Vol. 4, No. 4, December 2026, pp. 3239-3251
INTERNATIONAL JOURNAL OF
POST AXIAL
FUTURISTIC TEACHING AND LEARNING
https://journal.amorfati.id/index.php/postaxial||ISSN 3025-7549

Rapulung Local Wisdom and Changes in Students' Social Deliberation Skills: A Pretest–Posttest Study in Sociology

Mardiana [1a*], Anik Widiastuti [2b], Riko Septiantoko [3c]
[1,2,3] Universitas Negeri Yogyakarta, Indonesia
[a]mardiana.2026@student.uny.ac.id, [b]anikwidiastuti@gmail.com, [c]rikoseptiantoko@uny.ac.id

Article History:
Received:
09-10-2026
Revised:
10-10-2026
Accepted:
18-10-2026

Keywords:
Paired-samples t-test; Pretest–posttest; Rapulung; Social deliberation; Sociology education

***Correspondence Address:**
mardiana.2026@student.uny.ac.id

Abstract:
This study documents changes in students' social deliberation skills following Sociology learning integrating Rapulung, a local deliberative tradition of the Samawa community. A quantitative, one-group pretest–posttest design examined 36 paired scores from class XI B (XI IPS 1) at MAN 1 Sumbawa Barat. Analyses included descriptive statistics, the Shapiro–Wilk test on paired differences, a paired-samples t-test, Wilcoxon signed-rank test, and standardized effect size. The mean increased from 62.64 (SD = 9.48) to 77.61 (SD = 11.20), a gain of 14.97 points; 35 students improved and one declined by one point. The Shapiro–Wilk test did not reject normality of paired differences ($W = 0.971$; $p = 0.467$). The mean increase was statistically significant, $t(35) = -14.803$, $p < 0.001$, 95% CI [12.92, 17.03], with Cohen's $dz = 2.467$; the Wilcoxon result was consistent ($p < 0.001$). Mean individual normalized gain was 0.434. The contribution lies in combining paired statistical inference with individual change patterns to assess social deliberation in a culturally situated learning context. Because no comparison group was included, the observed changes cannot be attributed exclusively to Rapulung integration.

This is an open-access article under the CC–BY-SA license.

Introduction

Sociology education requires more than knowledge of conflict, integration, social change, and social differentiation. Students must also evaluate competing reasons when confronted with conflicting interests. In conflict situations, correctly identifying the causes of a problem does not necessarily entail listening to contrasting arguments or anticipating the social consequences of a decision. Social deliberation is therefore an important learning outcome that connects evidence-based reasoning, openness to dialogue, and responsibility for collective decisions. Hess and McAvoy (2015) regard the discussion of controversial issues as central to democratic education when students are given opportunities to assess arguments openly.

Social deliberation is understood as the capacity to develop shared judgments by framing problems, presenting arguments, assessing evidence, responding to objections, negotiating

alternatives, and making decisions that recognize consequences for different parties. It is not synonymous with verbal participation. Frequent contributions may indicate dominance, whereas a concise response that evaluates evidence and acknowledges objections may move a discussion toward a more defensible decision. This distinction is consistent with McAvoy and Hess (2013), who regard deliberative classroom discussion as democratic engagement with disagreement rather than a search for uniform answers.

Within Samawa society, Rapulung refers to a deliberative tradition that enables the exchange of perspectives and the pursuit of agreement on collective concerns. Dianto (2021) associates Rapulung with participation and collective consideration in Sumbawa's social life. At the regional policy level, West Sumbawa Regency Regulation No. 2 of 2021 concerning Bale Mediasi similarly recognizes deliberation and local wisdom in dispute resolution. This context matters because deliberative values can be introduced not merely as abstract principles but as social practices culturally familiar to learners.

Integrating Rapulung into Sociology learning can bridge conceptual knowledge and the practice of social reasoning. While examining social conflict, students need to distinguish stakeholders' positions from their underlying interests, scrutinize the grounds of claims, and assess solutions that do not simply privilege the dominant party. Connecting familiar social contexts with problems requiring collective judgment can create opportunities for meaningful participation. As emphasized by OECD (2025), the quality of classroom interaction depends not merely on communication frequency but on opportunities for students to develop ideas and build on others' thinking.

Rapulung's pedagogical significance lies in linking social engagement with argumentative discipline. Reaching consensus does not in itself establish high-quality deliberation if reasons remain untested or dissenting students lack opportunities to speak. Accordingly, instruction should attend to how students formulate reasons, respond to objections, identify the consequences of decisions, and consider underrepresented interests. This approach resonates with OECD (2019), which treats critical thinking as a practice of examining ideas rather than simply repeating answers.

Research on local wisdom provides valuable cultural and normative accounts of collective problem solving; for example, Dianto (2021) examines Rapulung in the context of regional development planning. Research on classroom deliberation, by contrast, has examined reason-giving, disagreement, and democratic participation in educational settings (McAvoy & Hess, 2013; Hess & McAvoy, 2015). These strands illuminate the value of participatory traditions, but neither alone describes the distribution of individual changes in an assessed Sociology learning outcome following the integration of Rapulung. The present study connects them through paired evidence rather than treating cultural familiarity as proof of learning effectiveness.

The empirical gap is particularly relevant to the study of social conflict. Recognizing the causes or forms of conflict is different from assessing competing claims, considering less powerful stakeholders, and justifying an acceptable collective response. A quantitative account can establish whether assessed performance differs across two observations, how consistently students change, and how large that difference is. It cannot, without a comparison condition, establish that a cultural teaching approach produced the difference. Thus, the novelty of this study lies in its outcome-focused, individual-pair analysis of social deliberation within a local-wisdom context, not in a claim of experimentally verified causality.

This study treats changes in social deliberation scores as an empirical question examined through paired measurements. The analysis extends beyond statistical significance to the magnitude of score differences, consistency of individual change, and distribution of relative gains. Accordingly, pretest–posttest differences can be interpreted at two levels: shifts in the

group's overall performance and heterogeneity in individual development. Distinguishing these levels prevents the group mean from concealing variation among students (Field, 2018).

This study examines the difference between students' social deliberation scores before and after Rapulung-integrated Sociology learning on social conflict. Specifically, it characterizes baseline and final achievement, tests paired differences, quantifies the magnitude and variability of change, and relates the pattern to theories of classroom deliberation. The contribution is a transparent account of 36 matched learning outcomes, including students whose gains were limited, that can inform culturally responsive assessment and more rigorous subsequent evaluations.

Method

2.1 Research design

This study used a quantitative, pre-experimental one-group pretest–posttest design. The same 36 students in class XI B (XI IPS 1) at MAN 1 Sumbawa Barat completed the two assessments around Sociology learning integrating Rapulung local wisdom. Every pretest result was matched to its corresponding posttest result, allowing within-student differences to be analyzed (Creswell & Creswell, 2018). The primary estimand was the mean difference between the two occasions. As the design had no untreated or alternative-instruction comparison group, this difference describes observed change rather than an isolated causal treatment effect.

Students were anonymized using codes S01–S36 so that each pretest score could be linked to the corresponding student's posttest score. Change was defined as the posttest score minus the pretest score. Positive values indicated improvement, negative values indicated a decline, and zero indicated no change for the relevant pair.

2.2 Participants and research variables

The analyzed class comprised 12 male students (33.33%) and 24 female students (66.67%). The outcome was a composite social deliberation performance score on a 0–100 scale at each assessment. The conceptual domain comprised five aspects: (1) constructing reasons and evaluating the evidentiary basis of arguments; (2) recognizing and responding to divergent positions; (3) anticipating consequences for parties affected by conflict; (4) weighing and negotiating alternatives; and (5) articulating a defensible collective decision. These aspects describe the construct represented by the total score. The dataset used in the statistical analyses contains paired composite scores rather than separate scores for each aspect; consequently, differential improvement by domain cannot be calculated from the reported results.

The two measurement occasions used the same score range to permit within-person comparison. Each recorded total was treated as a single composite observation, and no assumption was made that the five conceptual aspects contributed equally to students' gains. The paired-score dataset does not contain item responses or rater-level judgments from which item validity, internal consistency, or inter-rater agreement could be estimated. Accordingly, the inferential findings concern composite-score change and should not be read as evidence of psychometric equivalence across the five aspects.

Table 1. Characteristics of the participants

Characteristic	Frequency	Percentage
Male	12	33.33%
Female	24	66.67%
Total	36	100.00%

2.3 Framework for integrating Rapulung into instruction

Rapulung provided a culturally situated framework for discussing social conflict and reaching collective judgments. Its deliberative logic can be expressed as a sequence of five learning actions: examining a conflict and identifying the affected actors; making claims supported by reasons; listening and responding to different positions; weighing the likely consequences of competing solutions; and formulating an agreement that can be justified to the parties involved. The focus is not the speed of reaching consensus but the quality of reasoning preceding it. This account specifies the instructional logic through which Rapulung was connected to the social-conflict topic, without assuming that each action was separately scored.

In learning activities, a conflict case provides the common question for students to consider, competing positions provide material for reason-giving, and the deliberative stage directs attention to objections and the consequences of alternatives. The concluding stage places emphasis on the justification of a shared decision and on whose interests may have been overlooked. These steps make the connection between local deliberative values and Sociology concepts explicit. The data reported here consist of composite pretest and posttest scores; lesson-by-lesson exposure time and independent process-observation scores are not part of the quantitative outcomes analyzed in this article.

The analytic interpretation distinguishes the social deliberation construct from the measure available for inference. An improved total score is compatible with stronger deliberative performance, but without separately reported indicator scores and records of classroom interaction it cannot demonstrate how every learner reasoned, listened, or negotiated during discussion. The same distinction informs interpretation of both the effect size and the individual gains.

2.4 Data analysis

Analysis began by checking whether all observations formed complete pairs, followed by descriptive statistics covering means, medians, standard deviations, ranges, and confidence intervals. The normality assumption for the paired t-test was examined primarily on paired difference scores using the Shapiro–Wilk test. Kolmogorov–Smirnov results with the Lilliefors correction are reported as supplementary information. The emphasis on paired differences reflects the distribution relevant to testing mean change (Shapiro & Wilk, 1965; Field, 2018).

A two-sided paired-samples t-test was conducted at $\alpha = 0.05$, together with a 95% confidence interval and Cohen's dz effect size. The Wilcoxon signed-rank test assessed the consistency of the directional findings, while Pearson correlation described the association between measurements. Relative improvement was estimated using individual normalized gain, calculated as (posttest − pretest)/(100 − pretest). Mean individual N-gain was distinguished from gain based on the class mean because their denominators differ (Hake, 1998).

Results

3.1 Data completeness and score changes

All 36 students had complete scores at both measurement points (Table 2). Table 1 summarizes the group composition, and Table 12 presents each matched score pair. With no missing observations, the descriptive and paired inferential analyses used the same 36 students.

Table 2. Case processing summary

	Valid		Missing		Total	
Variable	N	Percent	N	Percent	N	Percent
Pretest	36	100.0%	0	0.0%	36	100.0%
Posttest	36	100.0%	0	0.0%	36	100.0%

	Valid		Missing		Total	
Variable	N	Percent	N	Percent	N	Percent
Difference	36	100.0%	0	0.0%	36	100.0%

The combined pretest score of 2,255 increased to 2,794 at posttest, representing a total gain of 539 points across all pairs. Scores improved for 35 students (97.22%), while one student (2.78%) declined by one point; no score remained unchanged. Thus, the rise in the class mean reflects improvement among nearly all students rather than an unusually large increase in just a few individuals (Table 3).

Table 3. Summary of pretest and posttest scores

Indicator	Pretest	Posttest	Change
Number of observations	36	36	36 pairs
Total score	2255	2794	+539
Mean	62.64	77.61	+14.97
Minimum	46	57	–
Maximum	82	100	–
Standard deviation	9.48	11.20	6.07
Students with improvement	–	–	35 (97.22%)
Students with decline	–	–	1 (2.78%)

3.2 Descriptive statistics for social deliberation skills

Table 4 summarizes the distributions of pretest, posttest, and within-student differences without repeating the raw values. The median and interquartile range complement the means and standard deviations, while the confidence intervals provide precision estimates for the population mean under the corresponding sampling assumptions. The paired difference (posttest minus pretest) is the outcome of primary interest.

Table 4. Descriptive statistics

Statistic	Pretest	Posttest	Gain (Post − Pre)
Valid N	36	36	36
Mean	62.64	77.61	14.97
Standard deviation	9.48	11.20	6.07
Median	60.50	74.00	16.50
Minimum	46	57	−1
Maximum	82	100	26
Interquartile range	11.75	16.25	8.00
95% CI of mean	59.43–65.85	73.82–81.40	12.92–17.03
Skewness	0.52	0.39	−0.52
Kurtosis	−0.43	−0.60	0.09

Mean achievement rose from 62.64 (SD = 9.48) to 77.61 (SD = 11.20), an increase of 14.97 points or approximately 23.90% relative to the baseline mean. The median rose from 60.50 to 74.00, indicating that the observed shift was not confined to unusually high scores. The posttest standard deviation was larger than the pretest standard deviation, so between-student variation remained evident. Two students reached 100 at posttest, while the lower end of the distribution increased from 46 to 57. These patterns establish the size and heterogeneity of score change, not an independent instructional effect.

3.3 Distributions of pretest, posttest, and change scores

The histograms in Figures 1–3 show a shift in the distribution toward higher posttest scores. Paired differences were predominantly positive, with a median improvement of 16.50 points. These visual patterns are consistent with the individual-level records, in which 35 differences

were positive and one was negative. Histogram interpretation was combined with normality testing of paired differences before conducting parametric inference.

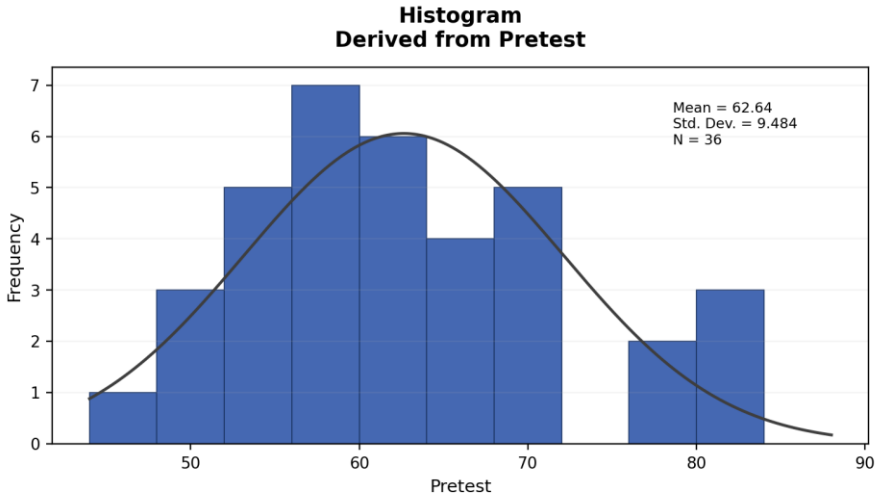


Figure 1. Pretest score histogram with a normal curve.

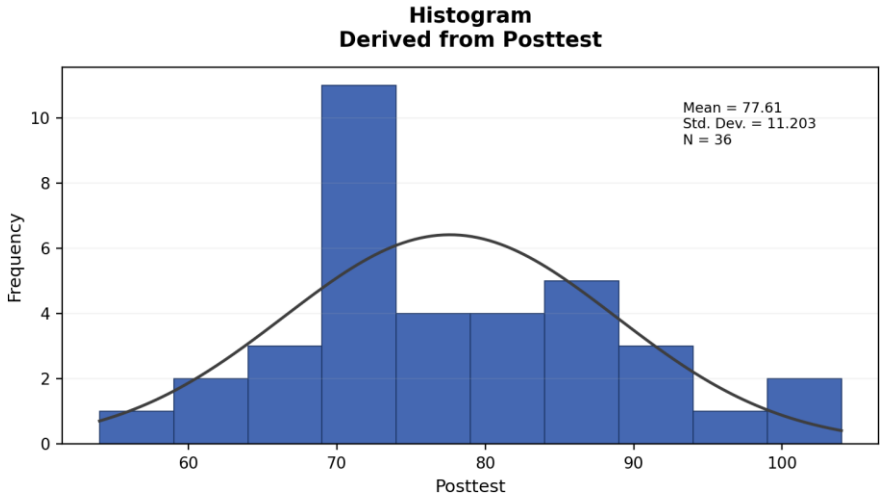


Figure 2. Posttest score histogram with a normal curve.

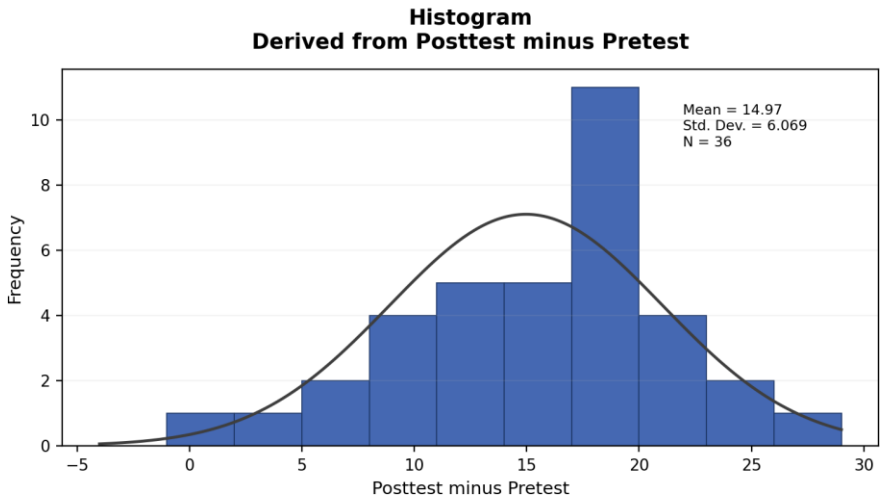


Figure 3. Histogram of posttest-minus-pretest differences.

3.4 Normality assessment of paired differences

Normality was assessed for pretest scores, posttest scores, and paired differences, summarized in Table 5 and illustrated by the difference-score Q–Q plot (Figure 4). For the paired t-test, it is the distribution of within-student differences, rather than the requirement that each set of scores separately be normal, that determines the relevant assumption.

Table 5. Tests of normality

	Kolmogorov–Smirnov[a]			Shapiro–Wilk		
Variable	Statistic	df	Sig.	Statistic	df	Sig.
Pretest	0.132	36	0.121	0.954	36	0.140
Posttest	0.154	36	0.033	0.961	36	0.229
Difference	0.131	36	0.131	0.971	36	0.467

a. Lilliefors significance correction; a = 0.05.

Shapiro–Wilk p-values were 0.140 for the pretest, 0.229 for the posttest, and 0.467 for paired differences. The corresponding Lilliefors-corrected Kolmogorov–Smirnov p-values were 0.121, 0.033, and 0.131. Although these procedures differed for the posttest, neither rejected the normality of paired differences at α = 0.05. The Q–Q plot in Figure 4 further contextualizes the formal test. Accordingly, the paired t-test was used for the primary inference and Wilcoxon testing for a complementary consistency check.

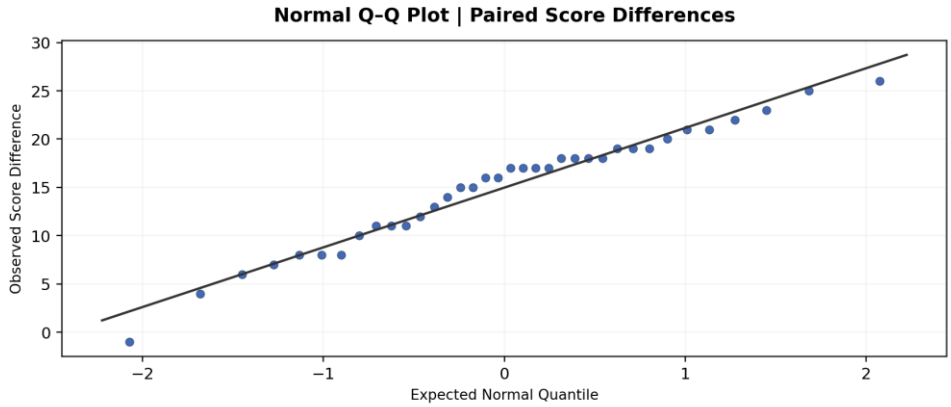


Figure 4. Normal Q–Q plot of paired score differences.

3.5 Correlations and paired-sample statistics

Table 6 confirms the pretest and posttest means and standard deviations used in the paired comparison; Table 7 reports their within-student association. The correlation describes the extent to which students retained a similar relative position between assessments, not the amount of improvement.

Table 6. Paired-samples statistics

Pair	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 Pretest	62.6389	36	9.48428	1.58071
Posttest	77.6111	36	11.20275	1.86712

Table 7. Paired-samples correlations

Pair	N	Correlation	Sig.
Pair 1 Pretest & Posttest	36	0.841	<.001

The strong positive pretest–posttest correlation ($r = 0.841$; $p < 0.001$) indicates that students with higher baseline scores generally retained relatively high scores at the second assessment. This relationship coexisted with improvement in the group: relative ranking remained comparatively stable while overall performance shifted upward. The correlation should therefore not be interpreted as a measure of improvement or evidence of identical gains across students.

3.6 Paired comparison and effect size

Table 8 uses pretest minus posttest as the paired difference, so an increase appears as a negative value. For readability, subsequent interpretation expresses the gain as posttest minus pretest. The change of sign does not alter the absolute test statistic, its significance, or the strength of evidence for a difference between occasions.

Table 8. Paired-samples test

Comparison	Mean diff.	SD (diff.)	95% CI	t (df)	p (2-tailed)
Pretest − Posttest	−14.972	6.069	[−17.026, −12.919]	−14.803 (35)	< 0.001

Note: Significance values smaller than 0.001 are reported as p < 0.001.

The paired-samples t-test found a difference between the two observations, $t(35) = -14.803$, $p < 0.001$ (Table 8). Expressed as a gain, the mean difference was 14.972 points, 95% CI [12.919, 17.026]. Cohen's $dz = 2.467$ (Table 9) quantifies the large standardized within-student difference

relative to the standard deviation of individual changes. Neither the size of the difference nor its statistical significance removes the possibility that factors other than the Rapulung-oriented learning context contributed to the increase.

Table 9. Paired-samples effect sizes

Measure	Standardizer	Point Estimate	Interpretation
Cohen's d (paired; dz)	6.06859	2.467	Effect size for paired score changes
Mean increase (Post − Pre)	–	14.972	Score-scale points

Table 10. Wilcoxon signed ranks

Posttest − Pretest	N	Mean Rank	Sum of Ranks
Negative Ranks	1	1.00	1.00
Positive Ranks	35	19.00	665.00
Ties	0	–	–
Total	36	–	–

Table 11. Wilcoxon test statistics

Indicator	Posttest − Pretest
Wilcoxon W	1.000
Z (normal approximation)	-5.219
Asymp. Sig. (2-tailed)	<.001

The Wilcoxon signed-rank test produced 35 positive ranks, one negative rank, and no ties ($Z = -5.219$, $p < 0.001$). Its agreement with the paired t-test shows that the conclusion concerning the direction of the change is not restricted to the parametric procedure. Substantively, the predominance of positive ranks reinforces the finding of widespread improvement, while the single one-point decline highlights continued individual variation.

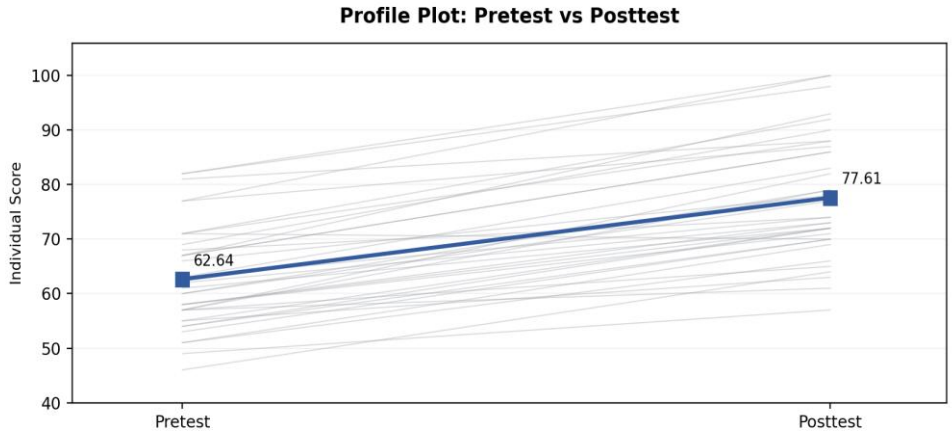


Figure 5. Individual score changes and group means.

3.7 Individual pretest and posttest scores

Table 12 reports the complete pretest–posttest pairs for S01–S36 and each student's normalized gain. Mean individual N-gain was 0.434, compared with 0.401 when calculated from the two class means, because the individual measure uses a different available-score denominator for each learner. Nine students had gains below 0.30, 22 recorded gains from 0.30 to below 0.70, and five attained 0.70 or above. The range included one small negative gain. These values are descriptive supplements to the paired tests, not separate evidence of intervention causality.

Table 12. Individual pretest and posttest scores (N = 36)

Code	Sex	Pretest	Posttest	Difference	N-Gain
S01	Female	54	70	+16	0.348
S02	Female	57	79	+22	0.512
S03	Male	67	93	+26	0.788
S04	Female	57	61	+4	0.093
S05	Male	58	72	+14	0.333
S06	Male	68	74	+6	0.188
S07	Female	63	83	+20	0.541
S08	Female	66	78	+12	0.353
S09	Female	67	86	+19	0.576
S10	Female	57	82	+25	0.581
S11	Female	51	66	+15	0.306
S12	Male	71	88	+17	0.586
S13	Female	55	72	+17	0.378
S14	Male	55	63	+8	0.178
S15	Female	67	86	+19	0.576
S16	Female	81	88	+7	0.368
S17	Male	51	69	+18	0.367
S18	Female	54	72	+18	0.391
S19	Male	49	57	+8	0.157
S20	Female	71	70	-1	-0.034
S21	Female	61	72	+11	0.282
S22	Female	77	87	+10	0.435
S23	Male	71	92	+21	0.724
S24	Male	63	74	+11	0.297
S25	Female	60	77	+17	0.425
S26	Female	82	98	+16	0.889
S27	Male	62	73	+11	0.289
S28	Male	58	73	+15	0.357
S29	Female	46	64	+18	0.333
S30	Female	69	90	+21	0.677
S31	Male	57	65	+8	0.186
S32	Female	82	100	+18	1.000
S33	Female	60	79	+19	0.475
S34	Female	77	100	+23	1.000
S35	Female	58	71	+13	0.310
S36	Female	53	70	+17	0.362

Discussion

4.1 Interpretation of change scores and social deliberation

The most informative finding is the convergence between the group-level shift and individual-level change. Mean achievement increased by 14.97 points, and 35 of the 36 paired differences were positive. These findings do not merely show that a few high achievers raised the group mean: change extended across most of the class, although its size varied markedly. In relation to the social deliberation construct, the result is consistent with better assessed reasoning about conflicting interests, but the total scores do not establish equivalent improvement in argument quality, listening, or agreement-building separately.

Cohen's dz of 2.467 indicates that the mean change substantially exceeded the variability of within-student differences. Its interpretation differs from that of a p-value: statistical significance assesses compatibility with a null hypothesis of no mean difference, whereas the effect size standardizes the magnitude of the observed change. The confidence interval for the mean gain (12.92–17.03 points) provides an estimate on the original score scale. The strength of the findings is therefore supported not only by a small p-value but also by the estimated magnitude, score dispersion, and direction of change across individual pairs.

The Rapulung framework has an identifiable theoretical connection to deliberative learning: students are encouraged to recognize affected actors, explain positions, respond to competing reasons, weigh alternatives, and justify shared decisions. Hess and McAvoy (2015) locate reason-giving and engagement with disagreement at the center of democratic classroom learning; McAvoy and Hess (2013) likewise distinguish deliberation from obtaining uniform answers. Meanwhile, Dianto (2021) situates Rapulung within collective decision-making in the Samawa setting. By bringing these perspectives together, the present study translates a local deliberative practice into an assessable learning outcome in Sociology, rather than treating local wisdom solely as cultural content.

Variation among learners also calls for more nuanced interpretation than a simple distinction between success and failure. Students with higher pretest scores have less room for absolute gain than those starting lower, so score-point increases and N-gain do not necessarily produce the same ranking. Two students reached the maximum score of 100, but this ceiling does not imply identical quality of dialogue. Likewise, a one-point decline for one learner cannot automatically be interpreted as a loss of deliberative ability without considering measurement uncertainty. Evaluation should integrate absolute change, relative gain, and evidence from the interaction process.

Sociologically, deliberation has two analytically distinct dimensions: participatory procedure and the quality of collective judgment. Equal speaking turns are insufficient if students merely repeat their group's position; conversely, rapid agreement may conceal unfairness when minority objections go unexamined. Rapulung can thus orient learning toward active listening, evaluation of decision consequences, and reflection on tensions between individual and collective interests. This interpretation aligns with McAvoy and Hess (2013), who connect classroom deliberation with responsible engagement across differing perspectives.

This alignment is especially relevant to the Sociology topic of social conflict. A response that accurately names conflicting parties may still be weak if it disregards less powerful actors or offers a solution without examining likely consequences. Rapulung supplies a culturally recognizable point of entry for these questions, while Sociology offers the analytic concepts needed to challenge facile consensus. The educational implication is therefore to evaluate the reasons supporting a decision and its fairness to affected groups, not merely the number of participants or the speed with which agreement is achieved (OECD, 2019; Hess & McAvoy, 2015).

4.2 Synthesis of findings and pedagogical implications

Several convergent results strengthen the description of within-group change: the mean and median shifted upward, the Shapiro–Wilk test did not reject normality of paired differences, the paired t-test and Wilcoxon test agreed in direction, and the standardized difference was large. The pretest–posttest correlation ($r = 0.841$) indicates that students' relative rankings were fairly stable even as their scores increased. This does not imply uniform benefits: nine students had individual N-gains below 0.30, and a single student showed a one-point decline. Examining the distribution is therefore as important as highlighting the overall mean.

For Sociology instruction, an important implication is the distinction between assessing the quality of the final decision and assessing the reasoning process that produces it. Tasks can focus on analyzing actors and interests, constructing testable arguments, responding to objections, evaluating social consequences, and reflecting on collective choices. This design provides clearer targets for deliberative learning than counting speaking turns or rewarding rapid consensus. For learners with smaller relative gains, specific feedback on the quality of reasons and responses to disagreement is more informative than labeling students solely by their final scores.

The design also has substantive limits. Repeating a test may improve familiarity with the tasks, students may mature or encounter related material outside the lesson, and other classroom experiences or teacher support may contribute to gains. Without a comparison group, these alternatives cannot be separated from Rapulung integration. In addition, the reported total scores do not reveal the validity and reliability of item-level judgments or the quality of observed interactions. Variation in task difficulty, scoring consistency, and ceiling effects for students who reached 100 may influence interpretation. These limitations qualify causal claims but do not negate the descriptive evidence of positive matched-score differences. Subsequent studies should report the duration and fidelity of instruction, instrument development, content validation, reliability, indicator-level results, and evidence from a comparable group.

Conclusion

In the 36 paired observations, mean social deliberation scores increased from 62.64 before learning to 77.61 afterward, corresponding to a 14.97-point gain. A positive change was observed for 35 students; one student's score decreased by one point. The paired t-test yielded $t(35) = -14.803$, $p < 0.001$, and the 95% confidence interval for the gain was 12.92–17.03 points (Cohen's $dz = 2.467$). The Wilcoxon finding was consistent with the parametric test, and the mean individual N-gain was 0.434. These statistics describe a substantial change within the assessed group.

Rapulung provides a theoretically relevant context for integrating reason-giving, engagement with disagreement, and collective judgment into Sociology learning about social conflict. This study contributes quantitative evidence on the distribution and magnitude of pretest–posttest change in that context. Its one-group design does not permit attribution of the change solely to Rapulung integration. Future evaluation should incorporate a comparison group, documented instructional duration, validated scoring rubrics, and observations of deliberative exchanges to establish how the different aspects of the construct develop.

Bibliography

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Routledge. https://www.routledge.com/Statistical-Power-Analysis-for-the-Behavioral-Sciences-2nd-Edition/Cohen/p/book/9780203771587

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE. https://www.sagepub.com/shop/buy-a-book/research-design-5-255675

Dianto. (2021). Hukum perencanaan pembangunan daerah berbasis Rapulung Sumbawa. JISIP (Jurnal Ilmu Sosial dan Pendidikan), 5(2). https://doi.org/10.36312/jisip.v5i2.1954

Field, A. (2018). Discovering statistics using IBM SPSS statistics (5th ed.). SAGE Publications. https://discover.library.unt.edu/catalog/b5961109

Hake, R. 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. https://doi.org/10.1119/1.18809

Hess, D. E., & McAvoy, P. (2015). The political classroom: Evidence and ethics in democratic education. Routledge. https://www.routledge.com/The-Political-Classroom-Evidence-and-Ethics-in-Democratic-Education/Hess-McAvoy/p/book/9781315738871

McAvoy, P., & Hess, D. (2013). Classroom deliberation in an era of political polarization. Curriculum Inquiry, 43(1), 14-47. https://doi.org/10.1111/curi.12000

OECD. (2019). Fostering students' creativity and critical thinking: What it means in school. OECD Publishing. https://doi.org/10.1787/62212c37-en

OECD. (2025). Unlocking high-quality teaching. OECD Publishing. https://doi.org/10.1787/f5b82176-en

Pemerintah Kabupaten Sumbawa Barat. (2021). Peraturan Daerah Kabupaten Sumbawa Barat Nomor 2 Tahun 2021 tentang Bale Mediasi. https://peraturan.bpk.go.id/Details/188770/perda-kab-sumbawa-barat-no-2-tahun-2021

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3/4), 591-611. https://doi.org/10.1093/biomet/52.3-4.591

UNESCO. (2017). Education for sustainable development goals: Learning objectives. UNESCO. https://www.unesco.org/en/articles/education-sustainable-development-goals-learning-objectives