

A COMPARISON OF THE EFFECTS OF ZIG-ZAG RUN AND SHUTTLE RUN TRAINING ON THE AGILITY OF MALE FUTSAL ATHLETES AT SMP NEGERI 11 MADIUN IN 2026

Vega Ramadhana Syahdewa^{1*}, Titin Kuntum Mandalawati², Pratama Dharmika Nugraha³

¹ Program Studi Ilmu Keolahragaan, Fakultas Ilmu Kesehatan dan Sains, Universitas PGRI Madiun, Madiun, Indonesia

vega_2204102005@mhs.unipma.ac.id; titin@unipma.ac.id; pratama.dharmika@unipma.ac.id

* Corresponding author: Vega Ramadhana Syahdewa

ARTICLE INFO	ABSTRACT
<i>Article history:</i> Received:- 21 August 2026 Revised:- 20 September 2026 Accepted:- 07 Oktober 2026 <i>Keywords:</i> Agility Futsal Illinois Agility Test Shuttle run Zig-zag run	This study compared changes in Illinois Agility Test (IAT) completion time following zig-zag and shuttle run training among male futsal athletes at SMP Negeri 11 Madiun. A two-group pretest–posttest quasi-experimental design included all 30 extracurricular athletes, allocated to matched groups of 15 using pretest-ranked ABBA ordinal pairing. The reported intervention followed a nominal six-week schedule, although session dates and the stated total require verification. IAT time was treated as planned change-of-direction performance. SPSS version 24 was used for paired and independent t-tests. Mean improvements were 1.2720 s for zig-zag ($t(14) = 13.065$, $p < 0.001$) and 1.3920 s for shuttle ($t(14) = 18.966$, $p < 0.001$). Approximate 95% confidence intervals reconstructed from summary statistics were 1.0632–1.4808 and 1.2346–1.5494 s, with paired Cohen's d_z values of 3.373 and 4.897. The between-group improvement difference was -0.1200 s (zig-zag minus shuttle), $t(28) = -0.984$, $p = 0.333$; its approximate 95% confidence interval was -0.3698 to 0.1298 s. No statistically significant between-group difference was detected, which does not establish equivalence. Both groups improved their IAT times, but the small nonrandomized sample, absence of an untreated control, and incomplete exposure reporting limit causal conclusions.

I. Introduction

Futsal requires repeated acceleration, deceleration, and rapid repositioning within a confined playing area. A review of its physical and physiological demands identifies the importance of conditioning alongside technical performance [1]. For school coaches, this makes running drills a practical component of training. However, the effectiveness of a particular drill must be evaluated through a defined outcome rather than inferred from the demands of the game.

The distinction between agility and change-of-direction (COD) speed is central to that evaluation. Agility includes responding to a stimulus, whereas a predetermined running course primarily measures planned COD speed [2], [3]. The Illinois Agility Test (IAT) used in this study does not contain an opponent-driven or externally signaled decision task. Consequently, the outcome is interpreted as IAT completion time, a measure of planned COD performance. The term agility in the unchanged title and the name of the test should be understood within that operational scope.

Zig-zag runs and shuttle runs provide different running patterns for practicing planned directional changes. Zig-zag drills involve navigating successive turns, whereas shuttle drills emphasize repeated reversals along a straight course. Comparative school-futsal research by Saputra and Bulqini [4] examined both methods over 12 meetings. That work provides local precedent, but it does not establish a universally superior drill across ages, training doses, and test protocols.

Movement geometry may help explain why different protocols produce different outcomes. A recent systematic review examines biomechanical factors associated with COD performance [5], while an earlier review shows that cutting demands depend on approach velocity and turning angle [6]. Neither source establishes that the two training programs in this study produce identical

physiological adaptations. Comparisons therefore need clear descriptions of exercise distance, turning configuration, volume, recovery, and outcome measurement.

The practical gap is whether a matched two-group comparison yields different changes in IAT time among male junior-high-school futsal participants when both groups follow the same nominal training schedule. The contribution is a direct comparison of absolute improvement, accompanied by uncertainty estimates, in a younger school setting than the vocational-school comparison [4]. This is a context-specific evaluation, not a claim that zig-zag and shuttle training have never been compared. The interpretation remains limited by nonrandom allocation and incomplete reporting of actual training exposure.

Preliminary training observations and a coach interview suggested that some players had difficulty changing direction during futsal activities. The manuscript did not specify an observation checklist, interview schedule, number of sessions observed, or systematic coding procedure. These reports are therefore presented as contextual motivation rather than quantified evidence of a population deficit. The recorded pretest means subsequently provide the numerical baseline for the intervention comparison.

The study aimed to assess pre–post changes in IAT completion time within the zig-zag and shuttle groups and to compare their mean improvements. The within-group hypotheses predicted lower posttest times. The between-group hypothesis tested whether mean improvement differed between programs using a two-tailed test; no directional superiority or equivalence hypothesis was prespecified. Practical conclusions were based on the between-group estimate and its confidence interval rather than on a comparison of within-group p-values.

II. Method

This was a two-group pretest–posttest quasi-experimental comparison. Both groups received active running interventions; there was no untreated or usual-training control. Allocation was matched rather than randomized. The original account attributed the lack of randomization to working within an established extracurricular cohort, but did not document the logistical restriction that prevented random allocation within that cohort. As a result, observed changes cannot be separated fully from routine futsal participation, maturation, test familiarity, or other concurrent influences.

The study took place at the futsal court of SMP Negeri 11 Madiun during April–May 2026. The manuscript describes a six-week program scheduled on Mondays, Wednesdays, and Fridays, from 14:30 to 15:00 WIB, with a nominal total of 18 sessions. The recorded dates were 20 April for pretest, 22 April–27 May for intervention, and 29 May for posttest. The intervention range contains 16 Monday/Wednesday/Friday dates, so the stated total and date range require reconciliation using the attendance log. Actual attendance, completed sessions, make-up sessions, and training adherence were not reported. Equal actual exposure cannot be established from the nominal schedule alone.

All 30 male participants in the school’s futsal extracurricular program were reported as the study population and were included through total sampling. The analysis tables contain 15 participants in each intervention group. Participants were ranked by pretest time from fastest to slowest and allocated using an ABBA ordinal-pairing pattern: for a consecutive block of four ranked athletes, the assignments were A, B, B, A. Group A received zig-zag training and Group B shuttle training. The source does not specify how ties or the final two participants were handled, who implemented the allocation, or whether allocation was concealed. Eligibility criteria, exact ages, training experience, and a participant-level dropout record were also unavailable. The 30 analyzed cases must not be interpreted as proof of complete attendance at all training sessions.

Zig-zag training was reported as a 30-m course using cones with 2-m and 4-m spacing. Shuttle training was reported as 30 m of accumulated running using 5-m and 10-m shuttle segments. Both programs were described as three sets of six to eight repetitions, with two to three minutes of recovery between sets. The stated intensity range was 60–85% of maximum capacity, but the reference capacity, measurement method, and monitoring procedure were not defined. Exact course geometry, short/long segment sequencing, between-repetition recovery, weekly dose, and progression rules remain undocumented. Table 1 identifies the weekly information required for reproducibility without inventing a progression. Intervention reporting should specify delivery and adherence, as recommended by TIDieR and CERT [7], [8].

Table 1. Weekly Training Progression Requiring Author Verification

Week	Sets / repetitions	Intensity	Recovery	Distance / progression
1	Not specified by week	Not specified by week	Not specified by week	Not specified by week
2	Not specified by week	Not specified by week	Not specified by week	Not specified by week
3	Not specified by week	Not specified by week	Not specified by week	Not specified by week
4	Not specified by week	Not specified by week	Not specified by week	Not specified by week
5	Not specified by week	Not specified by week	Not specified by week	Not specified by week
6	Not specified by week	Not specified by week	Not specified by week	Not specified by week

Note. The source gives only overall ranges: three sets, six to eight repetitions, 60–85% of an undefined maximum, two to three minutes between sets, and 30 m per reported course. The table is a completion record, not an invented intervention schedule. Insert actual weekly values from the training log before submission.

The outcome was IAT completion time in seconds, with shorter times indicating better planned COD performance. A digital stopwatch with a displayed resolution of 0.01 s was used; display resolution is not evidence of timing accuracy. The reported protocol began in a prone position, used a 10-m straight-running component and a central four-cone weaving course, and involved two trials with the faster time retained. Hachana et al. [9] provide primary research on IAT reliability and validity in male team-sport athletes, but that evidence does not establish reliability for this school sample. The exact width and cone spacing, warm-up, surface condition, footwear, rest between trials, timing personnel, and whether the same tester conducted pretest and posttest were not documented. Test–retest or inter-rater reliability for this sample was not reported; appropriate reliability reporting should specify the model and uncertainty of the estimate [10]. Population-specific normative cutoffs were not verified, so results are reported continuously without Excellent, Good, Average, Poor, or Very Poor labels.

Data were analyzed in SPSS version 24. Improvement was defined as pretest minus posttest time, so a positive value represents a faster posttest. Paired t-tests evaluated within-group changes, and the reported independent t-test compared mean improvements between groups at two-tailed $\alpha = 0.05$. Shapiro–Wilk tests were reported for pretest, posttest, and improvement scores. The improvement-score normality tests are relevant to both the paired and gain-score analyses; non-rejection with 15 cases per group is not proof of normality. Levene results were available for pretest and posttest, but not for improvements. The equal-variance assumption for the reported gain-score test therefore remains unverified. Outlier diagnostics and independence checks were not supplied. Since allocation used rank balancing across the entire cohort rather than documented one-to-one matched pairs, the independent comparison reproduces the source analysis; allocation-related dependence cannot be assessed from aggregate data.

The gain-score test estimates a direct comparison of improvement and is retained to reproduce the original analysis. Baseline mean times were descriptively balanced by the allocation procedure, but this does not remove confounding or justify ignoring baseline covariance. ANCOVA with posttest as outcome and pretest as covariate may provide a more efficient adjusted comparison [11]. It was not performed in this revision because individual paired scores and the covariance structure were unavailable; no adjusted result is invented.

Supplementary effect estimates and 95% confidence intervals were calculated during revision from the supplied rounded statistics [12]. For each paired comparison, $SE = \text{mean improvement}/t$, $CI = \text{mean improvement} \pm t(0.975, 14) \times SE$, and $d_z = t/\sqrt{15}$. For the between-group result, $SE = |\text{mean difference}/t|$, $CI = \text{mean difference} \pm t(0.975, 28) \times SE$, and $d_s = t \times \sqrt{(1/15 + 1/15)}$. These are approximate reconstructions, not outputs from a new raw-data analysis. A positive within-group effect indicates improvement; a negative between-group effect favors shuttle. SPSS values displayed as 0.000 are reported as $p < 0.001$.

Ethics and participant protection. School permission is acknowledged in the source manuscript, but institutional ethics approval or exemption, parental/guardian informed consent, child assent, confidentiality arrangements, and an approval identifier were not documented. The author must confirm and report the procedures actually followed; this revision does not assert approvals or consent that cannot be verified.

III. Results and Discussion

Results

The analysis included 15 participants per group. Table 2 summarizes IAT times before and after the intervention. All descriptive values are retained from the supplied manuscript; decimal punctuation and table terminology have been standardized to English.

Table 2. Results of Agility Descriptive Statistics of Futsal Athletes

Variable	Group	N	Minimum	Maximum	Mean	Std. Deviation
Pretest	Zig-zag Run	15	17.25	20.84	19.5160	1.07463
	Shuttle Run	15	17.58	20.85	19.5167	1.02113
Posttest	Zig-zag Run	15	16.15	19.79	18.2440	1.07259
	Shuttle Run	15	16.42	19.50	18.1247	0.93858

Mean pretest time was 19.5160 s in the zig-zag group and 19.5167 s in the shuttle group. These means were descriptively close, differing by approximately 0.0007 s, and their standard deviations were 1.07463 and 1.02113 s. Similar baseline means do not establish equivalence in age, experience, maturation, or other characteristics. Posttest means were 18.2440 and 18.1247 s, giving mean improvements of 1.2720 and 1.3920 s, respectively. No population-specific performance categories are assigned. Table 3 reports the available Shapiro–Wilk tests.

Table 3. Results of Agility Data Distribution Normality Test

Variable	Group	Shapiro–Wilk p	Interpretation
Pretest	Zig-zag Run	0.365	Normality not rejected
	Shuttle Run	0.414	Normality not rejected
Posttest	Zig-zag Run	0.589	Normality not rejected
	Shuttle Run	0.567	Normality not rejected

All reported pretest and posttest Shapiro–Wilk p-values exceeded 0.05. Thus, these tests did not reject univariate normality. The improvement-score results in Table 7 are more directly relevant to the t-tests of change. Small samples reduce the ability of normality tests to identify problematic departures. The available Levene tests are summarized in Table 4.

Table 4. Results of Agility Data Variance Homogeneity Test

Variable	Levene statistic	p	Interpretation
Pretest	0.008	0.931	Variance difference not detected
Posttest	0.231	0.634	Variance difference not detected

Levene p-values were 0.931 for pretest and 0.634 for posttest, providing no detected variance difference for those time points. These results do not verify equal variances of the improvement scores used in the between-group test. That diagnostic remains to be confirmed using the original output or raw paired observations.

The zig-zag within-group comparison is presented in Table 5. A positive mean difference indicates reduced IAT time after training.

Table 5. Results of Paired Sample t-Test for the Zig-Zag Run Group

Improvement (s)	95% CI (s)	t	df	p	Cohen's d _z
1.2720	1.0632–1.4808	13.065	14	< 0.001	3.373

The reported paired comparison showed a mean time reduction of 1.2720 s, $t(14) = 13.065$, $p < 0.001$. The approximate 95% confidence interval was 1.0632–1.4808 s and Cohen's d_z was 3.373. Both supplementary values were reconstructed from the rounded mean difference, t statistic, and

sample size. The change is a within-group observation; it does not establish an intervention effect against an untreated counterfactual.

The shuttle within-group comparison is presented in Table 6 using the same improvement-score convention.

Table 6. Results of Paired Sample t-Test for the Shuttle Run Group

Improvement (s)	95% CI (s)	t	df	p	Cohen's d _z
1.3920	1.2346–1.5494	18.966	14	< 0.001	4.897

The reported paired comparison showed a mean time reduction of 1.3920 s, $t(14) = 18.966$, $p < 0.001$. The approximate 95% confidence interval was 1.2346–1.5494 s and Cohen's d_z was 4.897. These supplementary values were reconstructed from rounded summary statistics. The standardized within-group change should not be interpreted as evidence that shuttle training was superior to zig-zag training.

Table 7 summarizes the Shapiro–Wilk tests of improvement scores, which are equivalent to paired differences up to a change of sign.

Table 7. Results of Agility Gain Score Normality Test

Group	Shapiro–Wilk p	Interpretation
Zig-zag Run	0.958	Normality not rejected
Shuttle Run	0.156	Normality not rejected

The gain-score normality tests did not reject normality in either group ($p = 0.958$ and $p = 0.156$). This supports proceeding with a provisional parametric summary but does not replace inspection of the individual scores. Table 8 presents the reported between-group comparison and supplementary effect estimate.

Table 8. Results of Independent Sample t-Test of Gain Scores

Difference (s)	95% CI (s)	t	df	p	Cohen's d _s
–0.1200	–0.3698–0.1298	–0.984	28	0.333	–0.359

The reported between-group improvement difference, defined as zig-zag minus shuttle, was –0.1200 s, $t(28) = -0.984$, $p = 0.333$. The reconstructed approximate 95% confidence interval was –0.3698 to 0.1298 s, and Cohen's d_s was –0.359. Negative values favor a larger mean improvement in the shuttle group. The interval includes zero and values favoring either program. The analysis failed to reject the null hypothesis of no mean difference; it did not establish equivalent effectiveness.

Discussion

Interpretation of the Zig-Zag Group Change

The zig-zag group completed the IAT faster at posttest by an average of 1.2720 s. The confidence interval describes a positive within-group change under the paired-test assumptions. The comparatively large d_z reflects change relative to variability in paired differences; it is not directly interchangeable with a between-group standardized effect [12]. The test combines straight running

and planned turns, so improvement may reflect several task components rather than agility as a purely reactive ability [3].

Repeated practice of acceleration, braking, and turning is a plausible contributor to the observed improvement. COD reviews identify turning angle, approach velocity, and whole-body movement strategies as relevant performance considerations [5], [6]. Nevertheless, muscle activation, eccentric strength, proprioception, and joint stability were not measured in this study. Those mechanisms cannot be presented as demonstrated adaptations or as explanations for a specific amount of improvement.

Interpretation of the Shuttle Group Change

The shuttle group showed a mean time reduction of 1.3920 s. Repeated running reversals plausibly practice stopping and reaccelerating, but this rationale does not isolate their effect from normal futsal training or familiarity with the IAT. The smaller standard error of the shuttle paired difference helps explain its larger t statistic and d_z ; a comparison of the two within-group p -values cannot determine which intervention was more effective.

The direction-reversal pattern differs from the sequential turns of a zig-zag course. The angle-velocity literature provides a basis for expecting differences in movement demands [6], but actual approach speeds and turn angles were not recorded here. No physiological or biomechanical monitoring was performed. Assertions about stretch-shortening-cycle adaptation, motor-unit recruitment, energy-system dominance, or injury prevention are therefore treated as untested possibilities rather than findings.

Between-Group Evidence and Practical Interpretation

No statistically significant between-group difference was detected. The mean estimate of -0.1200 s descriptively favors shuttle training, but its 95% confidence interval spans approximately 0.370 s in favor of shuttle and 0.130 s in favor of zig-zag. This range is more informative than describing the result simply as $p > 0.05$. Equivalent effectiveness would require a justified equivalence margin and a suitable equivalence analysis, neither of which was specified.

Both training methods involve planned directional changes, so overlapping task practice is a plausible interpretation of the pattern. It is not evidence that their physiological mechanisms or adaptations were identical. The prior school comparison [4] should likewise be interpreted within its own participants, intervention dose, and outcome. Differing within-group significance values do not by themselves establish differing training effects, and contextual similarity does not make results interchangeable across studies.

The practical importance of 0.1200 s cannot be determined without a justified performance threshold. Hachana et al. [9] report measurement properties in a different male team-sport sample; their detectable-change estimate should not be imported as a school-futsal decision threshold. Here, unknown timing error, incomplete adherence reporting, and sampling uncertainty further limit claims of superiority. Coaches may consider either drill within a broader program, but the present analysis does not show that they are interchangeable or that combining them is better.

Limitations and future research. The single-school male sample, nonrandom allocation, two active groups without an untreated control, and unavailable participant demographics limit causal interpretation and generalization. The inconsistent session count, missing weekly progression, and absence of adherence and concurrent-training records limit exposure assessment. Stopwatch-based IAT testing measures planned COD performance, not stimulus-driven agility. Reliability and performance norms were not established for this group. The gain-score variance assumption was not documented, and summary-statistic reconstruction cannot identify influential observations or verify every analytical decision. Future work should define the primary outcome and meaningful performance difference in advance, justify sample size in terms of precision or power [14], document actual training exposure, and use a randomized comparison when feasible. An ANCOVA sensitivity analysis is also warranted when individual baseline and follow-up observations are available.

IV. Conclusion

Both groups showed shorter IAT completion times at posttest: mean improvements were 1.2720 s for zig-zag and 1.3920 s for shuttle training (both paired $p < 0.001$). The between-group difference was -0.1200 s (zig-zag minus shuttle; approximate 95% CI -0.3698 to 0.1298 s; reported $p = 0.333$). No statistically significant between-group difference was detected, and equivalence or superiority was not demonstrated. These findings describe planned COD performance in a small nonrandomized male school cohort. The absence of an untreated control and incomplete exposure and measurement reporting require cautious interpretation.

Acknowledgment

The authors would like to express their sincere gratitude to the Principal of SMP Negeri 11 Madiun for granting permission to conduct this research. We also extend our deepest appreciation to the school's futsal coach for their invaluable assistance during the coordination of training sessions, as well as to all the male futsal athletes who participated with high dedication and enthusiasm as research subjects throughout the study.

References

- [1] N. Naser, A. Ali, and P. Macadam, "Physical and physiological demands of futsal," *Journal of Exercise Science & Fitness*, vol. 15, no. 2, pp. 76–80, 2017, doi: 10.1016/j.jesf.2017.09.001. <https://doi.org/10.1016/j.jesf.2017.09.001>.
- [2] J. M. Sheppard and W. B. Young, "Agility literature review: Classifications, training and testing," *Journal of Sports Sciences*, vol. 24, no. 9, pp. 919–932, 2006, doi: 10.1080/02640410500457109. <https://doi.org/10.1080/02640410500457109>.
- [3] S. Nimphius, S. J. Callaghan, N. E. Bezodis, and R. G. Lockie, "Change of direction and agility tests: Challenging our current measures of performance," *Strength and Conditioning Journal*, vol. 40, no. 1, pp. 26–38, 2018, doi: 10.1519/SSC.0000000000000309. <https://doi.org/10.1519/SSC.0000000000000309>.
- [4] T. D. Saputra and A. Bulqini, "Pengaruh model latihan zigzag run dan shuttle run terhadap peningkatan kelincahan pada pemain futsal SMKN 1 Trenggalek," *Jurnal Prestasi Olahraga*, vol. 7, no. 5, pp. 413–421, 2024. [Online]. Available: <https://ejournal.unesa.ac.id/index.php/jurnal-prestasi-olahraga/article/view/62847>.
- [5] U. Singh, A. S. Leicht, J. D. Connor, S. M. Brice, A. Alves, and K. Doma, "Biomechanical determinants of change of direction performance: A systematic review," *Sports Medicine*, vol. 55, no. 9, pp. 2207–2224, 2025, doi: 10.1007/s40279-025-02278-3. <https://doi.org/10.1007/s40279-025-02278-3>.
- [6] T. Dos'Santos, C. Thomas, P. Comfort, and P. A. Jones, "The effect of angle and velocity on change of direction biomechanics: An angle–velocity trade-off," *Sports Medicine*, vol. 48, pp. 2235–2253, 2018, doi: 10.1007/s40279-018-0968-3. <https://doi.org/10.1007/s40279-018-0968-3>.
- [7] T. C. Hoffmann et al., "Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide," *BMJ*, vol. 348, Art. no. g1687, 2014, doi: 10.1136/bmj.g1687. <https://doi.org/10.1136/bmj.g1687>.
- [8] S. C. Slade, C. E. Dionne, M. Underwood, and R. Buchbinder, "Consensus on Exercise Reporting Template (CERT): Explanation and elaboration statement," *British Journal of Sports Medicine*, vol. 50, no. 23, pp. 1428–1437, 2016, doi: 10.1136/bjsports-2016-096651. <https://doi.org/10.1136/bjsports-2016-096651>.
- [9] Y. Hachana, H. Chaabène, M. A. Nabli, A. Attia, J. Moualhi, N. Farhat, and M. Elloumi, "Test–retest reliability, criterion-related validity, and minimal detectable change of the Illinois agility test in male team sport athletes," *Journal of Strength and Conditioning Research*, vol. 27, no. 10, pp. 2752–2759, 2013, doi: 10.1519/JSC.0b013e3182890ac3. <https://doi.org/10.1519/JSC.0b013e3182890ac3>.
- [10] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *Journal of Chiropractic Medicine*, vol. 15, no. 2, pp. 155–163, 2016, doi: 10.1016/j.jcm.2016.02.012. <https://doi.org/10.1016/j.jcm.2016.02.012>.
- [11] A. J. Vickers and D. G. Altman, "Analysing controlled trials with baseline and follow up measurements," *BMJ*, vol. 323, no. 7321, pp. 1123–1124, 2001, doi: 10.1136/bmj.323.7321.1123. <https://doi.org/10.1136/bmj.323.7321.1123>.

- [12] D. Lakens, "Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs," *Frontiers in Psychology*, vol. 4, Art. no. 863, 2013, doi: 10.3389/fpsyg.2013.00863. <https://doi.org/10.3389/fpsyg.2013.00863>.
- [13] D. Lakens, "Equivalence tests: A practical primer for t tests, correlations, and meta-analyses," *Social Psychological and Personality Science*, vol. 8, no. 4, pp. 355–362, 2017, doi: 10.1177/1948550617697177. <https://doi.org/10.1177/1948550617697177>.
- [14] D. Lakens, "Sample size justification," *Collabra: Psychology*, vol. 8, no. 1, Art. no. 33267, 2022, doi: 10.1525/collabra.33267. <https://doi.org/10.1525/collabra.33267>.