

Integrating Lesson Study and Talking Stick Model for Enhancing Student Activity and Conceptual Understanding of Inverse Functions: Implications for AI-Supported Higher Mathematics Education

Puji Rahayu^{a,*}, Sawab Abdullah^b, Gusti Uripno^c, Mario Fahmi Syahrial^d, Arum Rahmawati^c

^{a,c,d,e} Universitas PGRI Ronggolawe, Jl. Manunggal 61, Tuban, Indonesia

^b Universitas Internasional Semen Indonesia (UISI), Gresik, East Java, Indonesia

* e-mail address: pujirahayu.mpd@gmail.com

Abstract

This paper reports on an effort to improve student learning in a tertiary mathematics course by combining two approaches: Lesson Study (LS) and the Talking Stick (TS) model. The focus was inverse functions, a topic where students at UNIROW Tuban consistently struggled — pre-study data showed 63% low participation and 68% misconception rates around domain and codomain. We used a Collaborative Action Research (CAR) design with two plan-do-see cycles and 30 third-semester students. Data came from observation, activity questionnaires, and pre/post-tests, analyzed both quantitatively and qualitatively. Results were encouraging. Participation climbed from 63% to 86%. Conceptual understanding scores rose from 68.3 to 84.7, a statistically significant gain ($t(29) = 11.45$, $p < 0.001$, Cohen's $d = 1.35$, $N\text{-Gain} = 0.74$). Domain/codomain errors dropped sharply, from 68% to just 12%. A strong correlation between participation and learning gains ($r = 0.82$) suggested the two were genuinely connected. We also discuss how the habits of reasoning developed through TS may prepare students to engage more critically with AI-assisted mathematical tools. The LS-TS combination worked here as a two-level model — improving teaching through reflection and improving learning through structured accountability. While AI tools were not directly implemented in this study, the findings highlight foundational cognitive practices — verbal reasoning, self-monitoring, and conceptual accountability — that are necessary for effective and critical engagement with AI-supported mathematical systems

Keywords: Lesson Study; Talking Stick; Inverse Functions; Student Activity; Conceptual Understanding; Collaborative Action Research; Computational Mathematics Education

1. Introduction

Teaching inverse functions at university level is harder than it looks. Many students can follow the steps: swap the variables, solve for the inverse, write the answer. What they often cannot do is explain why. They miss the underlying logic — why the function needs to be bijective, how domain and range constrain each other, what it actually means to reverse a mapping. Lecture-based instruction, for all its efficiency, tends to reward the ability to perform procedures over the ability to understand them. This has been a well-

documented problem in mathematics education for some time (Imad et al., 2023), and it is exactly the kind of gap that motivated this study.

In our own program at UNIROW Tuban, the picture was clear and troubling. Before any intervention, 68% of third-semester students made consistent errors on domain and codomain identification in inverse functions. More importantly, these errors kept appearing even after repeated instruction. That persistence is telling — it signals conceptual confusion, not just gaps in practice.

Lesson Study (LS) offers one well-established answer to this kind of problem. In LS, faculty work together: they plan a lesson as a team, one person teaches it while others observe, and then they meet to reflect on what happened and how to improve. The goal is not just a better lesson — it is better teaching practice, developed over time (Anfara et al., 2009; Poe et al., 2018; Wahman et al., 2020). The Talking Stick (TS) model addresses a different but related problem. In most classrooms, a handful of students dominate discussion. TS changes this by passing a physical stick — whoever holds it must speak. Everyone gets a turn. No one can stay silent indefinitely (Adiko & Djafar, 2022; Amir & Rizki, 2020). Together, LS and TS work on two levels: LS improves what teachers do, and TS changes what students are required to do in class.

Theoretically, this pairing draws on Vygotsky's (1978) social constructivism. The core idea is that learning happens through interaction, not in isolation. When students have to verbalize their thinking — especially in a structured, accountable way — they move from passive reception into active sense-making. LS ensures the teaching environment is designed to support this; TS ensures students actually do it. The framework predicts that this kind of social engagement produces deeper, more durable learning than lecture alone (Özbek et al., 2024)

We brought these two approaches together within a Collaborative Action Research (CAR) framework, where LS structured the Plan and See phases and TS drove the Do phase. What is missing from the existing literature is any empirical test of this specific combination — particularly its effect on inverse function misconceptions across iterative cycles at the tertiary level. This study fills that gap. It is a practical, improvement-focused study rather than a controlled experiment, and we want to be clear about that from the start. The three questions guiding the work were: (1) How does the LS-TS combination affect student activity across cycles? (2) How does it influence conceptual understanding of inverse functions? (3) What does the LS reflection process contribute to instructional improvement between cycles?

2. Method

2.1. Study Context and Participants

We used a two-cycle Collaborative Action Research design, following the Plan–Do–See sequence. The study ran during the odd semester of 2024/2025 in the Mathematics Education Study Program at UNIROW Tuban. Participants were 30 third-semester students from Class A, enrolled in Elementary Algebra with a focus on inverse functions. We chose this class deliberately — pre-study data showed a 63% activity rate and 68% misconception rate, which made the need for intervention concrete and visible. Each cycle had its own separately constructed pre-test as a within-cycle baseline. These scores are

not comparable across cycles; they exist only to anchor the N-Gain and paired T-test computations within each cycle.

2.2. Data Collection Instruments

Three instruments were constructed for this study. To measure conceptual understanding, we developed ten essay-format items in total — two pre-test and three post-test items per cycle — targeting three aspects of inverse function understanding: recognizing domain–range relationships, determining when a function is invertible, and explaining the inverse process itself. Three mathematics education specialists reviewed these items for content fit, and reliability testing confirmed Cronbach’s Alpha ≥ 0.80 across both cycles. Student activity was measured using a twenty-item questionnaire covering how students engaged visually, orally, through listening, in writing, and cognitively during class. Expert review confirmed its relevance and reliability (Cronbach’s Alpha ≥ 0.75). To capture what was actually happening in the classroom, we also used structured faculty observation sheets that recorded how often students and instructors interacted and how substantively students discussed the mathematics. Two observers rated independently; their agreement was strong ($\kappa = 0.89$), which gave us confidence that what was being recorded reflected what was actually occurring.

2.3. How Each CAR Cycle Was Implemented

Each cycle moved through three phases. In the planning phase, the teaching team worked together to design TS-based lesson sequences, anticipating where students were likely to struggle and building tasks specifically intended to surface those difficulties. In the implementation phase, inverse function lessons were delivered using the Talking Stick protocol: as the stick rotated through the group, each student took sequential responsibility for speaking. One team member taught while the others sat at the back observing and taking notes. After each lesson, the team gathered for a structured reflection session, reviewing their observation records, any video footage, and samples of student work. That review drove the specific changes carried forward into the next cycle

2.4. Analytical Approach

This study prioritizes ecological validity over experimental control, consistent with the improvement-oriented nature of CAR. The goal here is not to isolate a causal mechanism but to understand, in rich and contextually grounded terms, what shifted and why in this particular setting. Quantitative data were analyzed using descriptive statistics (means and percentages) and inferential analysis through a Paired Sample T-Test conducted separately for each cycle (significance level: $p < 0.05$). Effect size was calculated using Cohen’s d with SPSS 25. Normalized learning gain was computed following Hake (1998): $N\text{-Gain} = (\text{post} - \text{pre}) / (\text{max} - \text{pre})$, with values categorized as High (>0.70), Medium ($0.30\text{--}0.70$), and Low (<0.30). For the qualitative strand, we worked through observation field notes and team reflection records by coding inductively and then converging on recurring themes, following the six-phase analytical process Braun and



Clarke (2006) outline. Credibility was strengthened by triangulating across the three data sources rather than relying on any single one. The study received ethics approval from UNIROW’s Research Ethics Committee, and all participants gave written informed consent before data collection began.

3. Results and Discussion

3.1. Student Activity Enhancement

Results and interpretation are woven together below, since one is hard to make sense of without the other. Participation improved in both cycles, though not at the same rate or in the same way for every student. Table 1 shows the headline numbers.

Table 1. Percentage of Student Activity from Pre-Cycle to Cycle II

Cycle	Mean Score (Scale 100)	Activity (%)	Increase (%)	Active Students (n=30)
Pre-cycle	63.5	63	—	19
Cycle I	76.2	76	+13	23
Cycle II	85.7	86	+10	26

By Cycle II, 86% of students were actively participating. But the numbers alone do not tell the full story. Some students took to the Talking Stick quickly. Others needed the whole first cycle before they would respond without obvious nervousness. What we observed was not students who didn't understand the material — it was students who didn't know how to put their understanding into words. The TS structure addressed that directly, in a way that verbal encouragement from us simply could not.

Classroom observation records from Cycle II capture two instructive cases. Student M-12, previously among the least vocal, stated: “Because the stick was rotating, I had to be brave enough to speak. An inverse is a reversal: the original function goes from X to Y, so the inverse goes from Y to X” (Observation Notes, Cycle II). Student M-07 identified a metacognitive effect: “When I had to explain to others, I realized there were parts I did not fully understand yet” (Observation Notes, Cycle II). These two accounts exemplify the broader pattern through which TS-structured accountability rendered latent conceptual gaps visible and open to pedagogical intervention.

By Cycle II, it was not just the previously active students who improved — the gains were spread across the group. That breadth matters. It fits with what the ICAP framework would predict: when students are required to actively construct and interact rather than just receive, the benefits reach further (Natsis et al., 2018; Nicolajsen & Plank, 2015; Tithi et al., 2026). In a class as mixed in background as ours, that kind of broad reach is significant.

3.2. Conceptual Understanding Enhancement

Table 2. N-Gain Results of Conceptual Understanding per Cycle

Cycle	Mean Pre-test	Mean Post-test	N-Gain (g)	Category (R. Hake, 1998)
Pre-cycle	42.10	—	—	—
Cycle I	45.30	68.50	0.55	Medium
Cycle II	68.30	84.70	0.74	High

The Cycle II gains were statistically significant: $t(29) = 11.45$, $p < 0.001$, Cohen’s $d = 1.35$, N-Gain = 0.74 (High category, Hake, (1998)). A correlation of $r = 0.82$ between activity and learning gains suggested that the two were not just happening at the same time — they appear genuinely linked, which is consistent with the ICAP framework’s predictions about interactive engagement (Chi & Wylie, 2014).

The clearest sign of real conceptual change was the drop in domain/codomain errors: from 68% before the intervention to 12% by the end of Cycle II. Students who had previously just executed procedures were now explaining, out loud, why a function must be bijective for an inverse to exist. Not all of them got it perfectly — some still mixed up domain restrictions with range restrictions in Cycle I. But the conversations that TS generated brought these confusions into the open in a way that written tests never had. That, in itself, was valuable.

Contextualizing the present findings within the relevant literature: Amir and Rizki (2020) reported a 17.4-percentage-point increase in student activity using TS in secondary mathematics; Rahmawati and Widodo (2022) obtained an N-Gain of 0.61 through LS-integrated instruction with pre-service teachers. The present study’s results are numerically higher on both dimensions; however, this comparison must be approached with considerable caution given the substantial heterogeneity in research designs, educational levels, and participant profiles. A plausible interpretation is that the combined LS-TS architecture creates a mutually reinforcing dynamic that neither component could achieve alone. Freeman et al.’s (2014) meta-analytic effect size benchmark for active STEM learning ($d = 0.47$) provides an additional reference frame, though the single-site, non-comparative CAR design limits the interpretability of this juxtaposition. The observed effect size of $d = 1.35$ should be treated with caution; improvement-focused action research designs are inherently prone to inflating locally observed gains relative to what controlled experimental designs would yield.

These findings carry theoretical implications that extend beyond their local context. The observed transition from procedural participation to genuine conceptual reasoning — evidenced by the sharp reduction in domain/codomain errors and the emergence of spontaneous verbal explanation — suggests that structured accountability mechanisms such as TS may operationalize the transition from active to interactive cognitive engagement as conceptualized within the ICAP framework (Chi & Wylie, 2014). That is, TS does not merely increase participation; it appears to shift the quality of that participation from surface-level response to constructive and interactive knowledge-building. At the same time, the LS reflection cycles demonstrate that this shift does not occur automatically: it requires deliberate instructional scaffolding, iterative refinement, and collaborative faculty reflection. Together, the LS-TS model provides empirical



grounding for the theoretical proposition that structured social accountability, when coupled with reflective instructional design, can reliably move learners up the ICAP hierarchy from passive to interactive engagement — even in topics as conceptually demanding as inverse functions.

3.3. Lesson Study as a Reflective Mechanism

Table 3. LS Interventions and Scaffolding Changes Across Cycles

Cycle	Critical Finding (See Phase)	(See Finding)	Adjustment (Next Plan Phase)	Observed Result
I	Participation procedural; hesitant to reason aloud.	mostly students	Slowed TS rotation; replaced 30% of items with HOTS questions; revised rubric to weight justification.	Richer responses; more conceptual reasoning visible.
II	Students active but still waiting for stick rather than initiating.		Developed independent module; introduced reflective closure.	Early signs of more spontaneous engagement.

Table 3 captures the changes we made, but not the conversations behind them. After Cycle I, we had a long reflection session. The main thing we realized was this: the questions we were asking were procedural, so the answers we got were procedural. That sounds obvious in hindsight, but it took observing our own class to see it clearly. We replaced 30% of the tasks with higher-order questions. That was a real risk — we genuinely did not know how the students would respond. In Cycle II, responses were richer. But something remained: students were still waiting for the stick to reach them rather than volunteering. They had become engaged, but not yet independent. That gap is worth noting honestly.

Across the two cycles, something changed in how we thought about our role as instructors. Less about covering content, more about creating conditions where students could think out loud. That shift did not happen all at once — it accumulated through the See phases, through watching students, and through honest team discussion. Wenger (1998) describes this kind of change as professional identity transformation within communities of practice, and that description fits what we experienced.

3.4. Potential Alignment with AI-Supported Environments

We did not study AI tools in this research. But the connection to the title of this paper — and to the direction of the field — requires explicit engagement. The LS-TS model, as implemented here, developed specific cognitive habits in students: saying your reasoning out loud, owning your interpretation of a mathematical object, noticing when an answer feels wrong. These are not incidental by-products. They are the foundational practices that distinguish a student who can critically evaluate AI-generated output from one who simply accepts it. Tools like Wolfram Alpha or GeoGebra AI can produce plausible-

looking errors. A student trained to reason procedurally and silently is poorly equipped to catch them. A student trained through TS-style accountability — who has spent a semester being held responsible for explaining why, not just what — is in a fundamentally different position. The direct empirical test of this relationship remains for future research (McPhee & Jerowsky, 2025; Roll & Wylie, 2016). But the implication is clear enough to name: the LS-TS model may serve as a preparatory pedagogical framework for AI-integrated mathematics learning environments, building the very reasoning dispositions that AI-assisted instruction demands but cannot itself supply.

The instructional context provides immediate grounding for this alignment. University mathematics students increasingly rely on AI-assisted computation platforms — among them Wolfram Alpha and GeoGebra AI — that deliver high-powered outputs but are not immune to producing confident-sounding errors. A student whose educational experience has not systematically cultivated the habit of reasoning aloud and defending interpretations is poorly equipped to recognize or challenge such outputs when they are incorrect. The question of whether structured participation in LS-TS cycles builds precisely this kind of transferable critical capacity is left to future research. The connection is offered not as a causal claim but as a theoretically grounded proposition deserving of rigorous empirical follow-up in the emerging field of AI-ready mathematics pedagogy.

4. Conclusion

This study asked a practical question: can combining Lesson Study with the Talking Stick model, within a CAR framework, meaningfully improve how undergraduate students engage with and understand inverse functions? Based on two cycles of implementation with 30 students at UNIROW Tuban, the answer appears to be yes — with important qualifications.

Participation reached 86%. Conceptual scores improved significantly ($t(29) = 11.45$, $p < 0.001$; $N\text{-Gain} = 0.74$; $d = 1.35$). Domain/codomain misconceptions fell from 68% to 12%. The correlation between participation and learning ($r = 0.82$) suggests these outcomes were connected, not coincidental. We find these results meaningful, while being honest about their limits: this was a single-site study with no control group.

CAR is a local method. What worked here, in this class, with these students, will not automatically generalize. We also cannot rule out a Hawthorne effect: students knew they were being observed, and that awareness may have influenced how they behaved. Our observers were colleagues from the same program who had a stake in the study succeeding — that introduces bias risk, even with strong inter-rater agreement ($\kappa = 0.89$). These are real limitations. They do not erase the findings, but they do mean the next step is a quasi-experimental design with a control group, delayed post-tests, and external observers.

What makes the LS-TS model practical is that it does not require special equipment or radical curriculum change. It requires faculty willing to observe each other and be honest about what they see, and a classroom structure that holds every student accountable for speaking. Both of those things are achievable. The model improves with use — each cycle produces better information for the next. Crucially, the cognitive dispositions developed through this model — verbal reasoning, conceptual accountability, and reflective self-

monitoring — are precisely those that students will need to engage productively with AI-supported mathematical tools. Thus, the LS–TS model may serve as a preparatory pedagogical framework for AI-integrated mathematics learning environments, equipping students not just to solve problems but to reason about the solutions that AI systems offer them. For future research, the priority should be quasi-experimental studies with control groups and external observers, plus targeted empirical investigation of how TS-developed reasoning habits transfer to AI-assisted learning contexts. That connection is promising and, so far, untested.

Acknowledgements

The authors gratefully acknowledge the support of the Universitas Ronggolawe (UNIROW) Tuban, and Universitas Internasional Semen Indonesia (UISI) Gresik. We thank all students who participated in this study and all colleagues who contributed to the collaborative reflection process. This research was conducted during the 2024/2025 academic year as part of a faculty-initiated Collaborative Action Research program aimed at improving undergraduate mathematics instruction

References

- Adiko, H. S. S., & Djafar, M. (2022). Fostering the use of Talking Stick learning model on student learning outcomes. *Journal of Curriculum and Teaching*, *11*(4), 14–24. <https://doi.org/10.5430/jct.v11n4p14>
- Amir, Z., & Rizki, D. (2020). Pengaruh model pembelajaran Talking Stick terhadap keaktifan dan hasil belajar matematika siswa. *Jurnal Pendidikan Matematika Raflesia*, *5*(1), 34–41. <https://doi.org/10.33449/jpmr.v5i1.9890>
- Anfara, V. A., Lenski, S. J., & Caskey, M. M. (2009). Using the Lesson Study Approach to Plan for Student Learning. *Middle School Journal*, *40*(3), 50–57. <https://doi.org/10.1080/00940771.2009.11495587>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- 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>
- Imad, M., Reder, M., & Rose, M. (2023). Recasting the agreements to re-humanize STEM education. *Frontiers in Education*, *8*(October), 1–21. <https://doi.org/10.3389/feduc.2023.1193477>
- McPhee, S., & Jerowsky, M. (2025). Beyond technical skills: A pedagogical perspective on fostering critical engagement with generative AI in university classrooms. *Frontiers in Education*, *10*, 1593278. <https://doi.org/10.3389/feduc.2025.1593278>

- Natsis, A., Papadopoulos, P. M., & Obwegeser, N. (2018). Research integration in information systems education: Students' perceptions on learning strategies, skill development, and performance. *Journal of Information Technology Education: Research*, 17, 345–363. <https://doi.org/10.28945/4120>
- Nicolajsen, P. H., & Plank, B. (2015). Using knowledge components for collaborative filtering in adaptive tutoring systems. *Proceeding of the 8th International Conference on Educational Data Mining, EDM15*, 536–539. http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_86.pdf%5Cnpapers2://publication/uuid/17248F24-FCA8-4C3D-92D4-106C2096021F
- Özbek, T., Wekerle, C., & Kollar, I. (2024). Fostering pre-service teachers' technology acceptance – does the type of engagement with tool-related information matter? *Education and Information Technologies*, 29(5), 6139–6161. <https://doi.org/10.1007/s10639-023-12047-2>
- Poe, M., Inoue, A. B., & Elliot, N. (2018). The End of Isolation. In *The End of Isolation* (p. 3). The WAC Clearinghouse; University Press of Colorado. <https://doi.org/10.37514/per-b.2018.0155.1.3>
- R. Hake, R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74. <http://link.aip.org/link/?AJPIAS/66/64/1%5Cnpapers2://publication/doi/doi:10.1119/1.18809%5Cnhttp://www.mendeley.com/research/interactiveengagement-versus-traditional-methods-a-sixthousandstudent-survey-of-mechanics-test-data-for-introductory-physics-courses/%5Cnhttp://web.mit.edu/rsi/www/2005/misc/minipaper/papers/Hake.pdf%5Cnpapers2://publication/uuid/8ED66261-06F4-47E8-9F78-CAF601F68AF5>
- Rahmawati, D., & Widodo, S. A. (2022). Integrasi Lesson Study dalam pembelajaran matematika untuk meningkatkan kemampuan berpikir kritis mahasiswa calon guru. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 6(2), 1241–1255. <https://doi.org/10.31004/cendekia.v6i2.1546>
- Roll, I., & Wylie, R. (2016). Evolution and Revolution in Artificial Intelligence in Education. *International Journal of Artificial Intelligence in Education*, 26(2), 582–599. <https://doi.org/10.1007/s40593-016-0110-3>
- Tithi, S. D., Alam, N., Yasir, T., Shi, Y., Tian, X., Chi, M., & Barnes, T. (2026). *Adaptive Scaffolding for Cognitive Engagement in an Intelligent Tutoring System*. 1–14. <http://arxiv.org/abs/2602.07308>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wahman, M. L., Peplow, A., Kumar, R., & Refaei, B. (2020). Benefits of Using Lesson Study for SoTL, Cross-Disciplinary Research, and Assessment. *International Journal for the Scholarship of Teaching and Learning*, 14(1). <https://doi.org/10.20429/ijstol.2020.140102>
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511803932>