Journal of Education and Teaching
Vol. 1, No.1 (2026): P.66-74
https://journal.terbaik.or.id/index.php/JETINDO

Improving Fourth-Grade IPAS Achievement Through Time Token Cooperative Learning: A Two-Cycle Classroom Action Study

Ismiati Hulia
Department of Primary School Teacher Education, Faculty of Teacher Training and Education, Universitas Pohuwato, Indonesia
Corresponding author: Ismiati Hulia – e-mail : ismiatihulia@gmail.com

Article History	Received: 11/04/2026	Accepted: 10/05/2026	Published: 20/05/2026

ABSTRACT

Low participation and uneven opportunities to speak can restrict concept development in elementary integrated science and social studies (IPAS). This classroom action study examined whether a Time Token cooperative learning model improved fourth-grade students' achievement and classroom participation. The participants were 28 students at SDN 02 Tilamuta, Gorontalo Province, Indonesia. The intervention was implemented through two iterative cycles of planning, action, observation, and reflection. Achievement was measured at baseline, Cycle I, and Cycle II using teacher-developed classroom tests, while affective and psychomotor participation were recorded with observation rubrics. The original score matrix was reanalysed using descriptive statistics, the Friedman test, Wilcoxon signed-rank comparisons with Holm adjustment, Cochran's Q, exact McNemar tests, and rank-based effect sizes. Mean achievement increased from 47.14 at baseline to 58.93 in Cycle I and 80.71 in Cycle II. Mastery at the school criterion of 75 rose from 25.0% to 53.6% and 89.3%. The overall change was statistically significant, Friedman chi-square(2) = 24.40, p < .001, Kendall's W = .436. The baseline-to-Cycle I contrast was not significant after adjustment, whereas Cycle I-to-Cycle II and baseline-to-Cycle II improvements were significant. The results indicate that improvement emerged mainly after reflective refinement of the second cycle. Because the study lacked a control group and used short, non-equivalent classroom tests, the findings support local instructional improvement rather than a causal effectiveness claim.

Keywords: *classroom action research; cooperative learning; elementary education; IPAS; learning achievement; Time Token*

1. INTRODUCTION

Elementary IPAS integrates natural and social phenomena and requires students to explain observations, compare evidence, and connect concepts with daily experience. These learning processes are difficult to develop when instruction is dominated by teacher explanation and only a small number of students speak. Active learning research shows that students learn more when they engage with ideas rather than only receive information (Chi & Wylie, 2014; Freeman et al., 2014). Cooperative learning can strengthen this engagement by structuring interaction, accountability, and peer explanation (Gillies, 2016; Johnson & Johnson, 2009).

The classroom problem in this study was specific. Initial observation in Grade IV at SDN 02 Tilamuta indicated limited variation in instruction, low student participation, weak confidence in discussion, and poor mastery of IPAS concepts. The school set a minimum mastery criterion of 75, but only 7 of 28 students reached this threshold at baseline. A teaching strategy was therefore needed that would distribute speaking opportunities more evenly and connect participation with conceptual work.

Time Token is a structured cooperative technique in which each student receives speaking coupons and must surrender one coupon when contributing. The procedure limits domination, requires quieter students to participate, and makes turn-taking visible. Previous Indonesian studies have reported improvements in speaking participation and learning outcomes after Time Token implementation in elementary classrooms (Asnita & Khair, 2020; Siregar et al., 2020; Terry et al., 2024). More recent evidence also indicates that the model can support elementary science achievement when its rules are implemented consistently (Mawardi & Humaira, 2025).

However, much of the available classroom research reports only percentages across cycles and treats any numerical increase as proof of effectiveness. This approach can overstate the evidence because classroom action research has no untreated comparison group, repeated scores are dependent, and short teacher-made tests may not be equivalent across occasions. The present article addresses this weakness by reanalysing the original student-level score matrix with repeated-measures nonparametric tests, multiplicity adjustment, effect sizes, and a mastery-status sensitivity analysis.

This study aimed to determine how fourth-grade IPAS achievement and observed participation changed across baseline, Cycle I, and Cycle II after the implementation and reflective refinement of Time Token cooperative learning. The study asked two questions: (1) Did achievement scores and mastery status change across the three measurement occasions? (2) Did affective and psychomotor participation improve from Cycle I to Cycle II?

2. LITERATURE REVIEW

2.1 Time Token as Structured Cooperative Participation

Cooperative learning is more than placing students in groups. Its effectiveness depends on positive interdependence, individual accountability, promotive interaction, and explicit social processes (Johnson & Johnson, 2009). Research syntheses generally show positive achievement effects, but the size of the benefit varies with implementation quality, age, subject, and cultural context (Kyndt et al., 2013). Teachers must therefore structure who participates, what students do with one another's ideas, and how responsibility is distributed.

Time Token operationalises individual accountability through limited speaking coupons. Each contribution consumes one token, so frequent speakers cannot monopolise the discussion and quieter students retain a visible obligation to contribute. This design targets unequal talk distribution, a common obstacle in group learning. It also provides teachers with a simple behavioural scaffold for monitoring participation. Studies in

E-ISSN: xxxx-xxxx
P-ISSN: xxxx-xxxx
DOI :

Indonesian elementary settings have linked the technique with improved speaking skills, communication, and classroom activity (Asnita & Khair, 2020; Siregar et al., 2020).

2.2 Participation, Guidance, and Concept Learning

The ICAP framework distinguishes passive, active, constructive, and interactive engagement. Learning is expected to deepen when students generate explanations and respond to others rather than merely attend to information (Chi & Wylie, 2014). Time Token can move students from passive attendance toward constructive and interactive participation, but speaking alone is insufficient. Contributions must be tied to the target concept, evidence, or problem.

Guidance is therefore essential, especially for elementary learners. Meta-analytic evidence on inquiry-oriented learning shows that guided participation produces stronger outcomes than minimally supported activity (Lazonder & Harmsen, 2016). In the present intervention, tokens were embedded in teacher explanation, small-group discussion, examples of forces in daily life, written group records, presentation, feedback, and reflection. The second cycle further strengthened motivation, classroom management, and rule consistency. This iterative refinement is central to action research, which uses evidence from one cycle to redesign the next (Mertler, 2020).

2.3 Research Gap and Contribution

Existing Time Token studies often focus on whether final percentages exceed a success criterion. This study adds a more cautious contribution. It separates descriptive improvement from statistical evidence, accounts for repeated observations on the same students, reports effect sizes, examines mastery transitions, and identifies where improvement occurred. The analysis also treats weaknesses in the original classroom tests as validity constraints rather than concealing them.

3. RESEARCH METHOD

3.1 Design and Setting

The study used classroom action research with two cycles. Each cycle followed planning, action, observation, and reflection. The design was appropriate for improving an identified classroom practice, but it was not designed to estimate a causal treatment effect because all students received the intervention and no comparison class was observed (Mertler, 2020). The study was conducted from November to December 2025 at SDN 02 Tilamuta in Pentadu Barat Village, Tilamuta District, Boalemo Regency, Gorontalo Province, Indonesia. The instructional topic was Chapter 3, Forces Around Us, with emphasis on types of force and examples from daily life.

3.2 Participants

The participants were all 28 students enrolled in Grade IV, comprising 14 boys and 14 girls. Whole-class inclusion was consistent with the improvement purpose of classroom action research. The classroom, not a randomly selected population sample, was the unit of practice. Generalisation is therefore analytical and context-bound.

3.3 Intervention and Action Cycles

At baseline, students completed an achievement test before Time Token was used. In Cycle I, the teacher introduced the topic, organised small groups, distributed speaking coupons, and required students to surrender one coupon for each contribution of approximately 30 seconds. Students discussed examples of force, recorded group conclusions, and presented their work. Observation results and post-cycle scores were then reviewed.

The Cycle I reflection identified three implementation problems: some students were still reluctant to speak, token use was not fully orderly, and time management limited equal participation. Cycle II therefore added explicit motivation, closer teacher guidance, clearer turn-taking rules, and more systematic monitoring of token use. The study comprised three recorded classroom meetings: one baseline meeting and one action meeting for each cycle. Each action meeting included a 10-minute introduction, approximately 50 minutes of core activity, and a 10-minute closing phase.

3.4 Measures and Data Quality

Achievement was recorded at baseline, Cycle I, and Cycle II. The source thesis described teacher-developed multiple-choice tests with five items and used a mastery criterion of 75. Student-level scores were available for all 28 participants at all three occasions. Affective participation and psychomotor participation were rated during Cycles I and II on four ordered categories from low to very good. The observed indicators included involvement, discussion activity, group cooperation, and understanding of the task.

The available records did not document content-validity indices, item difficulty, discrimination, internal-consistency reliability, inter-rater reliability, or evidence that the three achievement forms were parallel. Moreover, baseline and Cycle II scores occurred in 20-point increments, whereas Cycle I scores occurred in 25-point increments. This pattern indicates different item structures or scoring rules. The analysis therefore treated the scores as bounded classroom achievement indicators rather than as a validated interval-scale test. Nonparametric repeated-measures procedures were selected, and mastery status was analysed separately as a robustness check.

3.5 Data Analysis

Descriptive analysis included the mean, standard deviation, median, interquartile range, and the number and percentage of students reaching 75. The Friedman test assessed overall change in achievement across three dependent occasions. Kendall's W quantified the omnibus effect. When the omnibus result was significant, Wilcoxon signed-rank tests using the Pratt treatment of zero differences compared baseline with Cycle I, Cycle I with Cycle II, and baseline with Cycle II. Holm adjustment controlled the familywise error rate. Rank-biserial correlation described pairwise effect magnitude.

Cochran's Q tested overall change in binary mastery status. Exact McNemar tests examined paired mastery transitions, followed by Holm adjustment. Affective and psychomotor observation ratings were compared between cycles with Wilcoxon signed-rank tests. Statistical significance was set at .05. The reanalysis was conducted in Python 3.13 using SciPy 1.17.0 and statsmodels 0.14.6. Effect sizes were interpreted in

E-ISSN: xxxx-xxxx
P-ISSN: xxxx-xxxx
DOI :

conjunction with classroom relevance rather than through rigid universal thresholds (Lakens, 2013).

3.6 Ethical Considerations

Administrative permission was obtained from the school for the classroom improvement activity. The source thesis used coded student identifiers. This article reports only de-identified aggregate results. The thesis did not provide a formal research ethics approval number or a documented parental-consent procedure. Accordingly, this manuscript does not claim institutional ethics approval. This documentation gap is stated transparently for editorial assessment.

4. RESULTS

4.1 Achievement and Mastery Across the Three Stages

Achievement improved across the study. The mean score rose by 11.79 points from baseline to Cycle I and by a further 21.79 points from Cycle I to Cycle II. The total baseline-to-Cycle II increase was 33.57 points. The distribution also became more concentrated near 80 in Cycle II, which reduced the standard deviation. Mastery increased by 28.6 percentage points in Cycle I and by another 35.7 percentage points in Cycle II.

Table 1. Descriptive Achievement Results Across the Action-Research Stages

Stage	Mean	SD	Median	IQR	Mastery, n (%)
Baseline	47.14	28.40	50	20-65	7 (25.0)
Cycle I	58.93	22.79	75	50-75	15 (53.6)
Cycle II	80.71	12.75	80	80-80	25 (89.3)

Note. N = 28 at each stage. IQR = interquartile range. Mastery was defined as a score of at least 75. Scores are classroom assessment indicators, not standardised test scores.

4.2 Repeated-Measures Tests

The Friedman test showed an overall difference among the three stages, chi-square(2) = 24.40, $p < .001$, Kendall's W = .436. Pairwise analysis clarified that the improvement was not uniform. The baseline-to-Cycle I score change did not remain significant after Holm adjustment. In contrast, Cycle I-to-Cycle II and baseline-to-Cycle II changes were statistically significant and showed strong directional effects. No student declined from baseline to Cycle II, although seven students had the same score.

Table 2. Pairwise Changes in Achievement Scores and Mastery Status

Outcome and contrast	Mean change	Test statistic	Holm p	Effect	Transition pattern
Score: Baseline-Cycle I	+11.79	W = 143.0	.098	r_rb = .357	16 up, 12 down
Score: Cycle I-Cycle II	+21.79	W = 40.0	.002	r_rb = .724	24 up, 4 down
Score: Baseline-Cycle II	+33.57	W = 0.0	< .001	r_rb = 1.000	21 up, 7 tied

Outcome and contrast	Mean change	Test statistic	Holm p	Effect	Transition pattern
Mastery: Baseline-Cycle I	+28.6 pp	McNemar	.077	12 gained, 4 lost	7 to 15 mastered
Mastery: Cycle I-Cycle II	+35.7 pp	McNemar	.043	13 gained, 3 lost	15 to 25 mastered
Mastery: Baseline-Cycle II	+64.3 pp	McNemar	< .001	18 gained, 0 lost	7 to 25 mastered

Note. Score comparisons used Wilcoxon signed-rank tests with Pratt handling of ties. Mastery comparisons used exact McNemar tests. All pairwise p values were adjusted within outcome using Holm's method. pp = percentage points; r_rb = rank-biserial correlation. Overall mastery change was significant, Cochran Q(2) = 19.52, p < .001.

4.3 Observed Participation

Affective participation increased from M = 2.36, SD = 0.99, median = 2 in Cycle I to M = 2.93, SD = 0.94, median = 3 in Cycle II. Sixteen students moved to a higher category and 12 remained unchanged; none declined. The paired change was significant, Wilcoxon p < .001, r_rb = 1.00. Psychomotor participation increased from M = 2.43, SD = 0.92, median = 2 to M = 3.00, SD = 0.77, median = 3. Fifteen students improved and 13 were unchanged; none declined, Wilcoxon p < .001, r_rb = 1.00. These results are consistent with greater engagement, but they should be interpreted cautiously because the observation rubrics had no reported inter-rater reliability.

5. DISCUSSION

5.1 Improvement Emerged After Reflective Refinement

The most important finding is not simply that the final score was higher. The pattern shows that the first implementation produced an incomplete and statistically uncertain change, whereas the second cycle produced the clearest improvement. This distinction matters for classroom action research. Cycle I functioned as an implementation and diagnostic phase. Students were still adapting to the speaking rules, several remained reluctant, and the teacher had not yet optimised time and coupon management. The Cycle II revisions addressed these barriers directly.

The results therefore support an implementation-sensitive interpretation. Time Token did not operate as an automatic technique. Improvement became pronounced when token rules, motivation, teacher guidance, and turn-taking were coordinated. This pattern aligns with cooperative learning research that emphasises structure and teacher facilitation rather than group work alone (Gillies, 2016; Kyndt et al., 2013). It also reflects the action-research logic that instructional knowledge develops through planned intervention, evidence, reflection, and redesign (Mertler, 2020).

5.2 Plausible Learning Mechanisms

Time Token likely supported learning through three connected mechanisms. First, it redistributed opportunities to speak. This reduced domination and made participation expected rather than optional. Second, it required students to verbalise examples and explanations about force. Generating an explanation is cognitively stronger than only listening, particularly when peers respond or add examples (Chi & Wylie, 2014). Third,

E-ISSN: xxxx-xxxx
P-ISSN: xxxx-xxxx
DOI :

the visible token structure helped the teacher monitor individual accountability within small groups.

The observation results support this participation pathway. Affective and psychomotor ratings improved, and no student moved to a lower category. Prior elementary studies have similarly reported that Time Token can improve speaking involvement and communication (Asnita & Khair, 2020; Siregar et al., 2020). Studies focused on school achievement also report positive outcomes when the model is used to replace one-way instruction with structured interaction (Mawardi & Humaira, 2025; Terry et al., 2024). The present study adds that the strongest gains followed improved implementation fidelity in the second cycle.

5.3 Practical Implications

Teachers who use Time Token should connect each speaking turn to a clear cognitive task. Suitable prompts include explaining why an object moves, classifying the type of force, comparing two examples, challenging a group conclusion, or applying the concept to a new situation. Tokens should regulate access to discussion, not reward speech without conceptual content.

Implementation should begin with explicit modelling. Teachers need to demonstrate an acceptable 30-second contribution, define how students respond to peers, and monitor unused tokens. Quiet students may require sentence starters or brief preparation time. Dominant students need alternative roles after their speaking tokens are exhausted, such as recording evidence or summarising the group. A short reflection after the first lesson should identify participation imbalances and guide the next lesson.

5.4 Limitations and Validity Threats

Several limitations constrain the conclusion. First, one class and 28 students were studied without a control group. Score improvement may partly reflect repeated testing, maturation, teacher attention, or differences in task difficulty. The evidence supports improvement in this classroom, not a general causal claim about Time Token.

Second, the achievement measures were weak. Only five items were reported, validity and reliability evidence was absent, and score increments differed between occasions. The forms may therefore have measured unequal content breadth or difficulty. The mastery-status analysis reduced sensitivity to score granularity and showed the same general pattern, but it cannot establish test equivalence.

Third, participation ratings were produced during implementation without reported observer training or inter-rater agreement. Expectancy bias is possible. Fourth, the intervention was brief and limited to one IPAS topic. Future research should use parallel test forms with at least 20 well-reviewed items, document item quality and reliability, use two independent observers, measure implementation fidelity, include a delayed test, and compare Time Token with another active-learning condition through a cluster or quasi-experimental design.

6. CONCLUSION

Fourth-grade IPAS achievement and observed participation improved across two Time Token action cycles at SDN 02 Tilamuta. Mean achievement increased from 47.14 to 80.71, and mastery rose from 25.0% to 89.3%. The detailed analysis showed that the decisive improvement occurred after Cycle I reflection led to stronger motivation, clearer token rules, closer guidance, and better time management in Cycle II. Time Token is therefore best understood as a structured participation tool whose value depends on implementation quality and concept-focused discussion. The absence of a comparison group and the weakness of the classroom tests prevent a causal or broadly generalised claim. The findings justify continued local use with improved assessment and stronger research controls.

ACKNOWLEDGEMENTS

The author thanks the leadership, classroom teacher, and Grade IV students of SDN 02 Tilamuta for supporting the classroom improvement activity.

AUTHOR CONTRIBUTIONS

Ismiati Hulia: Conceptualisation, methodology, investigation, data collection, classroom intervention, data curation, original thesis preparation, and responsibility for the final manuscript.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

DATA AVAILABILITY

The de-identified score matrix supporting the reanalysis is contained in the source thesis and may be requested from the author, subject to school authorisation and child-data protection requirements.

CONFLICTS OF INTEREST

The author declares no conflicts of interest related to this research, authorship, or publication.

GENERATIVE ARTIFICIAL INTELLIGENCE DISCLOSURE

OpenAI ChatGPT was used to support English-language editing, manuscript restructuring, and statistical reanalysis of the de-identified score table. It was not used to create participants, raw data, references, ethical approvals, or research findings. Responsibility for verification and submission remains with the author.

REFERENCES

Asnita, A., & Khair, U. (2020). Penerapan model pembelajaran Time Token untuk meningkatkan keterampilan berbicara siswa. ESTETIK: Jurnal Bahasa Indonesia, 3(1), 53-74. https://doi.org/10.29240/estetik.v3i1.1501

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. Educational Psychologist, 49(4), 219-243. https://doi.org/10.1080/00461520.2014.965823

E-ISSN: xxxx-xxxx
P-ISSN: xxxx-xxxx
DOI :

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences, 111(23), 8410-8415. https://doi.org/10.1073/pnas.1319030111

Gillies, R. M. (2016). Cooperative learning: Review of research and practice. Australian Journal of Teacher Education, 41(3), 39-54. https://doi.org/10.14221/ajte.2016v41n3.3

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. Educational Researcher, 38(5), 365-379. https://doi.org/10.3102/0013189X09339057

Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning: Do recent studies falsify or verify earlier findings? Educational Research Review, 10, 133-149. https://doi.org/10.1016/j.edurev.2013.02.002

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Frontiers in Psychology, 4, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. Review of Educational Research, 86(3), 681-718. https://doi.org/10.3102/0034654315627366

Mawardi, M., & Humaira, H. (2025). Application of the Time Token Arends learning model in science learning to improve students' learning outcomes in elementary school/MI. PIONIR: Jurnal Pendidikan, 14(2), 199-210. https://doi.org/10.22373/4cmy4c71

Mertler, C. A. (2020). Action research: Improving schools and empowering educators (6th ed.). SAGE Publications.

Siregar, R., Widowati, A., & Ali, M. (2020). Meningkatkan keterampilan berbicara siswa menggunakan model Time Token pembelajaran IPA kelas V sekolah dasar. Jurnal Gentala Pendidikan Dasar, 5(1), 28-41. https://doi.org/10.22437/gentala.v5i1.9426

Terry, K., Lesnussa, A., & Johannes, N. Y. (2024). Pengaruh model pembelajaran Time Token terhadap hasil belajar IPS bagi siswa kelas V di SD Negeri 61 Maluku Tengah. Primary Didactic: Jurnal Pendidikan Sekolah Dasar, 4(2), 77-86. https://doi.org/10.30598/primary-didactic.4.2.77-86