INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN PROGRESSIVE EDUCATION & DEVELOPMENT

OPEN ACCESS JOURNAL

Vol 15, Issue 4, (2026) E-ISSN: 2226-6348

Restricted-Round Performance and Winning Margins in Olympic Synchronised Diving, 2000–2024

Hussain Haleem

Faculty of Sports Science and Coaching, Universiti Pendidikan Sultan Idris, Tanjong Malim, Perak, Malaysia

Corresponding Author Email: kudahusen@gmail.com

DOI Link: http://dx.doi.org/10.6007/IJARPED/v15-i4/29104

Published Online: 25 October 2026

Abstract

Olympic synchronised diving begins with two restricted rounds assigned a degree of difficulty (DD) of 2.0, followed by free rounds whose scores combine difficulty and judged performance. This retrospective study examined 28 Olympic finals from 2000 to 2024: 224 pair performances and 1,200 round scores. The primary measure compared champion–runner-up relative margins between scoring blocks. Restricted-minus-free differences averaged 3.01 percentage points in platform (95% bootstrap CI 0.62 to 5.41) and 0.38 in springboard (−2.96 to 3.91). The pooled estimate was 1.69 (−0.46 to 3.84). A direct platform–springboard contrast, matched by Games and sex and summarised across seven Games, remained uncertain: 2.63 percentage points (95% t interval −1.54 to 6.79). Champions outscored runners-up on restricted rounds in 27 finals, with a mean counted-award advantage of 0.62 (95% CI 0.47 to 0.78). No champion's restricted dive fell below the descriptive 8.0 benchmark, compared with nine runner-up dives; these were not validated technical-error classifications. Perfect restricted scores would still have been insufficient to close the runner-up's final deficit in 16 finals with free scores unchanged. An official-book audit confirmed all 176 Beijing 2008 round scores and recovered DD for its 112 free dives; six further round scores were checked against a national Olympic committee report. The remaining 640 free-dive DD values were not recovered. The four-final DD analysis is reported separately rather than extrapolated across Games. These results provide historical scoring benchmarks, not evidence that restricted dives independently cause or predict Olympic success. Data, source-specific verification records and executable code accompany the manuscript.

Keywords: Synchronised Diving, Performance Analysis, Judged Performance, Olympic Games, Degree Of Difficulty, Winning Margins, Reproducibility

Introduction

Judged sports convert observed movement into numerical outcomes through rule-governed human evaluation, which makes their results relevant not only to performance analysts and coaches but also to social scientists concerned with how evaluations are formed and interpreted. Identifying where competition margins arise is a central task of performance analysis, but a score difference is not automatically an explanation of performance. Indicators

must be related to the structure of the sport and interpreted at the level their measurements support (Hughes & Bartlett, 2002; O'Donoghue, 2010). In diving, the total reflects both the difficulty assigned to the selected programme and the marks awarded to its performance. A pair can therefore gain points through more demanding dives, better judged performance, or a combination of the two. The same final margin can be assembled in different ways across a competition. Understanding that composition can inform descriptive benchmarking without establishing which training intervention would improve an outcome. Separating description, explanation and prediction also requires different analytical aims and validation designs (Shmueli, 2010). This distinction is particularly important when the athletes being compared are selected because they finished first and second. Their final positions are already consequences of the scores that the analyst seeks to explain.

These general principles take a specific form in Olympic synchronised diving, where the scoring rules themselves divide each final into two differently structured blocks. The event provides a useful within-competition contrast because its first two rounds have an assigned DD of 2.0. Later rounds are not subject to that fixed multiplier, although they remain governed by programme requirements. Women performed three free dives throughout the period examined here; men performed three in 2000 and 2004 and four from 2008. The Olympic judging panel changed from nine judges through Beijing 2008 to eleven from London 2012, but five awards remained in the final calculation after the prescribed exclusions (Beijing Organising Committee, 2008; London Organising Committee, 2012; Fédération Internationale de Natation [FINA], 2022). Their sum is multiplied by 0.6 and then by DD. The restricted block consequently holds the scoring multiplier constant between pairs. It does not make every selected dive biomechanically identical, remove judging influences, or separate execution from synchronisation. Its value is therefore as a comparatively standardised judged-performance measure rather than a direct measure of technical quality.

Previous diving research offers complementary approaches to interpreting scores. Emerson et al. (2009) examined nationality-related judging patterns at the 2000 Olympics, and Kramer (2017) analysed sequential effects in Olympic synchronised-diving scores. These studies show why recorded awards should not be treated as unmediated observations of movement quality. Luedeker and McGee (2022) linked judges' scores to dive attributes measured from video at a regional high-school one-metre springboard competition. Their setting differs substantially from Olympic synchronised diving, but the methodological distinction is relevant: technical attributes were measured rather than inferred solely from points. McGee (2025) also examined discrepancies between estimated diver competency and awarded scores using United States high-school one-metre springboard data. The present study addresses a different, narrower question: how restricted and free scoring blocks contribute to observed margins within Olympic finals. It does not attempt to estimate judging bias or validate a biomechanical performance measure.

Judging research also supplies a framework for understanding what aggregate awards cannot reveal. Social-cognitive accounts treat sports judgements as evaluations shaped by information processing rather than direct readouts of movement quality (Plessner & Haar, 2006; Plessner et al., 2023). In figure skating, experimental work has examined reputation-related differences in assessment (Findlay & Ste-Marie, 2004). Sequential comparison can also produce assimilation or contrast, so an identical scoring scale need not imply an invariant

judging context (Damisch et al., 2006). Serial-position effects have been investigated separately in figure-skating judgements (Bruine de Bruin, 2006). These studies concern different tasks, samples and competitive settings and cannot be transferred numerically to Olympic synchronised diving. Their relevance is to the interpretation of the outcome: a retained-award mean is a judged-performance variable, not a biomechanical measurement. The present analysis consequently avoids attributing every score difference to an underlying technical difference.

Other studies distinguish several potential sources of contextual variation that should not be collapsed into one generic notion of bias. Research on international gymnastics has considered judges' accuracy and nationality-related scoring (Heiniger & Mercier, 2021). Work in figure skating has examined reluctance to produce outlying evaluations (Lee, 2008). A difficulty-related judging pattern has also been studied in gymnastics, which is conceptually different from the explicit DD multiplier in the diving formula (Morgan & Rotthoff, 2014). Historical Olympic research has compared home advantage across different types of events (Balmer et al., 2003). None of these findings establishes the presence, direction or size of such an effect in the present finals. Estimating those effects here would require judge identities, individual awards and relevant performance covariates. Fixing restricted DD improves scoring comparability, but does not supply those missing measurements.

Synchronised performance also involves coordinating two athletes, rather than simply combining their individual scores. Joint-action research describes the perceptual and motor demands of coordinating with another performer (Sebanz et al., 2006). In a different setting, Wing and Woodburn (1995) examined coordination and consistency through measurements of rowing, illustrating the value of direct timing and force information rather than a single outcome score. That rowing study is not evidence about Olympic diving and is used only as a measurement analogy. Pressure-related performance disruption provides another possible influence on skilled action (Beilock & Carr, 2001; Hill et al., 2010). However, a low-scoring dive does not diagnose pressure, failed coordination or a specific execution fault. Those mechanisms would require separate observations, ideally linked to the judged awards. The score-based questions below are therefore narrower than a general explanation of synchronised sporting excellence.

Taken together, existing diving-performance studies have advanced understanding mainly at the level of the individual award, judge or diver. They have examined nationality-related and sequential patterns in Olympic scores (Emerson et al., 2009; Kramer, 2017), the correspondence between awarded scores and measured dive attributes or estimated diver competency (Luedeker & McGee, 2022; McGee, 2025), and the design of springboard training (Barris et al., 2013, 2014). Performance-analysis research has likewise emphasised that indicators should reflect the structure of the sport (Hughes & Bartlett, 2002). Within the studies reviewed here, however, the structural division of synchronised diving into a fixed-DD restricted block and a free block has not been used to decompose Olympic winning margins. It therefore remains unknown whether restricted rounds contribute proportionally more than free rounds to the margin between champions and runners-up, whether any such pattern differs between platform and springboard, and whether it holds across Olympic editions or depends on particular Games. It is also unclear how much of a final deficit restricted points

alone could recover once free-round margins are taken into account. The present study addresses this gap at the level of the Olympic final rather than the individual dive or judge.

Addressing this gap requires clarity about what different scoring patterns could, and could not, indicate. Three possible patterns motivate the analysis, although the available variables cannot adjudicate their mechanisms. A difficulty-oriented account would expect free programmes to create substantial separation because their scores include variation in DD. A precision-oriented account would expect restricted-round differences to matter proportionally more when the leading pairs are otherwise closely matched. A low-score account would focus on isolated performances that contribute less than a pair's other dives. However, a large free-round margin is compatible with higher difficulty, better judged performance, or both. Similarly, a restricted score below a chosen benchmark does not establish that a major technical error occurred. These accounts are therefore interpretive possibilities rather than three separately identified causal models. The analysis tests their observable scoring implications while retaining the distinction between score patterns and their possible causes.

The study was first prompted by a single case, the women's platform final at Rio 2016. Malaysia finished 9.66 points behind China, of which 6.00 points were conceded in the restricted block (Rio 2016 Organising Committee, 2016). This raised the question of whether restricted rounds contributed disproportionately to title margins across other platform finals. A single final, however, cannot show whether such a pattern is typical, event-specific or exceptional. The historical decomposition therefore advances beyond the Rio example, and beyond existing dive- and judge-level research, in three ways: it shifts the unit of analysis from individual dives or judges to the composition of the winning margin; it extends one case to all 28 Olympic finals from 2000 to 2024, analysing platform and springboard separately and comparing them directly; and it reports each estimate with final-level uncertainty, Games-level sensitivity analyses and a source-verification record. The original investigation examined platform first and subsequently applied the same primary measure to springboard. Because the sequence was retrospective and no prospectively registered protocol is supplied, the cross-event extension is not treated as an independent preregistered replication. Both event types are analysed transparently, including a direct comparison of their estimated block contrasts. The motivating final is retained in the historical census, and leave-one-Games-out analyses assess the influence of each Olympic edition. This avoids allowing one memorable final to stand as evidence for a universal pattern.

Building on this rationale, four questions organise the study. First, is the champion's relative advantage over the runner-up larger on restricted than on free rounds, and how uncertain is the difference between event types? Second, how often do eventual champions lead the restricted block, and how large is their counted-award advantage? Third, how does the composition of margins change across finishing positions? Fourth, how are low-score flags and observed score dispersion distributed among finalists? A further scoring-accounting analysis asks whether the runner-up could mathematically close the final gap through restricted points alone, without changing any free scores. The intended contribution is an auditable historical decomposition and a set of interpretable benchmarks. It is not a prediction model, an estimate of training effectiveness, or a claim that restricted dives independently determine Olympic success. To the author's knowledge, this is the first study

to decompose champion–runner-up margins into restricted and free scoring blocks across the complete census of Olympic synchronised-diving finals from 2000 to 2024, and to report that decomposition with source-specific verification and executable code. Its broader contribution to the social sciences is to show how outcomes built from aggregated human judgements can be separated into interpretable components while keeping description distinct from causal and predictive claims, a principle that applies equally to other judged and assessed performances. For coaches and sport educators, the study supplies transparent historical benchmarks grounded in the official scoring structure.

Methods

Design, sample and analysis status

The study was a retrospective observational analysis of the complete Olympic synchronised-diving final programme from Sydney 2000 to Paris 2024. It included seven men's and seven women's finals in each of 10 m platform and 3 m springboard. All 28 finals contained eight pairs, producing 224 pair performances rather than 224 necessarily distinct partnerships. There were 18 five-round finals and 10 six-round finals, yielding 448 restricted and 752 free round scores. Tokyo is labelled 2020 to match the Olympic edition, although the competition took place in 2021. The final was the primary analytical unit; individual round scores were not treated as independent observations of winning. Games-level sensitivity analyses retained the four finals within an Olympic edition together. The primary block comparison was inherited from the original investigation; the direct event comparison, reconstruction audit, score-ceiling calculation and simpler final-level gradient summaries were added during revision and are exploratory. The reporting distinguishes the analytical population, available measurements and exploratory additions, consistent with the transparency aims of observational-reporting guidance (von Elm et al., 2007); this is not a claim that an epidemiological checklist certifies the study's quality.

Data reconstruction, provenance and checks

Historical round scores were reconstructed from secondary event compilations, with the exact source titles, permanent revisions and pre-correction values preserved in the supplementary provenance register. This reconstruction is not represented as the previously unsupplied original working dataset. Fourteen field amendments affecting 11 pair records were inherited from the earlier manuscript: 12 round scores and two published totals. The present revision added a source-specific audit rather than treating arithmetic agreement as verification. All four Beijing 2008 synchronised finals were independently transcribed from the official detailed-result images, including 176 round scores, 32 final totals, dive codes and assigned DD values (Beijing Organising Committee, 2008). The relevant PDF pages are 67, 70, 73 and 76, each identified by its detailed-results report code in the audit register. All 176 scores and 32 totals agreed with the adopted reconstruction. Seven inherited amended fields were thereby confirmed against official detailed records; no analytic score was changed.

A separate check used the Canadian Olympic Committee's contemporaneous account of its men's platform pair at Paris 2024. The report states all six Canadian round scores, including the corrected fourth-round value of 83.25, and a total of 422.13 (Nichols, 2024). These six values matched the reconstruction, bringing fresh round-indexed comparisons to 182 of 1,200 scores. This NOC narrative is distinguished from an official detailed score sheet. World Aquatics (2024) separately corroborates Mexico's men's springboard dive 407C score of

85.68, but does not print the round number in that table; this value is therefore not counted as an additional round-indexed verification. Of the 14 inherited amended fields, seven were verified from the Beijing book, one from the NOC narrative, one had its dive-score value corroborated, and five remained without fresh source confirmation. Table 1 and the updated correction register identify each status. Detailed-score-sheet coverage remains incomplete for 24 finals, and no independent human double-entry audit is asserted.

Table 1
Adopted amendments and updated source-verification status

Final / pair	Transcribed value → adopted value	Fresh verification status in this revision
2004 P W USA	Round 5: 80.56 → 80.58	Not freshly verified; inherited arithmetic/source amendment retained.
2004 S W MEX	Round 5: 68.80 → 58.80	Not freshly verified; inherited arithmetic/source amendment retained.
2008 S W UKR	Round 3: 63.80 → 63.90	Verified: official Beijing book, PDF page 67.
2012 S M MEX	Round 4: 79.18 → 79.80	Not freshly verified; inherited arithmetic/source amendment retained.
2024 S M MEX	Round 4: 85.69 → 85.68	Value 85.68 corroborated for MEX dive 407C in World Aquatics (2024), PDF page 13; round label not printed.
2012 P W MEX	Round 3: 84.40 → 84.48 Total: 343.24 → 343.32	Both fields remain without fresh confirmation; the earlier manuscript's report-based amendment is retained.
2024 P M CAN	Round 4: 83.52 → 83.25	Verified: Nichols (2024), Canadian Olympic Committee; all six Canadian rounds checked.
2008 S W GER	Round 3: 68.40 → 69.60	Verified: official Beijing book, PDF page 67.
2008 S W AUS	Round 3: 68.40 → 68.04 Round 4: 67.14 → 67.50	Both fields verified: official Beijing book, PDF page 67.
2008 S W GBR	Round 3: 53.70 → 53.76 Round 4: 65.29 → 65.25	Both fields verified: official Beijing book, PDF page 67.
2008 S M AUS	Total: 393.80 → 393.60	Verified: official Beijing book, PDF page 70.

P = platform; S = springboard; M = men; W = women. Fourteen field amendments affect 11 pair records. Seven amended fields were freshly verified from official Beijing detailed sheets and one from the Canadian NOC report. One additional dive-score value was corroborated without an independently printed round label; five fields remain unverified in this revision. No adopted analytic score was changed.

The supplement preserves the initial correction log, the updated field-level verification register, the exact secondary-compilation history and the fresh official transcription. All 224 adopted totals reconcile, but the fresh audit covers only 182 round-indexed scores.

Reproducibility and arithmetic consistency must not be presented as complete primary-record certification.

Internal validation established the expected numbers of finals, pairs and rounds, unique pair–round records, agreement between the corrected round sums and all 224 corrected totals, and agreement between totals and finishing order. All 448 restricted scores were compatible with the required 0.6-point lattice. Under the retained-five-award calculation, this is a necessary arithmetic property when individual awards occur in half-point steps, not proof that every underlying mark is correct. The recovered Beijing scores also satisfied the corresponding lattice using their actual assigned DD values. Actual-DD validation was not claimed for the remaining free scores. All 196 numeric cells in the original manuscript's 28-final gold–silver table reproduced at their published precision. These checks demonstrate internal reproducibility from the adopted data, whereas the source audit establishes the more limited primary-record coverage described above.

Primary measures and their interpretation

For each pair, the restricted total R was the sum of rounds 1 and 2, and the free total F was the sum of the remaining rounds. The gold–silver difference in each block was expressed against the mean points scored by those two pairs in that block. Denoting the gold and silver totals in block b by G_b and S_b, the relative margin was $M_b = 100 \times (G_b - S_b) / [(G_b + S_b) / 2]$. The within-final contrast was $D = M_R - M_F$, in percentage points; positive values indicated a proportionally larger restricted-round advantage. A separate descriptive quantity divided the restricted point margin by the overall winning margin. This share can be negative or exceed 100% when the two block margins have opposite signs, and becomes unstable when the final margin is small. Relative margins avoid that particular denominator problem but remain functions of both pairs' observed points. They are not execution-normalised free-round scores, and they are not assumed to provide a conservative test because either pair may have the more difficult free programme.

On a restricted dive, the five retained awards sum to a value that is multiplied by 0.6 × 2.0. The mean counted award is therefore the dive score divided by 6, while the mean across the two restricted dives is R divided by 12. This refers to the awards retained in the official calculation, not the average of every judge's mark. It combines the counted execution and synchronisation awards rather than isolating either component. Differences in this metric are directly interpretable on the judges' 0–10 scale, subject to the measurement limitations above. The maximum restricted total is 120 points, corresponding to a counted-award mean of 10 on both dives. Neither the historical mean for champions nor the minimum observed champion score was treated as a necessary or sufficient performance standard for winning.

Low-score flags, dispersion and scoring headroom

A restricted-dive low-score flag was defined as a counted-award mean below 8.0, equivalently fewer than 48 points. This is an analytical benchmark, not an official classification of a major error. FINA's individual-award guidance labels 7.0–8.0 as good and 8.5–9.5 as very good; a threshold applied to a retained-award mean cannot be equated directly with a referee's technical-error decision (FINA, 2022). A free-dive score-loss flag identified points below 80% of the mean of the same pair's other free dives. That ratio can reflect programme difficulty as well as judged quality and is not an execution-only indicator. Alternative paired thresholds

were 7.5/75% and 8.5/85%, with strictly less-than comparisons throughout. These unvalidated thresholds were retained for descriptive comparability rather than presented as prospectively specified diagnostic cut-offs. Flag-based exclusion analyses condition on the observed scores and are interpreted as sensitivity descriptions, not estimates of the causal effect of eliminating mistakes.

Observed dispersion was summarised in two distinct ways. The standard deviation of individual restricted-dive counted-award means pooled performances within each finishing-position category. Because that combines different pairs, Games, sexes and event types, it does not identify within-pair technical consistency. A complementary descriptive measure was the absolute difference between the two restricted-dive counted-award means within each pair's final. Reliability requires attention to repeated measurements and the relevant sources of error (Atkinson & Nevill, 1998; Hopkins, 2000). Even this two-dive contrast is not a reliability estimate: the selected dives differ and repeated trials of the same dive are unavailable. For scoring headroom, the increase needed in the runner-up's two-dive mean to match the champion's final total was calculated as the final point gap divided by 12. This was compared separately in each final with 10 minus the runner-up's observed restricted mean. All free scores and the champion's scores were held unchanged, making the calculation an arithmetic ceiling rather than a realistic intervention or forecast.

Primary comparisons, direct event contrast and uncertainty

Restricted and free relative margins were compared within finals using two-sided Wilcoxon signed-rank tests, matched-pairs rank-biserial correlations (Kerby, 2014), and percentile-bootstrap confidence intervals for the mean contrast (Efron, 1979). Twenty thousand resamples of finals were drawn with replacement using a fixed random seed of 20260930. Sign tests separately evaluated the number of positive versus negative contrasts. The signed-rank test uses both signs and ranked magnitudes, whereas the sign test uses direction only; neither should be described simply as the p value for the observed count. Bootstrap mean intervals and signed-rank tests also address different statistical summaries; a location-shift interpretation of the signed-rank test additionally assumes a symmetric contrast distribution. Nominal p values are reported, with a two-test Holm adjustment (Holm, 1979) supplied for the platform and springboard signed-rank comparisons. The pooled analysis and all subsequent extensions are exploratory rather than additional confirmatory tests. No equivalence bound or smallest practically important difference was specified, so a non-significant result is not interpreted as evidence of no meaningful effect (Lakens, 2017; Wasserstein & Lazar, 2016).

The direct event contrast was constructed by matching platform and springboard finals within each Games and sex. Their D values were subtracted to give 14 matched contrasts, and the two sex-specific contrasts were averaged within each of seven Games. The overall mean difference was evaluated using a Student-t interval across those seven Games-level values, with six degrees of freedom, and a Games-resampled percentile interval. An additional paired interval across the 14 Games–sex contrasts is supplied for comparison, not selected according to which is most favourable. This directly addresses between-event uncertainty rather than inferring a difference from one small p value and one large p value (Gelman & Stern, 2006). For the primary pooled and event-specific means, Games-level t intervals, Games-resampled percentile intervals and leave-one-Games-out estimates were also calculated. The Games-

level t analysis assumes the edition means are sufficiently independent and approximately normally distributed; the bootstrap relies on exchangeability of the observed editions (Field & Welsh, 2007). Seven Games remain a small basis for either approach, and repeated competitors and national programmes may introduce dependence across editions.

Secondary descriptions and computational reproducibility

The champion was compared with each lower finishing position, generating seven block contrasts per final. Margins were averaged by finishing position, with bootstrap resampling at the final level rather than treating the 196 comparisons as independent matches. To summarise the gradient parsimoniously, separate ordinary least-squares slopes of restricted and free relative margin against finishing-position distance were calculated within every final and then averaged. Distance ranged from zero for silver to six for eighth; final-resampled and Games-resampled intervals are reported for these descriptive slope summaries. This two-stage approach replaces the original draft's exploratory mixed-model sequence and does not estimate an effect of opponent quality. Additional summaries covered silver–bronze margins, the champion's mean margin over the other seven finalists, within-round coefficients of variation, and the champion's standing after each round. All analyses used the supplied Python script with NumPy, pandas and SciPy; package versions, the random seed, input files and derived outputs are recorded in the reproducibility package (Harris et al., 2020; Virtanen et al., 2020). Figures were generated with Matplotlib (Hunter, 2007). The historical means are descriptive quantities for the enumerated record; inferential intervals concern hypothetical repetition beyond that record and depend on the stated assumptions.

The reproducibility package separates the raw reconstruction, inherited amendments, freshly transcribed official records and derived variables. Executable scripts preserve that sequence and fail when a source comparison or arithmetic validation does not agree. Input files, software versions, the random seed and numerical outputs are recorded, following reproducible-computation principles (Sandve et al., 2013). A source register and data dictionary identify the meaning and provenance of each field, supporting the reuse aims of the FAIR principles without claiming formal FAIR certification (Wilkinson et al., 2016). The original secondary-source history is retained as an audit trail rather than replaced by a misleading all-official provenance statement. Recovered DD is stored in an enriched copy; missing DD is left blank and marked explicitly. The primary all-Games calculations continue to use the same adopted score data, so their values are not changed by the partial DD extension.

Official-DD subset: Beijing 2008

The four Beijing finals provide a complete, source-verified DD subset for one Olympic edition: 32 pair performances, 64 restricted dives and 112 free dives. Their inclusion was determined by access to usable official detailed records, not random selection from the seven Games. For each recovered dive, the mean retained award was calculated as points divided by 3 × DD. The factor of three follows from five retained awards multiplied by 0.6 in the scoring formula. For each free programme, the analysis summarised the sum of assigned DD, the unweighted mean of the individual-dive retained-award means, and the total free points. These are distinct quantities: three times the product of the programme mean award and summed DD need not equal total points, because DD and awards vary together across dives. The gold–silver comparison is presented in Table 6, while all 32 pair summaries are supplied in the supplement. No inferential test or imputation was added for this accessibility-defined four-

final subset. Its purpose is to distinguish observed programme difficulty from DD-adjusted judged awards in the available records, not to establish a general difficulty mechanism across 2000–2024.

Results

Historical overview and primary block comparison

China won 23 of the 28 titles: 12 in platform and 11 in springboard. The remaining titles were won by Russia in the men's platform and women's springboard finals of 2000, Greece in men's springboard in 2004, and Great Britain in men's springboard in 2016 and men's platform in 2020. The mean gold–silver gap was 21.45 points, ranging from 1.23 to 52.98. Restricted rounds supplied a median 30% of the two leading pairs' points, with a range of approximately 22% to 33%. The full final-level decomposition is presented in Table 2, and the distributions of relative margins are shown in Figure 1. Restricted and free margins varied appreciably between finals within both event types. This variation is important because the averages do not describe a uniform pattern in every Olympic final.

In platform, relative margins averaged 8.33% on restricted rounds and 5.32% on free rounds, giving a mean difference of 3.01 percentage points (95% final-bootstrap CI 0.62 to 5.41; Table 3). The difference was positive in 11 of 14 finals; the signed-rank p value was 0.025 and the rank-biserial correlation was 0.68. The corresponding two-sided sign-test p value was 0.057. In springboard, restricted and free relative margins averaged 5.78% and 5.40%, giving a mean difference of 0.38 percentage points (−2.96 to 3.91). Seven of 14 differences were positive, with signed-rank and sign-test p values of 1.000. The two-test Holm-adjusted signed-rank p values were 0.049 for platform and 1.000 for springboard. The platform estimate was therefore more clearly positive in this record, but its strength should not be expressed as a categorical contrast with the springboard result.

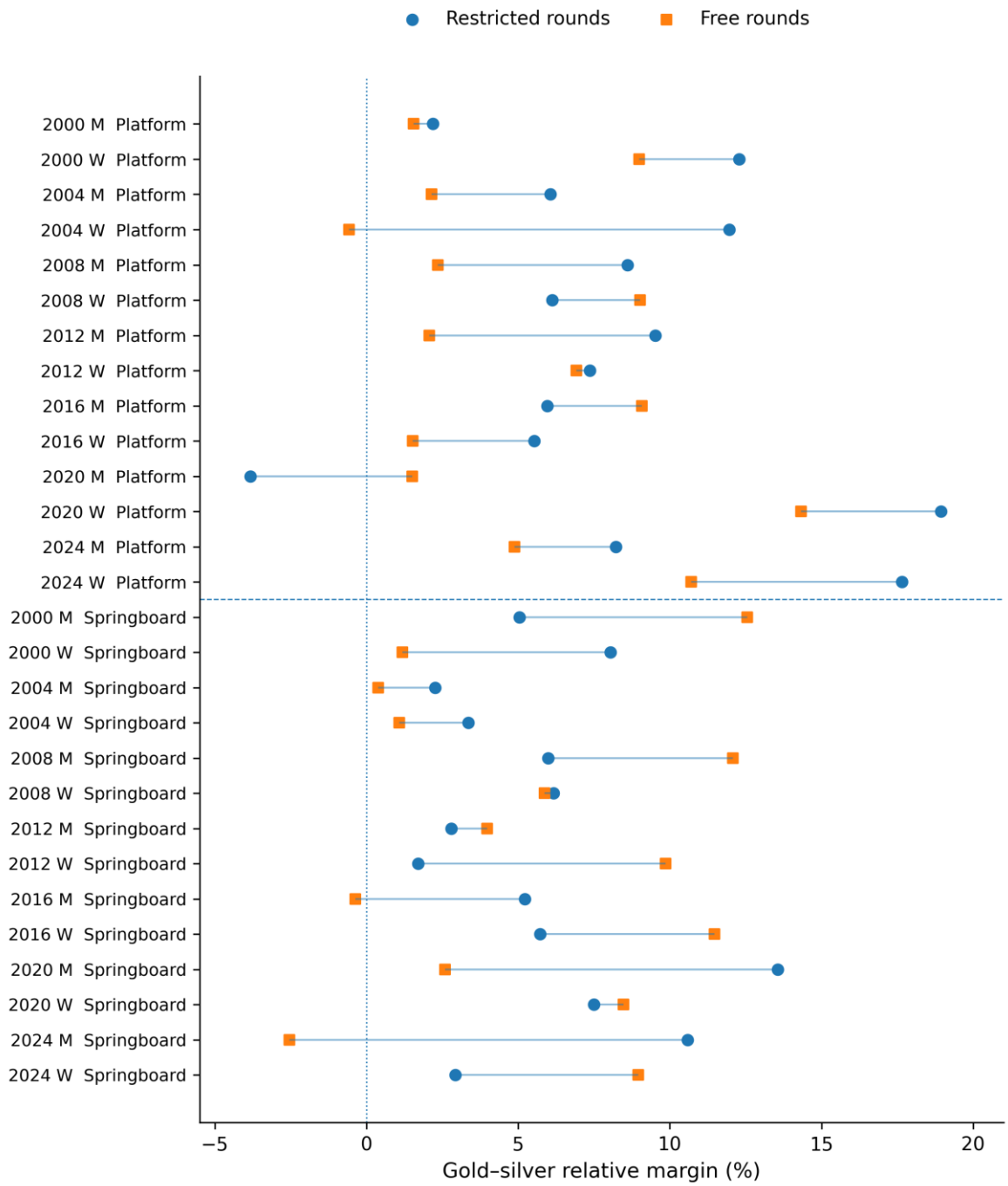


Figure 1 Gold–silver relative margins on restricted and free rounds in each Olympic final
Note. Each line links the two block margins from one final. Positive values favour the eventual champion. Restricted rounds use a fixed DD of 2.0; free scores retain variation in difficulty. M = men; W = women. The figure is descriptive and conditions on the eventual finishing order.

Table 2
Gold–silver score decomposition by final: 10 m platform

Final	Gold–silver	Final margin	Restricted margin	Free margin	Counted award gold / silver	Relative margin restricted / free (%)
2000 M	RUS–CHN	6.30	2.40	3.90	9.25 / 9.05	2.19 / 1.55
2000 W	CHN–CAN	33.09	13.20	19.89	9.50 / 8.40	12.29 / 8.99
2004 M	CHN–GBR	12.36	6.60	5.76	9.35 / 8.80	6.06 / 2.14
2004 W	CHN–RUS	11.22	12.60	−1.38	9.30 / 8.25	11.97 / −0.57
2008 M	CHN–GER	17.76	9.60	8.16	9.70 / 8.90	8.60 / 2.35
2008 W	CHN–AUS	28.38	6.60	21.78	9.25 / 8.70	6.13 / 9.01
2012 M	CHN–MEX	17.88	10.20	7.68	9.35 / 8.50	9.52 / 2.07
2012 W	CHN–MEX	25.08	7.80	17.28	9.15 / 8.50	7.37 / 6.91
2016 M	CHN–USA	39.87	6.60	33.27	9.50 / 8.95	5.96 / 9.08
2016 W	CHN–MAS	9.66	6.00	3.66	9.30 / 8.80	5.52 / 1.52
2020 M	GBR–CHN	1.23	−4.20	5.43	8.95 / 9.30	−3.84 / 1.50
2020 W	CHN–USA	52.98	19.20	33.78	9.25 / 7.65	18.93 / 14.32
2024 M	CHN–GBR	26.91	9.00	17.91	9.50 / 8.75	8.22 / 4.87
2024 W	CHN–PRK	43.20	18.00	25.20	9.25 / 7.75	17.65 / 10.70

Note. Margins are in points. M = men; W = women. Restricted rounds are rounds 1–2 at an assigned DD of 2.0. Counted-award mean = restricted total ÷ 12. Relative margin = 100 × the block margin ÷ the two pairs' mean points in that block. Negative block margins indicate that the eventual runner-up scored more points in that block. Values are recalculated from the explicitly labelled reconstruction.

Country codes used in this table: AUS Australia; CAN Canada; CHN China; GBR Great Britain; GER Germany; GRE Greece; ITA Italy; MAS Malaysia; MEX Mexico; PRK DPR Korea; RUS Russia; USA United States. Tokyo is labelled by its 2020 Olympic edition.

Table 2 (continued)
Gold–silver score decomposition by final: 3 m springboard

Final	Gold–silver	Final margin	Restricted margin	Free margin	Counted award gold / silver	Relative margin restricted / free (%)
2000 M	CHN–RUS	35.61	5.40	30.21	9.15 / 8.70	5.04 / 12.55
2000 W	RUS–CHN	11.04	8.40	2.64	9.05 / 8.35	8.05 / 1.19
2004 M	GRE–GER	3.33	2.40	0.93	8.95 / 8.75	2.26 / 0.38
2004 W	CHN–RUS	6.06	3.60	2.46	9.10 / 8.80	3.35 / 1.09
2008 M	CHN–RUS	47.10	6.60	40.50	9.45 / 8.90	5.99 / 12.07
2008 W	CHN–RUS	19.89	6.60	13.29	9.20 / 8.65	6.16 / 5.87

Final	Gold–silver	Final margin	Restricted margin	Free margin	Counted award gold / silver	Relative margin restricted / free (%)
2012 M	CHN–RUS	17.37	3.00	14.37	9.05 / 8.80	2.80 / 3.98
2012 W	CHN–USA	24.30	1.80	22.50	8.90 / 8.75	1.70 / 9.86
2016 M	GBR–USA	4.11	5.40	−1.29	8.85 / 8.40	5.22 / −0.37
2016 W	CHN–ITA	31.77	6.00	25.77	9.00 / 8.50	5.71 / 11.47
2020 M	CHN–USA	23.46	14.40	9.06	9.45 / 8.25	13.56 / 2.59
2020 W	CHN–CAN	25.62	7.20	18.42	8.30 / 7.70	7.50 / 8.47
2024 M	CHN–MEX	2.07	10.80	−8.73	8.95 / 8.05	10.59 / −2.54
2024 W	CHN–USA	23.04	3.00	20.04	8.65 / 8.40	2.93 / 8.95

Note. Margins are in points. M = men; W = women. Restricted rounds are rounds 1–2 at an assigned DD of 2.0. Counted-award mean = restricted total ÷ 12. Relative margin = 100 × the block margin ÷ the two pairs' mean points in that block. Negative block margins indicate that the eventual runner-up scored more points in that block. Values are recalculated from the explicitly labelled reconstruction.

Country codes used in this table: AUS Australia; CAN Canada; CHN China; GBR Great Britain; GER Germany; GRE Greece; ITA Italy; MAS Malaysia; MEX Mexico; PRK DPR Korea; RUS Russia; USA United States. Tokyo is labelled by its 2020 Olympic edition.

Across all 28 finals, the mean restricted-minus-free difference was 1.69 percentage points (95% final-bootstrap CI −0.46 to 3.84). Restricted margins were proportionally larger in 18 finals, with a signed-rank p value of 0.186, a sign-test p value of 0.185 and a rank-biserial correlation of 0.29. The direct platform-minus-springboard difference was 2.63 percentage points, with a Games-level t interval of −1.54 to 6.79 and p = 0.173. Its Games-bootstrap interval was −0.59 to 5.53; the paired t interval across the 14 Games–sex contrasts was −2.77 to 8.02. These direct intervals do not establish a between-event difference. They also do not establish equivalence between event types. The evidence supports reporting distinct estimates and their uncertainty rather than concluding that platform and springboard operate according to demonstrably different mechanisms.

Table 3
Primary block comparisons and uncertainty summaries

Measure	Platform (n = 14)	Springboard (n = 14)	Pooled (n = 28)
Finals, n	14	14	28
Restricted relative margin, mean (%)	8.33	5.78	7.05
Free relative margin, mean (%)	5.32	5.40	5.36
Restricted minus free, mean (percentage points)	3.01	0.38	1.69
95% final-bootstrap CI	0.62 to 5.41	−2.96 to 3.91	−0.46 to 3.84

Measure	Platform (n = 14)	Springboard (n = 14)	Pooled (n = 28)
Positive within-final contrasts	11/14	7/14	18/28
Wilcoxon W; two-sided p	17; 0.025	52; 1.000	144; 0.186
Matched-pairs rank-biserial correlation	0.68	0.01	0.29
Two-sided sign-test p	0.057	1.000	0.185
Holm-adjusted Wilcoxon p (two event tests)	0.049	1.000	Not in family
95% Games-level t CI (7 Games; df = 6)	0.25 to 5.77	−2.81 to 3.57	−0.44 to 3.83
95% Games-bootstrap CI	1.10 to 5.17	−2.00 to 2.70	0.20 to 3.33
Restricted counted award, gold / silver	9.33 / 8.59	9.00 / 8.50	9.17 / 8.55
Gold–silver award gap (95% final-bootstrap CI)	0.74 (0.48 to 0.99)	0.50 (0.37 to 0.66)	0.62 (0.47 to 0.78)

Direct event contrast. Platform minus springboard, matched by Games and sex: mean 2.63 percentage points; Games-level t CI −1.54 to 6.79, p = 0.173; Games-bootstrap CI −0.59 to 5.53. The 14-pair t CI is −2.77 to 8.02. This comparison does not establish a between-event difference.

Note. Percentile intervals use 20,000 resamples and seed 20260930. Final resampling and Games resampling use different assumptions; the pooled Games-bootstrap interval excludes zero, whereas the corresponding t interval includes it. Signed-rank tests and mean intervals summarise different features of the contrast distribution. Holm adjustment covers only the two event-specific signed-rank tests; remaining analyses are exploratory.

Restricted-round benchmarks and standings

The champion outscored the runner-up on restricted rounds in 27 of 28 finals: 13 of 14 platform finals and all 14 springboard finals. Champions averaged 9.17 per counted award compared with 8.55 for runners-up, a mean difference of 0.62 (95% final-bootstrap CI 0.47 to 0.78). The corresponding event-specific gaps were 0.74 in platform and 0.50 in springboard. Champions led the entire field outright after round 2 in 23 finals and held the highest or joint-highest restricted total in 24. The outright leader after rounds 1, 2, 3 and 4 eventually won in 22, 23, 24 and 26 finals, respectively. These are retrospective standings summaries, not independent forecasts or tests against a 50:50 chance of victory. The exception to the gold–silver restricted advantage was the 2020 men's platform final, in which Great Britain overcame a 4.20-point restricted deficit and won by 1.23 points.

Scoring headroom: what restricted points alone could change

The runner-up won the free block in three finals: women's platform in 2004 and men's springboard in 2016 and 2024. In these cases the champion's restricted margin exceeded the final winning margin because it offset a free-round deficit. In the other 25 finals, the champion also won the free block, so merely matching the champion's restricted total would still have left the runner-up behind. To match the champion's full total solely by raising restricted scores required a median increase of 1.96 counted-award points among those 25 runners-up. The

corresponding median over all 28 finals was 1.79; these are different denominators. Individual score-ceiling checks showed that even 120 restricted points would have left the runner-up below the champion's total in 16 of 28 finals (Figure 2). In the remaining 12, the ceiling was high enough arithmetically, including nine of the 25 finals in which the champion won both blocks. These calculations establish only scoring feasibility while holding all other scores fixed; they do not imply that perfect awards were realistically attainable.

Margin composition across finishing positions

The champion's restricted-minus-free relative margin declined from +1.69 percentage points against silver to −1.46 against bronze and approximately −16.0 against eighth place (Table 4; Figure 3). The within-final rank correlation between finishing-position distance and this contrast was negative in 27 of 28 finals, with a mean correlation of −0.68. Averaged within-final slopes were +0.89 percentage points per finishing place for restricted margins and +3.55 for free margins. Their difference was −2.66 (95% final-bootstrap CI −3.21 to −2.13; Games-bootstrap CI −3.45 to −1.97). In absolute terms, the champion's restricted advantage increased from 0.62 counted-award points against silver to 1.11 against eighth, while the free-point advantage increased from approximately 14.0 to 71.2. The within-round coefficient of variation averaged 5.9% for restricted rounds and 11.9% for free rounds, with the restricted value lower in every final. These gradients describe the composition of score-defined rankings, not an independent effect of lower opponent quality.

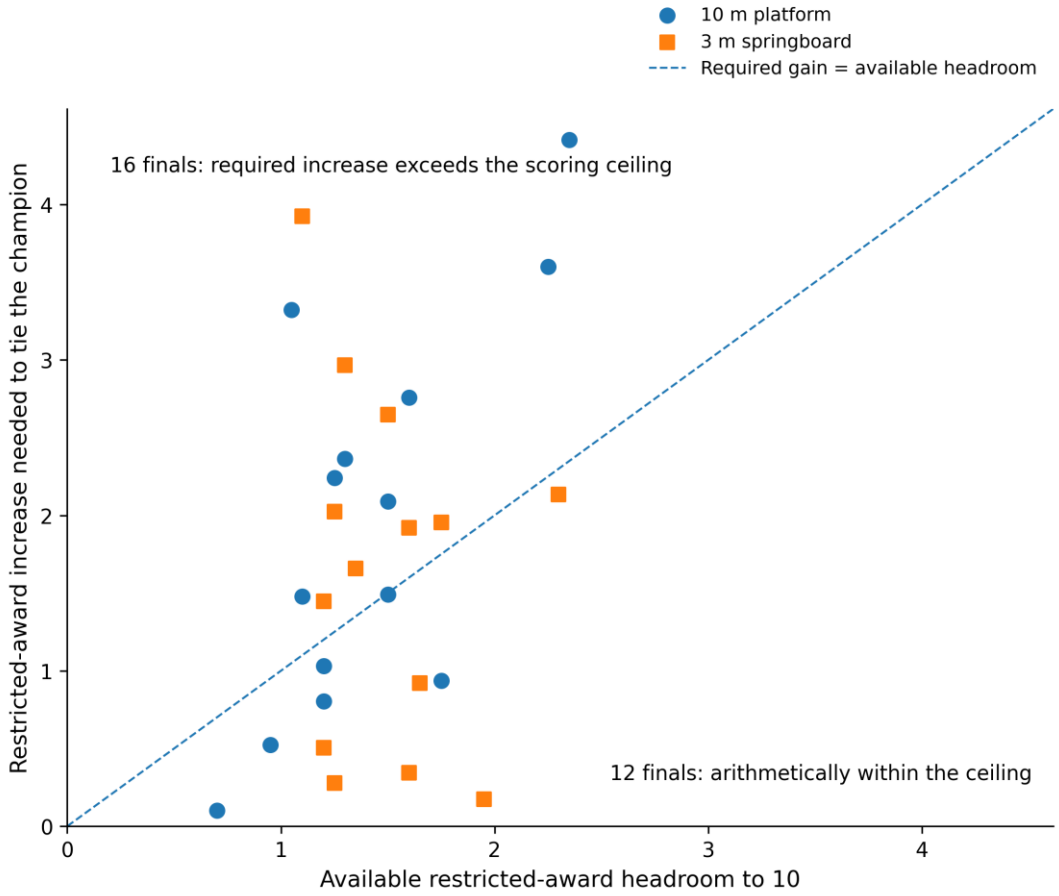


Figure 2 Restricted-score ceiling and the increase needed for the runner-up to tie the champion

Note. Each point is one final. The vertical coordinate is the final point gap ÷ 12. The horizontal coordinate is 10 minus the runner-up's restricted counted-award mean. Points above the diagonal require more improvement than the scoring ceiling permits (16 finals); those below it are arithmetically feasible (12 finals). All free scores and the champion's scores are held fixed. Feasibility is not evidence that the improvement was realistically achievable.

Table 4
Champion's margins across finishing positions and descriptive within-final slopes

Compared position	Restricted margin (%)	Free margin (%)	Pooled difference (95% CI), pp	Platform difference, pp	Springboard difference, pp
Silver	7.05	5.36	1.69 (−0.46 to 3.84)	3.01	0.38
Bronze	8.38	9.84	−1.46 (−3.43 to 0.56)	−0.94	−1.97
4	9.18	13.42	−4.23 (−6.19 to −2.28)	−4.10	−4.37
5	9.49	15.59	−6.10 (−8.22 to −3.97)	−4.83	−7.37
6	9.93	18.14	−8.21 (−10.50 to −6.03)	−6.45	−9.97
7	11.45	21.57	−10.13 (−13.57 to −6.47)	−6.74	−13.52
8	13.03	29.07	−16.04 (−20.09 to −12.08)	−14.03	−18.05

Note. Relative margins compare the champion with the stated position. Differences are restricted minus free; pp = percentage points. Intervals resample whole finals. Position is determined by these scores, so the table is a descriptive decomposition rather than a test of opponent quality.

Within-final slope summaries

Margin	Mean slope	95% final-bootstrap CI	95% Games-bootstrap CI
Restricted margin	0.89	0.52 to 1.25	0.48 to 1.28
Free margin	3.55	3.08 to 4.04	2.80 to 4.43
Restricted minus free	−2.66	−3.21 to −2.13	−3.45 to −1.97

Slope units are percentage points per finishing place below silver. A separate least-squares slope was calculated within each final and the 28 slopes were averaged. This is not a mixed-model coefficient and does not treat 196 champion–opponent comparisons as independent observations.

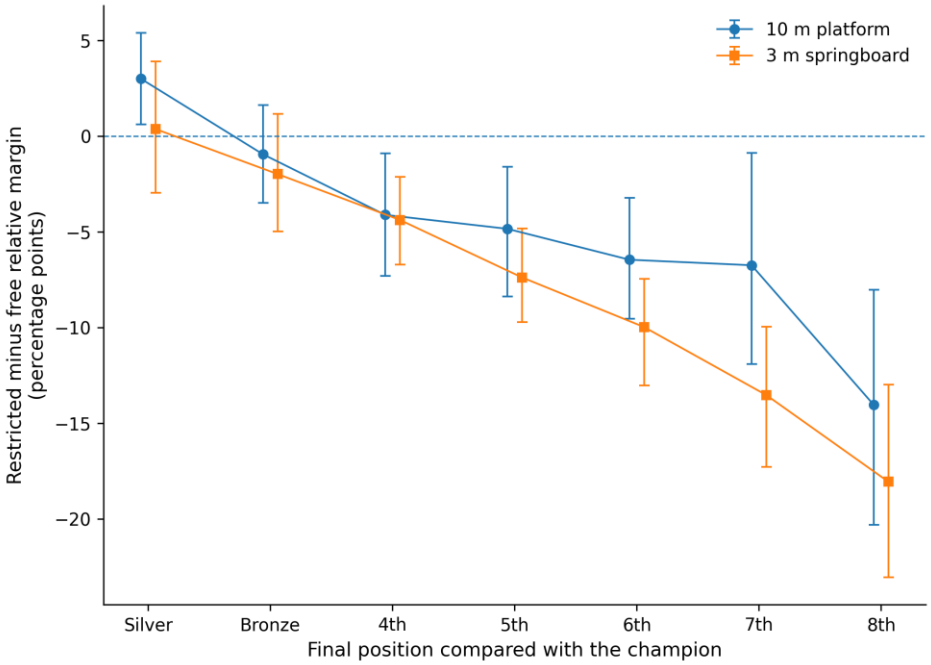


Figure 3 Restricted-minus-free relative margins across lower finishing positions

Note. Points are means across 14 finals per event and error bars are 95% percentile intervals from 20,000 resamples of whole finals. Positive values indicate a proportionally larger restricted-round advantage. No fitted mixed-model line is shown. Final positions are generated by the same competition scores, so this is a descriptive margin-composition gradient.

The same distinction was evident in the supplementary comparisons. Averaging the champion's seven pairwise relative margins within each final gave 9.79% on restricted rounds and 16.14% on free rounds. The restricted margin exceeded the free margin in only two finals under that comparison. Silver and bronze medallists were separated by mean relative margins of 1.33% on restricted rounds and 4.49% on free rounds. Thus, restricted-round advantages were particularly characteristic of champions relative to runners-up, but were not the main source of separation throughout the full field. That pattern cannot distinguish programme difficulty from judged free-dive quality. The supplementary outputs retain the corresponding exploratory tests without treating them as new independent confirmations of the primary hypothesis.

Low-score flags and observed dispersion

No champion recorded a restricted counted-award mean below 8.0 in 56 restricted dives, and none met the free-dive 80% score-loss criterion in 94 free dives (Table 5). Runners-up recorded nine restricted flags and two free flags, affecting eight of 28 pairs; bronze medallists recorded nine and two, affecting ten pairs. Among fourth- to eighth-placed pairs, 70 restricted and 20 free dives were flagged, affecting 60 of 140 pairs. The lowest champion restricted award was 8.30, while the lowest champion free-score ratio was 83.8%, not a validated minimum performance requirement. At the stricter 7.5/75% thresholds, neither gold nor silver pairs were flagged. At the more permissive 8.5/85% thresholds, four champions and 17 runners-up were flagged. The apparent separation between medallists therefore depended substantially on the chosen descriptive thresholds.

Table 5

Descriptive low-score flags and restricted-award dispersion

Position	Restricted dives, n	Restricted flags, n (%)	Free dives, n	Free flags, n (%)	Pairs flagged, n/N (%)
Gold	56	0 (0%)	94	0 (0%)	0/28 (0%)
Silver	56	9 (16%)	94	2 (2%)	8/28 (29%)
Bronze	56	9 (16%)	94	2 (2%)	10/28 (36%)
4th–8th	280	70 (25%)	470	20 (4%)	60/140 (43%)

A restricted flag is a counted-award mean < 8.0, not a validated technical-error classification. A free flag is a score < 80% of the mean of the same pair's other free dives; it retains difficulty effects. "Pairs flagged" counts performances with either type of flag, not the sum of flag counts.

Restricted-award dispersion

Position	Mean counted award	Pooled SD of individual dive means	Mean absolute gap between the pair's two dive means
Gold	9.17	0.36	0.34
Silver	8.55	0.45	0.32
Bronze	8.44	0.51	0.34
4th–8th	8.25	0.55	0.34

The pooled standard deviation mixes performances by different pairs, Games, sexes and event types. The two-dive gap is a different, within-final descriptive quantity; it is not a same-dive test–retest reliability measure. Champions had a narrower pooled score distribution but not a smaller average two-dive gap than runners-up.

Threshold sensitivity: number of gold / silver pairs flagged

Restricted / free cut-offs	Gold pairs flagged	Silver pairs flagged
< 7.5 / < 75%	0/28	0/28
< 8.0 / < 80%	0/28	8/28
< 8.5 / < 85%	4/28	17/28

Both thresholds change together in these sensitivity descriptions. Strictly less-than cut-offs are used. The exercise assesses dependence on the chosen benchmarks; it does not validate them.

The pooled standard deviation of individual restricted-dive awards was 0.36 among champions, compared with 0.45 for runners-up, 0.51 for bronze medallists and 0.55 for the remaining finalists. However, the mean absolute difference between a pair's two restricted-dive awards was 0.34 for champions and 0.32 for runners-up, with values of approximately 0.34 in both other categories. Champions therefore did not show a smaller average two-dive score contrast on this simple within-final measure. These results distinguish the narrower

observed award distribution among champions from a claim of superior within-pair technical consistency. Excluding the eight finals containing a gold or silver flag reduced the pooled block difference to 0.35 percentage points, with restricted and free relative margins of 5.69% and 5.34% (signed-rank p = 0.812). Six excluded finals were platform finals; among the eight remaining platform finals, the mean block difference was 1.73 percentage points and p = 0.148. Because selection was based on the component scores, this attenuation is descriptive and does not estimate what removing technical errors would cause.

Sensitivity to Games and analytical summary

The platform mean contrast remained positive after removal of each Games in turn, ranging from 2.14 to 3.57 percentage points. Its signed-rank p values ranged from 0.009 to 0.077, indicating that a binary significance label was sensitive to the inclusion of individual editions. Games-level t intervals were 0.25 to 5.77 for platform, −2.81 to 3.57 for springboard and −0.44 to 3.83 for the pooled estimate. The pooled Games-percentile-bootstrap interval, in contrast, was 0.20 to 3.33. With only seven Games, this disagreement cautions against selecting a single interval construction as decisive evidence of a general effect. The round-3 comparison retained a larger mean platform restricted margin, but cannot separate block from order because restricted dives always came first. The exploratory association between final closeness and the restricted-minus-free contrast was strong in springboard (Spearman ρ = −0.80) and weak in platform (ρ = 0.06). Both quantities reuse the same score components, so these associations are not evidence that closeness independently changes the value of restricted execution.

Difficulty and DD-adjusted awards in the verified 2008 subset

All 112 Beijing free dives had an explicitly recorded DD, ranging from 2.8 to 3.8. The gold–silver programme comparisons varied even within this single edition (Table 6). In men's platform, China's summed free DD was lower than Germany's, 13.8 versus 14.2, while its mean retained award was higher, 8.53 versus 8.08; China nevertheless held an 8.16-point free-round advantage. In women's platform, both leading pairs had a DD sum of 9.6, but China's mean retained award was 8.77 compared with Australia's 8.00, with a 21.78-point free gap. China's summed DD was slightly higher than the runner-up's in both springboard finals: 13.4 versus 13.3 for men and 9.0 versus 8.7 for women. Its mean retained awards were also higher in those finals, at 8.85 versus 7.93 and 8.63 versus 8.43, respectively. Thus, a gold–silver free-score advantage was not uniformly associated with a higher programme DD, even in the fully documented subset. This observation demonstrates why difficulty should not be inferred from total points alone. It does not quantify a causal contribution of difficulty or show that the 2008 pattern represents the other Games.

Table 6
Beijing 2008: source-verified free-programme difficulty and retained awards

Final	Gold–silver	Free dives per pair	Sum of DD gold / silver	Mean retained award gold / silver	Free margin (points)
Platform Men	CHN–GER	4	13.8 / 14.2	8.53 / 8.08	8.16
Platform Women	CHN–AUS	3	9.6 / 9.6	8.77 / 8.00	21.78
Springboard Men	CHN–RUS	4	13.4 / 13.3	8.85 / 7.93	40.50
Springboard Women	CHN–RUS	3	9.0 / 8.7	8.63 / 8.43	13.29

Note. Source: Beijing Organising Committee (2008), detailed results on PDF pages 67, 70, 73 and 76. The subset contains all eight pairs in each of four finals; this table presents the gold–silver comparisons. Full pair-level summaries are included in the supplement.

For a free dive, mean retained award = dive points ÷ (3 × assigned DD). The programme mean is the unweighted average across that pair's free dives; it is not total free points divided by three times the DD sum. The latter would be a DD-weighted award mean. Both programme difficulty and the awarded marks affect total points.

These four finals were selected by availability of usable official detailed records, not randomly. They supply 112 of the 752 free-dive DD values in the all-Games dataset. No inference about a general Olympic difficulty mechanism is drawn from this subset.

Discussion

A benchmark, not a general rule about where titles are decided

The study identified a repeated restricted-round advantage among Olympic champions, but did not establish a consistent disproportionate restricted-round advantage across both event types. This is a more specific conclusion than either saying that restricted dives decide titles or saying that they do not matter. Absolute contribution, disproportionate contribution and causal importance are different questions. Restricted rounds usually supplied around 30% of the leading pairs' points, yet could supply a larger or smaller share of their margin. A champion could win the restricted block without that block generating most of the final gap. Conversely, a small restricted advantage could offset a negative free-block margin in a close final. The useful contribution is therefore the scoring decomposition itself, together with transparent limits on the inferences drawn from it.

The 27-of-28 result is descriptively striking but conditioned on finishing position. High restricted scores contribute to a high total, so eventual champions are selected partly for success in the opening block. The same issue applies to the absence of low-score flags among champions and the frequency with which early leaders won. These observations do not demonstrate independent predictive value, necessity for gold, or a causal benefit from a particular preparation strategy. A prospective prediction study would need forecasts made before the relevant outcome and evaluated on separate competitions, with an appropriate comparison model. A causal training study would require measurements before and after an intervention and a design that addresses alternative explanations. The distinction between

explaining an outcome and predicting a new one is methodological, not simply a choice of statistical model (Shmueli, 2010). Those are different research tasks from reconstructing how recorded Olympic margins were assembled.

Between-event uncertainty and the limits of explanation

Platform's mean contrast was larger than springboard's, but the direct event comparison remained imprecise. This matters because comparing significance labels would exaggerate the distinction between the events (Gelman & Stern, 2006). The seven Games-level contrasts included both positive and negative values, and the intervals allowed differences of either sign. The platform result was also attenuated when finals containing low-score flags were omitted, although that exclusion is itself based on observed performance. It is reasonable to ask whether programme selection, synchronisation demands or the distribution of lower-scoring performances differs between platform and springboard. It is not possible to choose among those explanations from the present variables. The cross-event extension consequently broadens the descriptive evidence without demonstrating separate causal mechanisms for the two disciplines.

Incomplete free-dive DD remains the central obstacle to a whole-period explanatory analysis, but the recovered Beijing subset provides a limited advance. Its 112 free dives distinguish recorded programme difficulty from DD-adjusted retained awards for all 32 pairs in that edition. China's men's platform programme had a lower summed DD than Germany's while producing more free points, whereas the leading women's platform programmes had equal DD sums. These examples directly illustrate why a larger free score cannot be assumed to indicate a more difficult programme. They do not establish the relative causal effects of difficulty and execution, because selected dives, judging context and athlete capability also differ. Nor can one accessible Games establish whether elite free programmes generally converge in difficulty. Recovering the remaining 640 free-dive DD values is necessary before extending the same description across the historical census. Judge-level execution and synchronisation awards and independent video measures would still be required to connect those score components to particular technical characteristics (Luedeker & McGee, 2022).

Low scores are not synonymous with technical errors or inconsistency

Renaming the restricted threshold as a low-score flag is substantively important. An award below 8.0 may be competitively costly without corresponding to a major fault under the official judging criteria. Moreover, FINA's qualitative bands apply to individual awards, whereas the study's metric averages the five retained execution and synchronisation awards (FINA, 2022). Threshold sensitivity showed that apparently clear champion–runner-up separation changed when the cut-offs changed. A video-based coding scheme, documented referee decision or independently rated technical criterion would be needed to validate an error classification. Free-round flags have an additional ambiguity because raw points retain DD. These indicators are therefore useful for identifying performances for further review, not for diagnosing why a dive scored poorly.

The dispersion findings provide a second measurement lesson. Champions' restricted awards occupied a narrower pooled distribution, but their average difference between the two opening dives was not smaller than that of runners-up. The former statistic mixes between-pair and within-pair sources of variation, whereas the latter captures only a limited contrast

between two different selected dives. Neither should be presented as a complete measure of coordination reliability or pressure resistance. Research on elite performance variability shows why distinguishing the type and scale of variation matters (Atkinson & Nevill, 1998; Hopkins, 2000; Malcata & Hopkins, 2014). Synchronisation also involves the coordination of two performers, which cannot be recovered from a single aggregate points total (Sebanz et al., 2006). The revised interpretation therefore avoids attributing the champions' pooled score pattern to superior technical consistency without the repeated measurements needed to support that mechanism.

Practical interpretation and the Rio example

The counted-award transformation remains a useful practical output: dividing the restricted total by 12 expresses a pair's opening-block performance on an interpretable judging scale. Historical champion means of 9.33 in platform and 9.00 in springboard provide context, not selection thresholds or guaranteed targets. Coaches can compare a pair's repeated competition observations with such distributions while considering event, era, judging panel and programme. A low-score flag can identify a dive for video review, but the video and judge-level marks should determine what needs attention. Competition-representative practice is a reasonable coaching consideration supported by separate diving research, rather than an intervention whose effectiveness was tested here (Barris et al., 2013, 2014; Pinder et al., 2011). The present data do not identify the optimal balance of restricted- and free-dive training. That balance requires the pair's technical profile, programme difficulty, consistency across repeated attempts and response to training.

The individual ceiling calculation makes the same caution concrete. In 16 finals, improving restricted points alone could not mathematically erase the runner-up's deficit even at the maximum score. In the other 12, arithmetic feasibility did not mean that the required improvement was realistic. The Rio 2016 women's platform final illustrates a more moderate case: Malaysia averaged 8.80 per counted award on restricted dives compared with China's 9.30. Closing that 0.50-point award gap would have recovered 6.00 points but still left a 3.66-point deficit if the free scores stayed unchanged. The opening block supplied most of the observed margin, yet matching the champions there would not have produced a tie. This distinction is more useful for performance review than attributing the outcome solely to the restricted dives. It connects the motivating case to the wider record without converting a scoring identity into a training prescription.

Limitations

The analysis covers the complete Olympic final record for the stated period, but only seven Games and 28 finals. It is consequently a small and historically specific basis for generalisation, with 23 titles won by one national programme. Repeated divers, partnerships and national systems may create dependence beyond the Games-level clustering considered here. The census of this historical period removes sampling incompleteness within that frame, not uncertainty about future competitions or different competition levels. The study is retrospective, the motivating case was known, and no prospectively registered analysis protocol was available for this revision. Added event comparisons, score-ceiling calculations, threshold sensitivities and gradient summaries are exploratory. Confidence intervals and nominal p values should therefore be read as assumption-dependent summaries rather than certification of a stable sporting law.

Independent source verification remains partial. The official Beijing book supplied fresh checks for 176 round scores and all 32 pair totals, while an NOC narrative supplied six further round-score checks for one Paris pair. The other 1,018 round-indexed values were not freshly matched to a detailed primary record in this revision. The World Aquatics corroboration of one additional Mexican dive value does not resolve its round label independently and is recorded separately. Five inherited amended fields remain without fresh confirmation, including the 2012 women's platform runner-up's third-round score and final total. The original source values, adopted corrections and new verification status are retained so that later checking can identify precisely what would need recalculation. Successful sum checks and reproduction of an earlier table do not remove the possibility of a consistent transcription error. The statistical intervals are conditional on the adopted scores and do not incorporate uncertainty about the unaudited source records.

Difficulty data were recovered for 112 of 752 free dives (14.9%), all from Beijing 2008. The remaining 640 values and complete DD coverage for 24 finals were not recovered in this revision; this is a limitation of the assembled dataset and retrieval, not a claim that no such records exist. The four-final extension was selected by document accessibility and is reported separately rather than treated as a representative or randomly missing-data sample. No DD was inferred from a score or filled using a modern dive table. Individual execution and synchronisation awards were visible in the Beijing records but were not separately transcribed and analysed, and equivalent all-Games components and video measurements remain absent. DD-adjusted retained awards therefore remain judged-performance aggregates, not direct technical measurements. A fixed restricted multiplier also does not remove between-dive programme differences or judging context. The nine- to eleven-judge change and the fact that restricted rounds always occur first remain further comparability constraints.

The 8.0 and 80% flags are unvalidated analytical thresholds, and their sensitivity analyses do not validate a technical-error diagnosis. The two-dive award contrast is not a longitudinal measure of consistency or a same-dive repeatability estimate. Final rankings, early standings and score margins share their component scores, limiting causal and predictive interpretation. The source audit and expanded literature improve transparency without changing those design limitations. A full official-record audit, completion of historical DD and separately measured technical or judging components would permit stronger analyses. External championship data would also be needed to evaluate how well these Olympic benchmarks transfer beyond the enumerated period.

Conclusion

Restricted-round performance was a recurrent characteristic of Olympic synchronised-diving champions from 2000 to 2024, but its disproportionate contribution to winning margins was not established consistently across both event types. The platform estimate was positive, the springboard estimate was smaller and uncertain, and the direct event contrast did not establish a difference between them. Champions almost always exceeded runners-up in the restricted block, yet these scores also helped define who became champion. Low-score flags should therefore be treated as descriptive review indicators rather than validated technical errors or conditions for gold. The score-ceiling analysis further showed that restricted improvements alone could not erase the runner-up's deficit in more than half of the finals,

even under an unrealistically perfect restricted total. The study's contribution is an auditable set of historical benchmarks and scoring decompositions, strengthened by a complete official-record and DD subset for Beijing 2008. That subset does not resolve the remaining all-Games measurement gap. Claims about difficulty selection, technical consistency, prediction or training effectiveness require measurements and designs beyond those supplied by this Olympic score record.

Declarations

Disclosure statement

No potential conflict of interest was reported by the author.

Ethics and data scope

The study used publicly available competition results. It involved no direct participant recruitment, intervention or collection of private personal information. No institutional ethics-approval or exemption number is asserted in this manuscript.

Data and code availability

The accompanying reproducibility package supplies the reconstructed scores, exact compilation-source history, inherited correction log, fresh Beijing official transcription, Canadian NOC checks, an updated field-level verification register, and executable Python scripts. An enriched dataset adds source-backed DD and dive codes for Beijing 2008 while leaving the other free-dive DD values missing. The package reproduces the main numerical results from the adopted all-Games score data and the separately reported four-final DD summaries. It is a documented reconstruction, not the previously unsupplied original working file. Fresh verification covers 176 official-book scores and six NOC-reported scores; all remaining source limitations are retained explicitly. The source and reference registers accompany the dataset so that publication citations are not confused with evidence of a complete official-score audit.

Relationship to prior work

The Rio 2016 women's platform final provided the motivating case. The present manuscript extends that question across the stated Olympic period and distinguishes the original analytical question from the exploratory analyses added during revision.

Use of generative AI

Generative AI assisted with language and structural revision, source-table and official-record transcription, literature retrieval, statistical-code preparation and checking. The author retains responsibility for verifying the source records, code, references, declarations and interpretations before submission. The record-level comparisons described in the Methods do not represent an assertion that a separate human double-entry audit has been completed.

References

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, 26(4), 217–238. https://doi.org/10.2165/00007256-199826040-00002

Balmer, N. J., Nevill, A. M., & Williams, A. M. (2003). Modelling home advantage in the Summer Olympic Games. *Journal of Sports Sciences*, 21(6), 469–478. https://doi.org/10.1080/0264041031000101890

Barris, S., Davids, K., & Farrow, D. (2013). Representative learning design in springboard diving: Is dry-land training representative of a pool dive?. *European Journal of Sport Science*, 13(6), 638–645. https://doi.org/10.1080/17461391.2013.770923

Barris, S., Farrow, D., & Davids, K. (2014). Increasing functional variability in the preparatory phase of the takeoff improves elite springboard diving performance. *Research Quarterly for Exercise and Sport*, 85(1), 97–106. https://doi.org/10.1080/02701367.2013.872220

Beijing Organising Committee for the Games of the XXIX Olympiad (2008). *Diving: Official results book*. LA84 Foundation Digital Library. https://digital.la84.org/digital/api/collection/p17103coll8/id/65064/download

Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure?. *Journal of Experimental Psychology: General*, 130(4), 701–725. https://doi.org/10.1037/0096-3445.130.4.701

Bruine de Bruin, W. (2006). Save the last dance II: Unwanted serial position effects in figure skating judgments. *Acta Psychologica*, 123(3), 299–311. https://doi.org/10.1016/j.actpsy.2006.01.009

Damisch, L., Mussweiler, T., & Plessner, H. (2006). Olympic medals as fruits of comparison? Assimilation and contrast in sequential performance judgments. *Journal of Experimental Psychology: Applied*, 12(3), 166–178. https://doi.org/10.1037/1076-898X.12.3.166

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1), 1–26. https://doi.org/10.1214/aos/1176344552

Emerson, J. W., Seltzer, M., & Lin, D. (2009). Assessing judging bias: An example from the 2000 Olympic Games. *The American Statistician*, 63(2), 124–131. https://doi.org/10.1198/tast.2009.0026

Fédération Internationale de Natation (2022). *FINA diving rules 2022–2025*. FINA. https://resources.fina.org/fina/document/2022/11/30/3f3df9e6-c41a-4853-b62e-6d57ff750150/1_Diving-Technical-Rules.03.12.2022_clean.pdf

Field, C. A., & Welsh, A. H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(3), 369–390. https://doi.org/10.1111/j.1467-9868.2007.00593.x

Findlay, L. C., & Ste-Marie, D. M. (2004). A reputation bias in figure skating judging. *Journal of Sport and Exercise Psychology*, 26(1), 154–166. https://doi.org/10.1123/jsep.26.1.154

Gelman, A., & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. *The American Statistician*, 60(4), 328–331. https://doi.org/10.1198/000313006X152649

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. (2020). Array programming with NumPy. *Nature*, 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Heiniger, S., & Mercier, H. (2021). Judging the judges: Evaluating the accuracy and national bias of international gymnastics judges. *Journal of Quantitative Analysis in Sports*, 17(4), 289–305. https://doi.org/10.1515/jqas-2019-0113

Hill, D. M., Hanton, S., Matthews, N., & Fleming, S. (2010). Choking in sport: A review. *International Review of Sport and Exercise Psychology*, 3(1), 24–39. https://doi.org/10.1080/17509840903301199

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70. https://www.jstor.org/stable/4615733

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, 30(1), 1–15. https://doi.org/10.2165/00007256-200030010-00001

Hughes, M. D., & Bartlett, R. M. (2002). The use of performance indicators in performance analysis. *Journal of Sports Sciences*, 20(10), 739–754. https://doi.org/10.1080/026404102320675602

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. *Comprehensive Psychology*, 3, Article 11.IT.3.1. https://doi.org/10.2466/11.IT.3.1

Kramer, R. S. S. (2017). Sequential effects in Olympic synchronized diving scores. *Royal Society Open Science*, 4(1), Article 160812. https://doi.org/10.1098/rsos.160812

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355–362. https://doi.org/10.1177/1948550617697177

Lee, J. (2008). Outlier aversion in subjective evaluation: Evidence from World Figure Skating Championships. *Journal of Sports Economics*, 9(2), 141–159. https://doi.org/10.1177/1527002507299203

London Organising Committee of the Olympic Games and Paralympic Games (2012). *Diving: Official results book*. Games of the XXX Olympiad. https://web.archive.org/web/20130511023903/http://www.london2012.com/mm/Document/sport/General/01/40/33/53/DIVINGRESULTSBOOKV2_Neutral.pdf

Luedeker, B., & McGee, M. (2022). Relationship between judges' scores and dive attributes from a video recording of a diving competition. *PLOS ONE*, 17(8), Article e0273374. https://doi.org/10.1371/journal.pone.0273374

Malcata, R. M., & Hopkins, W. G. (2014). Variability of competitive performance of elite athletes: A systematic review. *Sports Medicine*, 44(12), 1763–1774. https://doi.org/10.1007/s40279-014-0239-x

McGee, M. (2025). Maximizing diver score by examining discrepancies in diver competency and judges' marks. *Journal of Sports Analytics*, 11. https://doi.org/10.1177/22150218251403307

Morgan, H. N., & Rotthoff, K. W. (2014). The harder the task, the higher the score: Findings of a difficulty bias. *Economic Inquiry*, 52(3), 1014–1026. https://doi.org/10.1111/ecin.12074

Nichols, P. (2024, July 29). *Wiens & Zsombor-Murray dive to bronze in men's 10m synchro at Paris 2024*. Canadian Olympic Committee. https://olympic.ca/2024/07/29/wiens-zsombor-murray-dive-to-bronze-in-mens-10m-synchro-at-paris-2024/

O'Donoghue, P. (2010). *Research methods for sports performance analysis*. Routledge. https://doi.org/10.4324/9780203878309

Pinder, R. A., Davids, K., Renshaw, I., & Araújo, D. (2011). Representative learning design and functionality of research and practice in sport. *Journal of Sport and Exercise Psychology*, 33(1), 146–155. https://doi.org/10.1123/jsep.33.1.146

Plessner, H., & Haar, T. (2006). Sports performance judgments from a social cognitive perspective. *Psychology of Sport and Exercise*, 7(6), 555–575. https://doi.org/10.1016/j.psychsport.2006.03.007

Plessner, H., Ermark, F., Schütz, L.-M., & Schweizer, G. (2023). Sports performance judgments—An update from a social cognitive perspective. *Asian Journal of Sport and Exercise Psychology*, 3(1), 13–23. https://doi.org/10.1016/j.ajsep.2023.01.002

Rio 2016 Organising Committee for the Olympic and Paralympic Games (2016). *Diving: Results book*. Olympic World Library. https://library.olympics.com/Default/detailstatic.aspx?RSC_BASE=SYRACUSE&RSC_DOCID=165312

Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten simple rules for reproducible computational research. *PLOS Computational Biology*, 9(10), Article e1003285. https://doi.org/10.1371/journal.pcbi.1003285

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Shmueli, G. (2010). To explain or to predict?. *Statistical Science*, 25(3), 289–310. https://doi.org/10.1214/10-STS330

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., Vandenbroucke, J. P., & STROBE Initiative (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *PLOS Medicine*, 4(10), Article e296. https://doi.org/10.1371/journal.pmed.0040296

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, Article 160018. https://doi.org/10.1038/sdata.2016.18

Wing, A. M., & Woodburn, C. (1995). The coordination and consistency of rowers in a racing eight. *Journal of Sports Sciences*, 13(3), 187–197. https://doi.org/10.1080/02640419508732227

World Aquatics (2024). *Paris 2024 diving results report*. World Aquatics. https://resources.fina.org/fina/document/2024/10/03/28802278-c220-48cc-996a-d4fd7200d9ce/Paris-2024-Diving-Results-Report-V2.pdf