Bulletin of the Transilvania University of Braşov, Series IX: Sciences of Human Kinetics, 18*(67), 57–66. https://doi.org/10.31926/but.shk.2025.18.67.1.7

Ciorăşteanu, R. M., Mijaica, R., & Balint, L. (2026). Ideomotor representation, motor control, and body aesthetics in juvenile rhythmic gymnasts: Between reciprocal support and differentiated trajectories within a personalized training program. *Frontiers in Sports and Active Living, 8*, 1874791. https://doi.org/10.3389/fspor.2026.1874791

 © 2027 ARD Asociación Española

Confederação Brasileira de Ginástica. (2026, January). *Regulamento técnico específico 2026—Ginástica rítmica*. https://sge.cbginastica.com.br/_uploads/orgaoAnexo/1AdG8UPi4Z1_K8WMNee9Q-QuQ9YRwvSir.pdf

Donti, O., Bogdanis, G. C., Kritikou, M., Donti, A., & Theodorakou, K. (2016). The relative contribution of physical fitness to the technical execution score in youth rhythmic gymnastics. *Journal of Human Kinetics, 51*(1), 143–152. https://doi.org/10.1515/hukin-2015-0183

Douda, H. T., Toubekis, A. G., Avloniti, A. A., & Tokmakidis, S. P. (2008). Physiological and anthropometric determinants of rhythmic gymnastics performance. *International Journal of Sports Physiology and Performance, 3*(1), 41–54. https://doi.org/10.1123/ijspp.3.1.41

Fédération Internationale de Gymnastique. (2024). *Rhythmic gymnastics code of points 2025–2028*. https://www.gymnastics.sport/publicdir/rules/files/en_1.1%20-%20RG%20Code%20of%20Points%202025-2028.pdf

Fédération Internationale de Gymnastique. (2026). *Technical regulations 2026*. https://www.gymnastics.sport/publicdir/rules/files/en_1.1%20-%20Technical%20Regulations%202026.pdf

Freire-Maceiras, R., Fernández-Villarino, M., & Sierra-Palmeiro, E. (2023). Components of performance in individual rhythmic gymnastics. *Retos, 49*, 16–21. https://doi.org/10.47197/retos.v49.97981

Freire-Maceiras, R., Sierra-Palmeiro, E., & Fernández-Villarino, M. (2025). Influence of Code of Points change on individual rhythmic gymnastics performance. *Science of Gymnastics Journal, 17*(2), 331–339. https://doi.org/10.52165/sgj.17.2.331-339

Gaspari, V., Bogdanis, G. C., Panidi, I., Konrad, A., Terzis, G., Donti, A., & Donti, O. (2024). The importance of physical fitness parameters in rhythmic gymnastics: A scoping review. *Sports, 12*(9), 248. https://doi.org/10.3390/sports12090248

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Kritikou, M., Donti, O., Bogdanis, G. C., Donti, A., & Theodorakou, K. (2017). Correlates of artistry performance scores in rhythmic gymnasts. *Science of Gymnastics Journal, 9*(2), 165–177. https://doi.org/10.52165/sgj.9.2.165-177

Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., & Bobo-Arce, M. (2017). Judging in rhythmic gymnastics at different levels of performance. *Journal of Human Kinetics, 60*(1), 159–165. https://doi.org/10.1515/hukin-2017-0099

Lemos, J. F. S., Figueiredo, L. S., De Oliveira, V., Gonçalves, L. C. O., Andrade, C. M. B., & De Oliveira Castro, H. (2026). From execution to difficulty: Age-category shifts in judging determinants of podium success in group rhythmic gymnastics. *Team Sports Studies, 2*, 100011. https://doi.org/10.1016/j.tss.2026.100011

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

Liu, T., Wang, L., & Sun, Y. (2025). Kinematic analysis of free vertical split with 720° turn in elite Chinese rhythmic gymnastics. *Sensors, 25*(9), 2667. https://doi.org/10.3390/s25092667

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon et al. (Eds.), Advances in Neural Information Processing Systems *30* (pp. 4765–4774). Curran Associates.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Vol.* 1. *Statistics* (pp. 281–297). University of California Press.

Nogueras, M., Floria, P., & Ferro-Sánchez, A. (2026). Identification of kinetic efficacy variables for the rhythmic gymnastics pike jump to monitor performance. *Biomechanics, 6*(1), 19. https://doi.org/10.3390/biomechanics6010019

Örs, B. S. (2020). The effect of difficulty and execution scores on total ranking during 2019 Rhythmic Gymnastics World Championships. *African Educational Research Journal, 8*(Suppl. 1), S37–S42. https://doi.org/10.30918/AERJ.8S1.20.005

Sierra-Palmeiro, E., Bobo-Arce, M., Pérez-Ferreirós, A., & Fernández-Villarino, M. A. (2019). Longitudinal study of individual exercises in elite rhythmic gymnastics. *Frontiers in Psychology, 10*, 1496. https://doi.org/10.3389/fpsyg.2019.01496

Toledo, E. de, & Antualpa, K. (2016). The appreciation of artistic aspects of the Code of Points in rhythmic gymnastics: An analysis of the last three decades. *Revista Brasileira de Educação Física e Esporte, 30*(1), 119–131. https://doi.org/10.1590/1807-55092016000100119

Zhou, C., Cai, Y., Li, K., Zhang, X., & Tang, C. (2025). Research on visual search strategies during apparatus throw-and-catch in group rhythmic gymnastics. Frontiers in Psychology, 16, 1700617. https://doi.org/10.3389/fpsyg.2025.1700617

This work is licensed under a Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

 © 2027 ARD Asociación Española