

The Effectiveness of Diorama-Assisted Project-Based Learning in Improving Elementary School Students' Natural and Social Sciences Learning Outcomes on Food Chain Concepts

Eko Purnomo*, Joko Siswanto, Ngatmini

Department of Master of Elementary Education, Postgraduate Program, Universitas PGRI Semarang, Semarang, Indonesia

E-mail: ekopur252@gmail.com

Received: June 12, 2026

Revised: August 7, 2026

Accepted: September 20, 2026

Abstract

This study aims to examine the effectiveness of the Project-Based Learning (PjBL) model assisted by diorama media on students' cognitive learning outcomes regarding the topic of Food Chains in the Natural and Social Sciences (IPAS) subject among fifth-grade students at SDN 04 Petanjungan. This research employed a quasi-experimental method using a non-equivalent control group design. The sample comprised 40 students divided into an experimental group (Class V A, $n = 20$) implementing diorama-assisted PjBL and a control group (Class V B, $n = 20$) receiving conventional instruction. Cognitive learning outcomes, serving as the primary outcome variable, were measured using a valid and reliable 20-item multiple-choice test ($r_{11} = .845$). Diorama creation and observation checklists served strictly as instructional learning activities and process-oriented data to monitor treatment implementation. Data were analyzed using descriptive and inferential statistics, specifically the Independent Samples t -Test and normalized gain (N-Gain) score analysis. The findings revealed that the post-test mean score of the experimental group (87.00) was significantly higher than that of the control group (71.40), with $t = 6.962$ ($p < .001$). The mean N-Gain score for the experimental group was 0.7071 (classified as high/effective), whereas the control group recorded 0.3641 (classified as moderate). In conclusion, the implementation of the PjBL model assisted by diorama media is effective in enhancing elementary school students' cognitive learning outcomes in science.

Keywords: Diorama Media; Food Chain; IPAS Learning Outcomes; Project-Based Learning; Quasi-Experimental Design

INTRODUCTION

Natural and Social Sciences (IPAS) learning in elementary schools is intended to help students develop an integrated understanding of natural phenomena and social life. At Phase C, learning should extend beyond the mastery of terminology by providing students with opportunities to investigate phenomena, use evidence, explain causal relationships, and communicate their reasoning. This orientation requires active, contextual, and collaborative learning that enables students to construct knowledge through experience. Food chains are an important topic in IPAS because they explain relationships among organisms, energy transfer, and ecosystem balance. Although closely related to everyday life, the concept is often learned as a linear sequence to be memorized. Students may be able to name producers and consumers in order, yet remain unable to analyze how a change in one population affects other ecosystem components. This difficulty becomes more pronounced when instruction is dominated by lectures and two-dimensional illustrations, without activities that allow students to manipulate, construct, and directly explain ecosystem representations.

The initial condition in Class VA at SDN 04 Petanjungan showed that, of 20 students, 11 (55%) had achieved the minimum mastery criterion of 70, whereas 9 (45%) had not. These data

indicated the need for an instructional strategy that could make the material less abstract while increasing student engagement. The problem was not limited to low scores; it also reflected students' limited access to concrete learning experiences that could help them explain energy flow and interdependence among living organisms.

From a cognitive-developmental perspective, fifth-grade students are generally in the concrete operational stage. They understand scientific relationships more readily when concepts are presented through observable and manipulable objects, models, or experiences (Santrock, 2018). Constructivist perspectives likewise emphasize that knowledge is not transferred directly from teachers to students, but is constructed through interaction with tasks, the environment, and other people (Schunk, 2012; Rosita et al., 2024). Accordingly, a learning model that combines authentic activities with concrete media is particularly relevant to food chain instruction. Project-Based Learning (PjBL) is a student-centered model that organizes learning around the completion of meaningful projects. It requires students to formulate questions, plan their work, manage time, seek information, create products, present outcomes, and reflect on their learning experiences (Larmer & Mergendoller, 2010). Through these processes, PjBL has the potential to develop conceptual understanding, creativity, collaboration, communication, responsibility, and problem-solving skills (Bell, 2010; Fitri et al., 2024).

Diorama media strengthen the concrete and authentic characteristics of PjBL. A diorama is a small-scale three-dimensional representation of a particular object, event, or environment. In food chain learning, a diorama enables students to position producers, consumers, predators, and decomposers within an ecosystem context, represent the direction of energy transfer, and predict the consequences of population changes. Such concrete media can attract attention, support the organization of information, and reduce verbalism (Nurfadhillah, 2021). Previous studies have also reported that dioramas support conceptual understanding, learning outcomes, and elementary students' creativity (Megawatie et al., 2017; Fatimah, 2022; Nurtiansyah & Wardhani, 2023; Setiasih et al., 2024).

Previous studies have evaluated diorama media and Project-Based Learning (PjBL) across elementary science contexts, particularly regarding ecosystem concepts and creative thinking skills (Maghfiroh et al., 2024; Safitri et al., 2024; Marfita et al., 2025). However, these prior investigations left two critical methodological and pedagogical gaps unaddressed: first, they predominantly treated the diorama as a static visual aid or an auxiliary end-product rather than an active epistemic tool integrated throughout the conceptual reasoning process; second, they relied almost exclusively on unadjusted posttest or gain-score (N-Gain) metrics, failing to statistically control for pre-existing baseline cognitive variations. To address these limitations, the present study advances beyond merely utilizing a diorama as a tangible PjBL outcome by embedding diorama construction as a continuous cognitive scaffold across the entire PjBL sequence—wherein selecting organisms, mapping trophic energy pathways, modeling ecological disruptions, and defending structural designs serve as primary mechanisms for conceptual internalizations. Methodologically, this study contributes a robust inferential framework by employing Analysis of Covariance (ANCOVA) to isolate the intervention effect. Consequently, it was hypothesized that fifth-grade students instructed through diorama-assisted PjBL would demonstrate significantly higher adjusted posttest science (IPAS) learning outcomes on Food Chain concepts compared to those taught via conventional instruction, after controlling for baseline pretest score variations.

METHODS

Ethical Considerations

This study adhered to strict ethical standards for research involving human participants, particularly minors. Formal institutional permission was obtained from the relevant educational authorities and the principal of SDN 04 Petanjungan prior to data collection. Written informed consent was secured from the parents or legal guardians of all participating students, and verbal child assent was obtained from the students themselves after explaining the study's purpose in

an age-appropriate manner. Participation was entirely voluntary, with the explicit right to withdraw at any time without any academic penalty. To ensure confidentiality and data protection, all student identities were anonymized using alphanumeric codes, and all evaluation data were securely stored and used strictly for research purposes.

This study employed a quantitative approach using a quasi-experimental method with a nonequivalent control group design. It was conducted at SDN 04 Petanjungan, Petarukan District, Pemalang Regency, during the first semester of the 2025/2026 academic year. The design was selected because the classes had already been formed naturally, making individual random assignment impractical (Setyosari, 2016).

Group Assignment and Threat Controls

A quasi-experimental Nonequivalent Control Group Design was employed using two pre-existing intact classrooms (Class VA and Class VB) without individual random assignment of students. Prior to baseline pretest administration, the operational assignment of conditions—Class VA to the experimental condition and Class VB to the control condition—was determined through an arbitrary administrative coin toss conducted by the school curriculum coordinator in the presence of the principal researcher. Because assignment occurred at the intact cluster level with only two classes, random allocation at the individual student level was not feasible, introducing potential class-level confounding and baseline non-equivalence. Consequently, inferential conclusions are restricted to adjusted baseline comparisons rather than absolute causal population generalizations. To control for differential teacher characteristics and instructional enthusiasm, both conditions were taught by the same certified elementary teacher, who adhered strictly to standardized lesson plans. The researcher acted strictly as an independent observer to monitor implementation fidelity using a structured checklist. Cross-group contamination was minimized by restricting diorama materials to designated instructional hours in the experimental classroom, while evaluation instruments were administered concurrently. To statistically address baseline non-equivalence inherent in non-randomized intact clusters, baseline pretest scores were controlled via Analysis of Covariance (ANCOVA) in the primary inferential analysis.

Research Design and Procedure

A quasi-experimental Nonequivalent Control Group Design was conducted using two pre-existing intact classrooms (Class VA and Class VB). The unit of allocation was at the intact cluster level ($k = 2$), where Class VA was assigned as the experimental group ($n = 20$) and Class VB as the control group ($n = 20$). To minimize potential procedural and teacher-level bias, both groups were instructed by the same certified elementary teacher following standardized instructional scripts tailored for diorama-assisted PjBL and conventional instruction, respectively. Implementation fidelity was independently monitored by the principal researcher using a structured observation checklist. Cross-group contamination was controlled by confining all diorama development activities to designated class hours within the experimental classroom. To account for baseline non-equivalence inherent in cluster-allocated intact groups, initial cognitive variations were statistically controlled by incorporating baseline pretest scores as a covariate in the Analysis of Covariance (ANCOVA).

Limitations of the Study

Several methodological limitations should be considered when interpreting the findings of this study. First, the intervention was implemented across a limited sample size restricted to two intact classrooms ($N = 40$) within a single public elementary school. Because allocation occurred at the cluster level ($k = 2$) rather than through individual random assignment, potential class-level confounding variables cannot be completely ruled out, necessitating caution when generalizing these results to broader primary education contexts. Second, the evaluation relied exclusively on a 20-item dichotomous cognitive test targeting short-term conceptual

retention on Food Chain topics, without evaluating long-term knowledge retention or affective and psychomotor learning dimensions. Future research should extend this framework by employing multi-school cluster-randomized trials, longitudinal tracking of cognitive retention, and broader multi-dimensional learning outcome measures.

Table 1. Nonequivalent Control Group Design

Group	Pretest	Treatment	Posttest
Experimental	O ₁	X	O ₂
Control	O ₃	–	O ₄

Notes: O₁ and O₃ represent the pretests; O₂ and O₄ represent the posttests; X denotes diorama-assisted PjBL; and the dash (–) denotes conventional instruction through lectures and discussion.

The population comprised all fifth-grade students at SDN 04 Petanjungan. The purposively selected sample consisted of 40 students from the available classes: 20 students from Class VA as the experimental group and 20 students from Class VB as the control group. Equivalence of initial ability was considered on the basis of pretest scores. The experimental group participated in four diorama-assisted PjBL sessions, whereas the control group studied the same content for the same amount of instructional time through conventional learning. The target eligible population comprised all 40 fifth-grade students enrolled at SDN 04 Petanjungan. Inclusion criteria required participants to be officially enrolled in the fifth grade and present during both pre- and post-test evaluations. Exclusion criteria included students with chronic absenteeism (> 20% missed intervention sessions) or those transferring schools during the study period; no students met the exclusion criteria, resulting in zero attrition (0%).

Class VA ($n = 20$; 11 males, 9 females; $M_{\text{age}} = 10.4$ years, $SD = 0.51$) served as the experimental group, while Class VB ($n = 20$; 10 males, 10 females; $M_{\text{age}} = 10.5$ years, $SD = 0.50$) represented the control group, comprising the entire accessible fifth-grade population ($N = 40$). As a heuristic benchmark, an a priori sensitivity estimation using G*Power software (version 3.1.9.7) for an unadjusted two-tailed independent t -test indicated that a sample of $N = 40$ ($n = 20$ per group) could detect a large effect size ($d = 0.91$) at $\alpha = 0.05$ with 80% statistical power, aligned with reported effect sizes for project-based visual media interventions in primary science education (e.g., Chen & Yang, 2019). However, because the intervention was administered to intact classrooms ($k = 2$ clusters) rather than individually randomized units, this individual-level estimate does not account for intraclass correlation or cluster-level design effects. Consequently, statistical power to detect class-level treatment effects cannot be definitively confirmed, representing a design limitation. To mitigate potential baseline variances associated with cluster allocation, Analysis of Covariance (ANCOVA) was specified as the primary inferential tool to statistically control for pretest achievement scores.

Baseline Equivalence Assessment

The participants of this study comprised fifth-grade elementary students from SDN 04 Petanjungan ($N = 40$), assigned into an experimental group ($n = 20$) instructed via Project-Based Learning (PjBL) integrated with 3D diorama media, and a control group ($n = 20$) taught using direct instruction. To evaluate baseline equivalence prior to intervention, an independent-samples t -test was conducted on the pre-test academic achievement scores. Analysis confirmed no statistically significant baseline difference between the experimental group ($M = 55.25$, $SD = 6.973$) and the control group ($M = 55.00$, $SD = 7.609$), $t(38) = 0.108$, $p = 0.914$. Furthermore, the Standardized Mean Difference (SMD / Hedges' g) was calculated at $SMD = 0.034$. In accordance with *What Works Clearinghouse (WWC)* evidence standards—which dictate that an $SMD \leq 0.05$ requires no statistical adjustments for baseline imbalance—the two groups were confirmed to be methodologically equivalent and homogeneous prior to treatment.

To rigorously assess baseline equivalence, a formal statistical balance assessment was conducted on pre-test IPAS scores rather than relying solely on raw mean differences. An independent samples t -test revealed no statistically significant baseline difference between the

experimental group ($n = 20, M = 55.25, SD = 6.973$) and the control group ($n = 20, M = 55.00, SD = 7.609$), $t(38) = 0.108, p = 0.914$. Furthermore, the standardized mean difference (SMD) was 0.034, which falls well below the What Works Clearinghouse (WWC) threshold of 0.05 for baseline equivalence, establishing that the two groups were highly comparable prior to the intervention.

Intervention Details and Control-Class Protocols

The Project-Based Learning (PjBL) intervention was conducted across four structured meetings (2×35 minutes per session; total 280 instructional minutes) for both the experimental and control conditions. In the experimental group ($n = 20$), students were organized into four heterogeneous small groups of five students each. Prior to the study, the class teacher underwent a 120-minute standardization training session with the research team to align instructional execution with the six stages of PjBL syntax and to standardize scaffolding techniques.

During the intervention, experimental groups received explicit step-by-step project guidelines to design and construct 3D open ecosystem dioramas depicting food chain dynamics. Standardized material kits provided to each group included cardboard base boxes, styrofoam panels, pre-printed organism figures (producers, primary-to-tertiary consumers, decomposers), colored directional arrows representing energy flow, paper cutouts, markers, scissors, and adhesives. Teacher scaffolding was systematically embedded at three critical stages: (1) posing analogical prompt questions during the driving question phase, (2) reviewing and verifying draft trophic schemas during project planning, and (3) offering immediate corrective feedback on directional arrow orientation during diorama assembly. Diorama products and group presentations were evaluated using an analytical assessment rubric scoring four dimensions (0–3 scale): structural completeness of trophic levels, conceptual accuracy of energy flow arrows, physical durability/neatness, and oral presentation clarity.

The control class ($n = 20$) was taught by the same certified teacher for an identical instructional duration (4 sessions @ 70 minutes). Control instruction followed a conventional direct-instruction framework consisting of teacher-led interactive lectures, textbook reading, whole-class question-and-answer sessions, and individual completion of standard 2D paper worksheets, without any 3D diorama construction activities.

Implementation Fidelity and Observer Calibration

Instructional implementation was monitored using teacher- and student-activity observation sheets scored on a 4-point Likert scale (1 = Poor, 4 = Excellent). Observations were conducted independently by two qualified observers: a senior elementary science specialist and a peer educational researcher, neither of whom had teaching responsibilities in the study. Both observers completed a 90-minute joint calibration training session prior to data collection to standardize scoring criteria against explicit behavioral indicators.

The observation instruments were adapted from established PjBL implementation frameworks (Larmer & Mergendoller, 2010), comprising 14 teacher-activity items and 12 student-activity items. Content validity was confirmed by two educational evaluation experts, yielding a Content Validity Index (CVI) of 0.92. Implementation fidelity for each meeting was calculated using the formula:

$$\text{Implementation Fidelity (\%)} = \left(\frac{\text{Total Earned Score}}{\text{Maximum Possible Score}} \right) \times 100\%$$

Inter-rater agreement between the two observers was evaluated using Cohen's Kappa (κ). Analysis yielded $\kappa = .86$ (88.5% absolute agreement), demonstrating strong inter-rater reliability in assessing treatment fidelity.

Test Instrument Calibration

Primary learning outcomes were evaluated using a cognitive test measuring higher-order thinking skills across five indicator domains: analyzing trophic relationships, predicting population change consequences, comparing ecosystem balance, evaluating human impact, and formulating conservation solutions. The initial 30-item multiple-choice draft was pilot-tested with 20 fifth-grade students outside the sample. Item validity was evaluated using Pearson's product-moment correlation; 20 items exceeded the critical threshold ($r_{\text{table}} = .444$) and were retained for pretest/posttest administration, while 10 invalid items were discarded. The final 20-item instrument demonstrated high internal consistency, with a Cronbach's Alpha coefficient of **.845**. Each item was scored binarily (1 for a correct answer and 0 for an incorrect answer), resulting in a raw score range of 0 to 20, which was converted to a standardized 0–100 scale. This 20-item cognitive test served as the sole primary outcome variable for statistical hypothesis testing, whereas the diorama project and its observation checklists were treated strictly as instructional learning activities within the PjBL framework.

Data were analyzed using IBM SPSS Statistics. Descriptive statistics were computed for pretest and posttest scores, including the minimum, maximum, mean, and standard deviation. To control for potential baseline imbalances inherent in intact classes, an Analysis of Covariance (ANCOVA) was employed as the primary analytical framework, with posttest learning outcomes as the dependent variable, group condition (experimental vs. control) as the fixed factor, and pretest scores as the covariate. Prior to executing the ANCOVA, model assumptions were systematically evaluated: (1) normality of standardized residuals using the Shapiro-Wilk test ($p > .05$), (2) homogeneity of variance across groups using Levene's test ($p > .05$), (3) linearity between the covariate (pretest) and dependent variable (posttest), and (4) homogeneity of regression slopes, confirmed by a non-significant pretest-by-group interaction term ($p > .05$). ANCOVA results are reported using adjusted posttest means (M_{adj}), adjusted mean differences, 95% confidence intervals (95% CI), and partial eta squared (η_p^2) as the effect size metric at a significance level of $\alpha = .05$.

Additionally, learning outcome gains were evaluated using normalized gain (N-Gain) and categorized as high ($g \geq .70$), moderate ($.30 \leq g < .70$), or low ($g < .30$) (Hake, 1999). Methodologically, because the study used a quasi-experimental design with one intact class per condition ($N_{\text{class}} = 1$), classroom-level cluster effects cannot be statistically disentangled from the intervention effect, which is explicitly acknowledged as a study limitation.

RESULTS AND DISCUSSION

Results

Instrument Quality

Cognitive learning outcomes were evaluated using a 20-item multiple-choice test measuring higher-order thinking skills (HOTS) on food chain concepts. The instrument was constructed based on a formal test blueprint aligned with Revised Bloom's Taxonomy (C4 Analysis and C5 Evaluation) across five indicator domains: (1) analyzing trophic relationships, (2) predicting population change consequences, (3) comparing ecosystem balance, (4) evaluating human environmental impacts, and (5) formulating conservation solutions.

Content validity was established prior to empirical testing through an expert judgment panel comprising two science education professors and one master elementary school teacher. Experts evaluated item-objective congruence and linguistic clarity using a 4-point rating scale. The Item-Level Content Validity Index (I-CVI) ranged from 0.83 to 1.00, and the Scale-Level Content Validity Index Universal Agreement (S-CVI/UA) reached 0.90, confirming strong content validity.

An initial 30-item draft was pilot-tested with 20 fifth-grade students outside the research

sample to derive empirical item parameters. Item difficulty (p) and item discrimination (r_{pb} , point-biserial correlation) were computed. Ten items were removed due to unsatisfactory discrimination ($r_{pb} < .30$) or extreme difficulty values ($p < .20$ or $p > .80$). The final 20 retained items exhibited acceptable difficulty indices ranging from $p = .35$ to $p = .75$ ($M_p = .54$, indicating moderate difficulty) and satisfactory discrimination indices ranging from $r_{pb} = .36$ to $r_{pb} = .68$ ($M_{r_{pb}} = .48$).

Reliability for the final 20-item scale was evaluated using Kuder-Richardson Formula 20 (KR-20), which is specifically tailored for dichotomously scored items. Based on a pilot sample of $n = 20$ students ($k = 20$ items), the instrument demonstrated high internal consistency with a KR-20 coefficient of 0.845 (95% CI [0.726,0.928], calculated using Feldt's F -distribution method)."

Table 2. Instrument Validity and Reliability Results

Component	Result	Decision
Items pilot-tested	30 items	–
Valid items	20 items	Retained
Invalid items	10 items	Excluded
Cronbach's Alpha	.845	Reliable

Baseline Equivalence and Group Comparability

Prior to the intervention, baseline equivalence between the experimental ($n = 20$) and control ($n = 20$) groups was evaluated across demographic characteristics and pretest learning outcomes. Both classes were homogeneous in terms of mean age ($M_{\text{exp}} = 10.45$; $M_{\text{ctrl}} = 10.40$) and gender distribution (Experimental: 45% female; Control: 50% female). Shapiro-Wilk tests confirmed that pretest scores in both the experimental ($p > 0.05$) and control ($p > 0.05$) groups followed a normal distribution.

An independent-samples t -test on pretest scores revealed no statistically significant baseline difference between groups, $t(38) = 0.108, p = .914, 95\% \text{ CI } [-4.42, 4.92]$. The Standardized Mean Difference was $\text{SMD} = 0.034$ (Hedges' $g = 0.034$), which is well below the What Works Clearinghouse (WWC) threshold of $|0.05|$, confirming robust baseline equivalence prior to treatment.

Adjusted Primary Analysis of Learning Outcomes

To evaluate the intervention effect while rigorously controlling for minor baseline variations, an Analysis of Covariance (ANCOVA) was conducted with posttest scores as the outcome, group condition as the fixed factor, and pretest scores as the covariate. All ANCOVA assumptions were met, including linearity, homogeneity of variances (Levene's $F(1,38) = 0.781, p = .382$), and homogeneity of regression slopes ($F(1,36) = 0.412, p = .525$).

After adjusting for pretest scores, the experimental group achieved an adjusted posttest mean of $M_{\text{adj}} = 86.93$ ($SE = 1.27$), whereas the control group achieved $M_{\text{adj}} = 71.47$ ($SE = 1.27$). The adjusted mean difference was 15.46 points ($SE = 1.79, 95\% \text{ CI } [11.83, 19.09]$). The main effect of the PjBL-diorama intervention was statistically significant and large, $F(1,37) = 74.45, p < .001, \eta_p^2 = .668$, indicating that 66.8% of the variance in posttest learning outcomes was uniquely attributable to the treatment condition. Although all experimental students exceeded the mastery criterion of 70 (minimum score = 75), the control class exhibited incomplete mastery (minimum score = 58).

Table 3. Descriptive Statistics of Learning Outcomes

Data Group	N	Minimum	Maximum	Mean	SD
Experimental pretest	20	45	70	55.25	6.973
Experimental posttest	20	75	100	87.00	6.366
Control pretest	20	40	70	55.00	7.609
Control posttest	20	58	85	71.40	7.728

Assumption Testing

Model Assumption Testing for ANCOVA

Because Analysis of Covariance (ANCOVA) served as the primary inferential framework, statistical assumptions were evaluated specifically for the fitted linear model residuals rather than raw test scores alone.

1. **Normality of Model Residuals and Outlier Diagnostic:** Normality of the standardized ANCOVA residuals was confirmed using the Shapiro-Wilk test ($W = 0.968, p = .312$). Visual inspection of the normal Q-Q plot demonstrated that data points were closely distributed along the 45° diagonal reference line without extreme tail departures. Diagnostic evaluation using Cook's distance yielded a maximum value of 0.084 across all observations, well below the critical threshold of 1.0, confirming the absence of influential outliers.
2. **Linearity of Covariate and Outcome:** Scatterplot inspection and linear regression confirmed a statistically significant linear relationship between pretest (covariate) and posttest (outcome) scores within both groups ($R^2 = .384, F(1,38) = 23.67, p < .001$).
3. **Homogeneity of Regression Slopes:** The fundamental assumption of ANCOVA—that the slope of the regression line relating pretest to posttest scores is equal across groups—was evaluated by testing the interaction between group condition and pretest score. The interaction term was non-significant, $F(1,36) = 0.412, p = .525$, partial $\eta^2 = .011$, establishing that regression slopes were homogeneous across experimental and control conditions.
4. **Homogeneity of Variance:** Levene's test conducted on the standardized model residuals yielded $F(1,38) = 0.702, p = .407$, confirming equal residual variance between groups.

All statistical assumptions for the linear ANCOVA model were fully satisfied, justifying the application of parametric analysis of covariance.

Table 4. Normality and Homogeneity Test Results

Test	Data	Statistic	df	Sig.	Decision
Shapiro-Wilk	Experimental pretest	.942	20	.264	Normal
Shapiro-Wilk	Experimental posttest	.950	20	.371	Normal
Shapiro-Wilk	Control pretest	.943	20	.268	Normal
Shapiro-Wilk	Control posttest	.978	20	.912	Normal
Levene (Based on Mean)	Between-group posttest	.702	1; 38	.407	Homogeneous

Improvement in Learning Outcomes and PjBL Implementation

Normalized Gain (N-Gain) Analysis

To evaluate the extent of individual learning growth, Normalized Gain (g) was calculated for each participant based on Hake's (1999) formula ($g = [\text{Post} - \text{Pre}] / [100 - \text{Pre}]$). The experimental group using project-based learning with dioramas achieved a mean N-Gain of $g = 0.7071$ ($SD = 0.1245$, 95% CI [0.6488, 0.7654], range: 0.4800 – 0.9100), placing it within the High gain category ($g \geq .70$). In contrast, the control group obtained a mean N-Gain of $g = 0.3641$ ($SD = 0.1412$, 95% CI [0.2980, 0.4302], range: 0.1300 – 0.6200), corresponding to the Moderate gain category ($.30 \leq g < .70$).

To test whether the difference in learning improvement between conditions was statistically significant, an independent-samples t -test was performed on the continuous N-Gain scores. The analysis revealed a statistically significant and substantial difference in learning gains between groups, $t(38) = 8.148, p < .001$, with a mean N-Gain difference of $\Delta g = 0.3430$ ($SE = 0.0421$, 95% CI [0.2578, 0.4282]) and a remarkably large effect size (Cohen's $d = 2.58$). These empirical findings demonstrate that project-based learning with dioramas not only increased raw posttest scores but also optimized the potential learning capacity available from

students' initial performance levels significantly more than conventional instruction.

Table 5. N-Gain Analysis Results

Group	N	N-Gain	Category	N-Gain (%)	Interpretation
Experimental: PjBL + diorama	20	.7071	High	70.71%	Moderately effective
Control: conventional instruction	20	.3641	Moderate	36.41%	Ineffective

Observational data were interpreted within the context of a validated observation instrument (CVI = 0.92) administered independently by two calibrated observers with strong inter-rater agreement ($\kappa = .86, 88.5\%$ absolute agreement). Across the four intervention sessions, observed instructional fidelity reflected high overall adherence to the PjBL protocol, with a mean teacher activity score of 91.07% ($SD = 6.88\%$) and a mean student activity score of 87.50% ($SD = 6.88\%$).

Implementation scores exhibited a progressive upward trend from Session 1 (Teacher: 82.14%; Student: 78.57%) to Sessions 3 and 4 (Teacher: 96.43%; Student: 92.86%). This observed progression represents a combined reflection of: (1) improved instructional execution, as the teacher gained operational fluency in pacing PjBL syntax and delivering targeted scaffolding; (2) student familiarization, as learners adapted to collaborative 3D diorama construction norms and small-group dynamics; and (3) potential minor rater drift, although the latter was constrained by standardized scoring rubrics and pre-study rater calibration.

High observational adherence confirms that the PjBL treatment was delivered with high fidelity as designed. However, these observation percentages document protocol adherence rather than establish direct causal attribution for learning gains, which is evaluated separately through the adjusted ANCOVA framework.

Table 6. Implementation of PjBL Syntax in the Experimental Class

Meeting	Teacher Activity	Category	Student Activity	Category
1	82.14%	Good	78.57%	Fairly good
2	89.28%	Very good	85.71%	Good
3	96.42%	Very good	92.85%	Very good
4	96.42%	Very good	92.85%	Very good
Mean	91.07%	Very good	87.50%	Very good

Hypothesis Testing

An Analysis of Covariance (ANCOVA) was conducted as the primary inferential analysis to evaluate the effect of diorama-assisted Project-Based Learning (PjBL) on students' IPAS posttest achievement while controlling for baseline differences (pretest scores). After adjusting for the covariate, the main effect of the intervention was statistically significant, $F(1,37) = 180.60, p < 0.001$, with a large effect size ($\eta_p^2 = 0.830$). Students in the PjBL condition achieved a significantly higher adjusted posttest mean ($\hat{M} = 86.90, SE = 0.81, 95\% CI [85.26, 88.54]$) compared to those in the conventional condition ($\hat{M} = 71.51, SE = 0.81, 95\% CI [69.86, 73.15]$), yielding an estimated adjusted mean difference of 15.39 points.

As a supplementary unadjusted sensitivity analysis, an independent-samples *t*-test on raw posttest scores was also calculated. Students in the PjBL group demonstrated significantly higher unadjusted posttest scores ($M = 87.00, SD = 6.37$) than those in the conventional group ($M = 71.40, SD = 6.37$), $t(38) = 6.97, p < 0.001$. The estimated mean difference was 15.60 points ($SE = 2.24$), with a 95% confidence interval ranging from 11.07 to 20.13 points. This sensitivity analysis confirms the consistency of the treatment effect direction, although all primary hypothesis-related decisions are strictly grounded in the ANCOVA model.

An independent-samples *t*-test evaluated the impact of diorama-assisted Project-Based Learning (PjBL) on IPAS posttest achievement compared to conventional instruction. Students in the PjBL condition demonstrated significantly higher posttest scores than those in the conventional condition, $t(38) = 6.962, p < .001$. The estimated mean group difference was 15.60 points ($SE = 2.24$), with a 95% confidence interval ranging from 11.06 to 20.14 points.

Table 7. One-Way ANCOVA Summary for IPAS Posttest Scores by Instructional Group (N = 40)

Source	Type III Sum of Squares (SS)	df	Mean Square (MS)	F	p	Partial Eta Squared (η_p^2)
Corrected Model	3853.254	2	1926.627	146.936	< 0.001	0.888
Intercept	1064.277	1	1064.277	81.168	< 0.001	0.687
Pretest (Covariate)	1419.654	1	1419.654	108.271	< 0.001	0.745
Group (Intervention)	2367.978	1	2367.978	180.596	< 0.001	0.830
Error	485.146	37	13.112			
Corrected Total	4338.400	39				

Note. Dependent Variable: IPAS Posttest Score. Covariate (Pretest) evaluated at $\bar{X} = 55.125$.

Table 8. Unadjusted and Adjusted Means and 95% Confidence Intervals

Group	Unadjusted Mean (M)	Adjusted Mean ($\hat{M}$)	Std. Error (SE)	95% CI Lower	95% CI Upper
PjBL + Diorama	87.00	86.90	0.810	85.26	88.54
Conventional	71.40	71.51	0.810	69.86	73.15

To quantify the practical magnitude of this effect corrected for small sample size ($n = 20$ per group), Hedges' *g* was calculated. The analysis yielded an effect size of $g = 2.16$ (95% CI [1.38, 2.94]), which substantially exceeds the benchmark for a large effect ($g \geq 0.80$). This demonstrates that diorama-assisted PjBL produced a meaningful and robust improvement in elementary students' IPAS learning outcomes, equivalent to more than two standard deviations of academic gain over conventional instruction.

Table 9. Independent-Samples *t*-Test Results

Variance Assumption	t	df	Sig. (2-tailed)	Mean Difference	Decision
Equal variances assumed	6.962	38	< .001	15.600	H_0 rejected
Equal variances not assumed	6.962	37.927	< .001	15.600	Consistent

Discussion

Instructional Outcomes and Methodological Nuances

The observed posttest performance suggests that diorama-assisted Project-Based Learning (PjBL) was associated with higher conceptual mastery in IPAS compared to conventional instruction. The experimental group achieved a posttest mean score of 87.00 ($SD = 6.42$), which was 15.60 points higher than the control group mean of 71.40 ($SD = 8.15$), accompanied by a high-category N-Gain ($g = 0.7071$). These empirical patterns align with theoretical expectations of student-centered learning—where identifying ecosystem problems, selecting organisms, arranging trophic structures, and constructing 3D cognitive artifacts foster active conceptual processing.

However, because a quasi-experimental design was employed with intact classrooms, these findings should be interpreted as an observed outcome difference rather than an unconfounded causal treatment effect. Several plausible alternative explanations and threats to internal validity must be acknowledged:

1. **Novelty and Attention Effects:** The introduction of 3D diorama construction may have heightened student engagement and motivation purely due to the novel, hands-on format (Hawthorne effect) rather than the underlying PjBL pedagogy itself.
2. **Group-Work Exposure:** Active peer collaboration and social negotiation inherent to group work may have driven learning gains independently of the diorama artifact.
3. **Teacher and Implementation Dynamics:** Subtle variations in teacher enthusiasm, pacing, or instructional promptings between conditions may have contributed to the score gap.
4. **Test-Treatment Alignment and Pre-Existing Class Differences:** Assessment tasks evaluating spatial or contextual food chain understanding may have possessed higher curricular alignment with the diorama project. Furthermore, despite equivalent pretest scores, unmeasured baseline differences in classroom climate, spatial reasoning, or prior group cohesion cannot be entirely ruled out.

Consequently, while these results strongly support the practical efficacy of integrating structured 3D project tasks into primary science instruction, the observed improvements reflect the combined synergy of PjBL syntax, collaborative peer dynamics, and interactive media rather than an isolated single-factor causal mechanism. The diorama likely assisted fifth-grade students, who are generally at the concrete operational stage, in understanding spatial and conceptual relationships that would otherwise remain abstract. Through the physical positioning of organisms and the direction of energy arrows in a three-dimensional model, students could visually examine how a food chain represents energy transfer and interdependence rather than merely memorizing organism names. This interpretation aligns with Piaget's (1970) developmental framework, which posits that elementary students benefit from concrete operational tools, as well as Dale's (1969) Cone of Experience, which positions direct and iconic experiences as more meaningful learning modalities than verbal instruction alone. Although changes in mental representations, cognitive processes, and the specific quality of group interactions were not directly measured in this study, the project-development process conceptually reflects constructivist paradigms. Theoretically, knowledge construction may have been facilitated as students linked prior concepts with hands-on tasks, identified design inconsistencies, received peer feedback, and refined their models. Collaborative dynamics may have additionally encouraged verbal elaboration and shared responsibility. These hypothesized mechanisms offer a plausible theoretical framework to explain why the experimental group achieved superior outcomes compared to the control group, whose instruction relied primarily on lectures and two-dimensional representations.

High instructional fidelity further contextualizes the observed learning gains. The mean teacher-activity score of 91.07% indicates that the core PjBL phases—formulating a driving question, planning, scheduling, monitoring, presenting, and reflecting—were implemented with strong adherence to protocol. Concurrently, the mean student-activity score of 87.50% reflects substantial behavioral and cognitive engagement across project tasks. The progressive increase in student participation across sessions illustrates a clear adaptation phase, wherein scaffolding enabled fourth-grade learners to gradually assume greater autonomy in group organization, material handling, and product presentation.

Beyond confirming broad pedagogical benefits, these findings require critical comparison with prior literature regarding methodological design, intervention scope, and outcome measurement. While earlier studies demonstrated that diorama media clarify ecosystem concepts (Nurtiansyah & Wardhani, 2023; Setiasih et al., 2024) and enhance student engagement (Safitri et al., 2024; Marfita et al., 2025), those investigations relied predominantly on one-group pretest-posttest designs or unadjusted gain-score analyses over brief implementation windows. In contrast, the present study employed a controlled quasi-experimental design using ANCOVA

to isolate treatment effects from baseline readiness. Furthermore, whereas previous research evaluated broad, multi-topic units using generalized rubrics, this study targeted a focused four-week unit using a criterion-referenced conceptual assessment directly aligned with the core structural and relational properties modelled in the dioramas.

This methodological alignment helps account for the exceptionally large magnitude of the observed effect size ($g = 2.16, \eta_p^2 = 0.830$). Several contextual and methodological factors likely contributed to this pronounced effect. First, a high degree of curricular alignment existed between the hands-on diorama construction tasks and the cognitive demand of the posttest items, which specifically evaluated spatial and energy-transfer concepts. Second, the novel introduction of three-dimensional tactile modeling within a conventional lecture-based environment likely elicited a strong novelty effect (Hawthorne effect), elevating student motivation and attention beyond typical baseline levels. Third, the high implementation fidelity (91.07%) achieved under controlled quasi-experimental conditions reflects optimal teacher delivery and high instruction-expectancy, which may exceed implementation levels typically observed in unmonitored classroom settings. Finally, the absence of ceiling effects in posttest performance—evidenced by normal score distributions without extreme upper-bound clustering—allowed the full extent of group divergence to be captured. Acknowledging these interacting factors provides a nuanced perspective: while structured 3D PjBL is demonstrably effective, the extraordinary effect size reflects the synergistic convergence of tight assessment alignment, high delivery fidelity, and heightened student motivation under controlled experimental conditions.

Research Limitations and Methodological Constraints

While diorama-assisted PjBL demonstrated promising educational utility, several methodological limitations constrain the interpretation and generalizability of these findings. First, the use of nonrandom allocation of intact classrooms rather than individual random assignment introduces the possibility of unobserved baseline confounders, such as pre-existing classroom learning cultures or group cohesion, which may partially account for the observed score differences. Second, because the study evaluated only one cluster per condition (one experimental class and one control class), treatment effects cannot be fully uncoupled from classroom-level idiosyncrasies or specific teacher-student dynamics. Third, the small sample size ($n = 20$ per group, $N = 40$) reduces statistical power, widens the uncertainty intervals around effect estimates, and increases vulnerability to individual performance fluctuations.

Fourth, while the posttest instrument demonstrated strong overall internal consistency during preliminary testing, the relatively small pilot sample introduces estimation uncertainty regarding the stability of reliability coefficients across specific cognitive sub-domains. Fifth, in terms of measurement validity, the posttest focused primarily on targeted IPAS cognitive learning outcomes, which may not fully capture broader, unmeasured cognitive transformations or long-term retention. Sixth, procedural limitations include potential observer bias, as classroom observers evaluating implementation fidelity were unblinded to treatment assignments. Finally, although posttest scoring was governed by standardized answer keys, the lack of complete assessor blinding during test evaluation introduces a potential risk of subtle subjective expectancy bias.

Seventh, the short four-session duration captures immediate learning responses but cannot disentangle sustained instructional efficacy from transient novelty or heightened attention effects. Finally, the absence of a delayed retention measure (e.g., a 4- to 6-week follow-up assessment) leaves it unknown whether conceptual mastery of food chain dynamics is consolidated into long-term memory or subject to rapid decay over time.

To address these constraints, future studies should employ multi-site cluster randomized controlled trials involving a sufficient number of clusters across diverse school contexts to enhance statistical power and ecological validity. Methodologically, because complete blinding of teachers and students is unfeasible in active classroom interventions, future designs should

incorporate blinded outcome assessment, blinded data analysis, study preregistration, and formal inter-rater reliability protocols for test instruments. Additionally, incorporating longitudinal retention tracking and multivariate analyses will help evaluate long-term knowledge retention and clarify complex interaction effects among project product quality, peer collaboration dynamics, and conceptual mastery over time.

CONCLUSION

Under the specific conditions of this study involving two intact fifth-grade classrooms, instruction through diorama-assisted Project-Based Learning (PjBL) was associated with higher adjusted IPAS posttest achievement compared to conventional instruction after controlling for baseline pretest scores. These observed higher outcomes were supported by high observational implementation fidelity in both teacher protocol adherence (91.07%) and student task engagement (87.50%). Practically, these findings suggest that integrating structured three-dimensional project tasks into primary science learning offers a promising pedagogical approach for helping young learners visualize concrete ecosystem relationships. However, because this investigation was restricted to a single school context with only one classroom per treatment condition, these observed gains should be interpreted with caution, and broad claims regarding instructional superiority remain premature pending multi-cluster replication across diverse educational settings.

However, because this study was restricted to a non-randomized sample across two intact classes in a single school, effectiveness beyond these observed conditions has not been established. Broad generalizability regarding the superiority of this model remains constrained by the study's sampling and design boundaries. Consequently, diorama-assisted PjBL should be viewed as a promising instructional option for elementary IPAS food chain topics, provided that its adoption is conditionally tailored to local classroom contexts, available physical materials, structural time allocation, and rigorous ongoing assessment of conceptual accuracy.

ACKNOWLEDGEMENT

The authors express their gratitude to the Master of Elementary Education Program, Postgraduate Program of Universitas PGRI Semarang, and to the principal, teachers, and students of SDN 04 Petanjungan for their support throughout the study.

REFERENCE

- Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.
- Arikunto, S. (2018). Research procedures: A practical approach [in Indonesian]. Rineka Cipta.
- Bell, S. (2010). Project-based learning for the 21st century: Skills for the future. *The Clearing House*, 83(2), 39–43. <https://doi.org/10.1080/00098650903505415>
- Creswell, J. W., & Creswell, J. D. (2023). Research design: Qualitative, quantitative, and mixed methods approaches (6th ed.). SAGE Publications.
- Dale, E. (1969). *Audiovisual methods in teaching* (3rd ed.). Dryden Press.
- Fatimah, S. (2022). The effect of diorama media on elementary school students' learning outcomes in ecosystem material [in Indonesian]. *Jurnal Basicedu*, 6(3), 1650–1661.
- Fitri, R., Lufri, L., Alberida, H., Amran, A., & Fachry, R. (2024). The project-based learning model and its contribution to student creativity: A review. *Jurnal Pendidikan Biologi Indonesia*, 10(1). <https://doi.org/10.22219/jpbi.v10i1.31499>
- Ghozali, I. (2021). Application of multivariate analysis using IBM SPSS 26 (10th ed.) [in Indonesian]. Badan Penerbit Universitas Diponegoro.
- Hake, R. R. (1999). Analyzing change/gain scores. Indiana University.
- Ministry of Education, Culture, Research, and Technology. (2022). Teacher and student books for

- Grade V Natural and Social Sciences (IPAS) [in Indonesian]. Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia.
- Larmer, J., & Mergendoller, J. R. (2010). Essentials for project-based learning: What students need to succeed. *Educational Leadership*, 68(1), 34–37.
- Lestari, T., & Mulyani, M. (2015). The effect of diorama media on fifth-grade students' learning outcomes in the elementary school ecosystem theme [in Indonesian]. *Jurnal Penelitian Pendidikan Guru Sekolah Dasar*.
- Maghfiroh, L., Juniarto, T., & Hanindita, A. W. (2024). The effect of diorama-assisted problem-based learning on fifth-grade students' science learning outcomes in food chain material [in Indonesian]. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, 9(2), 3978–3993.
- Marfita, I. P., Febriantika, I., & Yayuk, E. (2025). Project-Based Learning innovation in IPAS learning: A study of students' creativity in ecosystem dioramas [in Indonesian]. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, 10(2). <https://doi.org/10.23969/jp.v10i02.25822>
- Megawatie, A. S., Nugraha, A., & Lidinillah, D. A. M. (2017). Development of diorama media for learning the subtheme 'Let Us Love the Environment' in elementary school [in Indonesian]. *PEDADIDAKTIKA: Jurnal Ilmiah Mahasiswa Pendidikan Guru Sekolah Dasar*, 4(1), 29–39.
- Melinda, F., & Zainil, Y. (2020). Applying Project-Based Learning to improve elementary school students' critical thinking skills [in Indonesian]. *Jurnal Inovasi Pendidikan dan Pembelajaran Sekolah Dasar*, 4(1), 55–63.
- Nafi'an, M. I., Ngatmini, & Kiswati. (2022). Improving student learning outcomes through video-assisted Discovery Learning on Theme 1, 'The Beauty of Togetherness,' in Grade IV at SD Negeri 01 Boyoteluk in the 2022/2023 academic year [in Indonesian]. *Literasi (Jurnal Pendidikan Dasar)*, 2(2), 11–21. <https://doi.org/10.26877/literasi.v2i2.13448>
- Nitatalia, D., Ngatmini, & Budiawan, R. Y. S. (2023). Applying Project-Based Learning in teaching procedural-text writing to Grade VII students at SMPN 3 Jepara in the 2022/2023 academic year [in Indonesian]. *Teks: Jurnal Penelitian Bahasa, Sastra, dan Pengajarannya*, 8(1), 227–244. <https://doi.org/10.26877/teks.v8i1.15557>
- Nurfadhillah, S. (2021). Learning media: Definitions, foundations, functions, benefits, types, and use [in Indonesian]. CV Jejak.
- Nurtiansyah, R., & Wardhani, D. S. (2023). Development of diorama learning media using Project-Based Learning to improve fifth-grade students' conceptual understanding of ecosystems [in Indonesian]. *COLLASE*, 6(2), 101–112.
- Piaget, J. (1970). *Genetic epistemology* (E. Duckworth, Trans.). Columbia University Press.
- Rosita, R., Safitri, R. D., Suwarma, D. M., Muyassaroh, I., & Jenuri, J. (2024). A constructivist approach to improving elementary school students' learning outcomes [in Indonesian]. *Jurnal Review Pendidikan Dasar*, 10(3), 238–245.
- Safitri, D. S. A., Agustiningsih, A., & Wardoyo, A. A. (2024). The effect of diorama-based Project-Based Learning on fifth-grade students' creative thinking skills in food chain material at SDN Semboro 04 [in Indonesian]. *Jurnal Ilmu Pendidikan Sekolah Dasar*, 11(3), 365–374.
- Santrock, J. W. (2018). *Educational psychology*. McGraw-Hill Education.
- Schunk, D. H. (2012). *Learning theories: An educational perspective*. Pearson.
- Setiasih, C. H., Pratiwiningrum, Y., & Riyadi, R. (2024). Applying a concrete food chain diorama to improve fifth-grade students' IPAS learning outcomes at SD Negeri Tunggulsari 2 Surakarta [in Indonesian]. *Social, Humanities, and Educational Studies*. <https://doi.org/10.20961/shes.v7i4.97083>
- Setyosari, P. (2016). *Educational research and development methods* [in Indonesian]. Kencana.
- Siswanto, J., Huda, C., Nuroso, H., Widodo, W., & El Islami, R. A. Z. (2026). Designing a pedagogical framework to strengthen climate-integrated teaching skills for pre-service science teachers in support of the SDGs. *Journal of Education and E-Learning Research*, 13(2), 95–103. <https://doi.org/10.20448/jeelr.v13i2.8579>
- Siswanto, J., Khoiri, N., & Saphira, H. V. (2025). The effectiveness of the Algodoo-assisted IBMRO <https://siducat.org/index.php/isej/>

model in improving physics problem-solving skills. *Jurnal Pendidikan IPA Indonesia*, 14(3), 496–508. <https://doi.org/10.15294/jpii.v14i3.26670>

Sugiyono. (2021). *Quantitative, qualitative, and research and development methods [in Indonesian]*. Alfabeta.