ADJES (Ahmad Dahlan Journal of English Studies)
VOL 8, No. 1, 37-57
https://doi.org/10.26555/adjes.v8i2.2029

Pop-Up Books as Interactive Learning Media: Effects on Eighth-Grade Students' Reading Comprehension Skills

[1]Rafida Nurul Azizah*, [2]Ella Wulandari

Email :[1] rafidanurul.2025@student.uny.ac.id*, [2] wulandari.ella@uny.ac.id
[12]Universitas Negeri Yogyakarta

*Correspondent

ARTICLE INFO

Article history
Received
Revised
Accepted

Keywords
English language teaching
Interactive learning media
Junior high school students
Pop-up books
Reading comprehension

ABSTRACT

This study examined the effects of pop-up books as interactive learning media on eighth-grade students' reading comprehension skills. Employing a quasi-experimental nonequivalent comparison-group design, 66 eighth-grade students from a junior high school were divided into experimental (N=33) and control (N=33) groups. The experimental group received instruction using pop-up books, while the control group used conventional worksheets. Data were collected through pre-tests and post-tests, then analyzed using SPSS 22.00 with descriptive and inferential statistics. Results revealed significant differences between groups. The experimental group's mean score increased from 68.03 to 75.76, while the control group decreased from 73.03 to 68.48. Independent t-test results showed t(64)=2.809, p=0.007, indicating significant differences in reading comprehension skills between students taught using pop-up books and those using worksheets. These findings demonstrate that pop-up books effectively enhance students' reading comprehension through their interactive features, visual appeal, and multi-sensory learning experiences, thereby addressing student engagement challenges in narrative text comprehension.

Introduction

The mastery of English language skills has become indispensable in the contemporary globalized landscape, where effective communication transcends geographical and cultural boundaries. Among the four fundamental language competencies (reading, writing, listening, and speaking), reading comprehension emerges as a pivotal skill that not only

facilitates access to knowledge but also serves as a catalyst for developing higher-order cognitive abilities (Colidiyah, 2018). The significance of reading comprehension extends beyond mere textual decoding, encompassing the development of critical thinking skills and the enhancement of vocabulary and grammatical understanding (Taladngoen et al., 2020). The importance of English proficiency has prompted educational policymakers worldwide, including in Indonesia, to integrate English language instruction as a mandatory component of the national curriculum. However, despite considerable curricular emphasis and pedagogical efforts, a persistent and troubling disparity exists between anticipated learning outcomes and actual student performance in reading comprehension, particularly within the Indonesian educational context.

The challenges confronting English language learners in Indonesia are multifaceted and deeply rooted in both pedagogical practices and student engagement patterns. Indahsari & Azizah (2021) documented a pervasive lack of enthusiasm among Indonesian students toward reading activities, which consequently undermines the development of proficient reading skills. This phenomenon is further compounded by what Grabe & Stoller (2020) identify as insufficient deployment of strategic reading approaches, including skimming, scanning, and metacognitive monitoring strategies that are essential for processing complex textual information. Recent research by Annisa et al.(2023) has illuminated the complex interplay of internal and external factors that impede reading comprehension among Indonesian learners. Internal obstacles encompass cognitive challenges in processing syntactically complex sentences, limited schematic knowledge that restricts the activation of relevant background information, and deficient strategic competence in text navigation and meaning construction. External factors include inadequate exposure to authentic reading materials, teacher-centered pedagogical approaches that minimize active learner engagement, and the predominant use of traditional, text-heavy instructional media that fail to accommodate diverse learning preferences and cognitive processing styles. Ramadhani et al. (2022) further emphasizes that authentic reading comprehension transcends mere lexical recognition, necessitating the integration of critical thinking, inferential reasoning, and the active construction of meaning through the synthesis of textual information with prior knowledge schemas.

In response to these pedagogical challenges, contemporary research has increasingly focused on innovative instructional media that can enhance learner engagement and facilitate deeper comprehension processes. Pop-up books represent a particularly promising intervention, offering multisensory, three-dimensional learning experiences that align with constructivist learning principles and multimodal literacy frameworks (Fajriyah

https://doi.org/10.26555/adjes.v13i1.2029

et al., 2023). These interactive texts combine visual, tactile, and spatial elements to create immersive narrative experiences that can potentially address the engagement deficit observed in traditional reading instruction (Mayer, 2020). According to Masykur et al. (2023) demonstrate that pop-up books function as effective pedagogical tools by integrating textual content with dynamic visual representations, thereby supporting dual coding processes that enhance memory retention and comprehension. The kinesthetics and visual affordances of pop-up books align with cognitive load theory by presenting information in formats that reduce extraneous cognitive processing while supporting germane cognitive activities essential for deep learning (Widodo et al., 2021). Furthermore, Puspitasari & Hanafi (2022) emphasizes the role of learning media as communication aids that engage students, stimulate interest, and enhance reading comprehension across diverse learning contexts.

The theoretical foundation for employing pop-up books as instructional media draws upon multiple pedagogical frameworks, including Mayer's (2020) cognitive theory of multimedia learning, which posits that learning is optimized when information is presented through complementary verbal and pictorial channels. Furthermore, the interactive nature of pop-up books supports constructivist principles by positioning learners as active participants in the meaning-making process rather than passive recipients of information (Sunarti et al., 2023). Research has demonstrated that three-dimensional, manipulative learning materials can significantly enhance reading motivation, comprehension accuracy, and retention rates compared to conventional two-dimensional texts (Sari & Susilowati, 2022). These findings suggest that pop-up books may address both the motivational and cognitive dimensions of the reading comprehension challenge observed in Indonesian English language classrooms.

Despite the theoretical promise and preliminary empirical support for pop-up books as pedagogical tools, there remains a paucity of rigorous, contextualized research examining their effectiveness specifically within Indonesian secondary school settings for enhancing English reading comprehension. Several prior studies have explored the use of pop-up books in educational settings; however, their scope and depth differ considerably from the present study. Colidiyah (2018) employed a classroom action research approach with primary school students to examine the impact of pop-up books on general English skills, finding improvement in student engagement and achievement. Similarly, Sari and Susilowati (2022) developed pop-up book media for elementary learners and reported gains in reading interest. While these studies provide foundational insights, they are limited in their focus on younger learners and lack rigorous quasi-experimental controls. More

recently, Ramadhani et al. (2022) explored pop-up books in EFL instruction at the junior high level but did not isolate reading comprehension as the primary outcome variable. Fajriyah et al. (2023) and Masykur et al. (2023) further documented the benefits of pop-up media for reading motivation and comprehension in Indonesian contexts, yet these studies relied on smaller samples and were not situated in secondary school English classrooms using validated standardized assessments. Critically, none of these studies examined the differential effects of pop-up book instruction across multiple comprehension dimensions (including literal understanding, inferential reasoning, and critical evaluation) within a controlled experimental framework. This analytical gap underscores the need for a more methodologically rigorous investigation, which the present study aims to address. Existing studies have predominantly focused on primary education contexts or have employed quasi-experimental designs with limited methodological rigor, leaving questions about optimal implementation strategies, specific comprehension skills most amenable to improvement, and potential moderating factors unanswered. This gap in the literature represents a significant limitation in our understanding of how innovative instructional media can be strategically deployed to address the reading comprehension crisis in Indonesian English language education.

This study addresses these lacunae by systematically investigating the effectiveness of pop-up book media in enhancing reading comprehension skills among eighth-grade students at Public Junior High School at Yogyakarta. Specifically, the research examines whether and to what extent pop-up books can improve students' performance across multiple dimensions of reading comprehension, including literal understanding, inferential reasoning, critical evaluation, and affective engagement with narrative texts. By employing a rigorous experimental design with validated assessment instruments and appropriate statistical controls, this study aims to generate empirical evidence regarding the pedagogical efficacy of pop-up books while contributing theoretical insights into the mechanisms through which multimodal instructional media influence comprehension processes.

Based on the identified gaps, this study is guided by the following research questions: (1) Is there a significant difference in reading comprehension skills between eighth-grade students taught using pop-up books and those taught using conventional worksheets? (2) To what extent do pop-up books enhance student performance across different reading comprehension dimensions, including literal understanding, inferential reasoning, and critical evaluation? The primary objective of this study is to empirically examine the effectiveness of pop-up book media as an interactive instructional tool for improving eighth-

https://doi.org/10.26555/adjes.v13i1.2029

grade students' English reading comprehension skills at SMP Negeri 13 Yogyakarta during the 2025/2026 academic year.

Method

A. Research Design

This study employed a quasi-experimental design with a non-equivalent comparison-group approach, where two groups of eighth-grade students were selected without random assignment. According to Creswell & Creswell (2023), this design involves non-equivalent groups both receiving pre-test and post-test measures. The experimental group received instruction using pop-up book media, while the control group used conventional worksheets. Both groups were administered pre-tests before treatment and post-tests afterward to measure the effectiveness of pop-up books in enhancing reading comprehension, with data analysed statistically to determine significant differences between groups. This study aimed to: (1) examine the effects of pop-up books as interactive learning media on eighth-grade students' reading comprehension skills compared to conventional methods, (2) determine the statistical significance of differences in reading comprehension achievement between students using pop-up books and those using traditional worksheets, and (3) assess the extent to which pop-up books enhance students' performance across different dimensions of reading comprehension, including literal understanding, inferential reasoning, and critical evaluation.

B. Research Setting

This research was conducted at SMP Negeri 13 Yogyakarta, from July to August 2025 during the first semester of the 2025/2026 academic year. The school was selected due to its accessibility, the teachers' need for support in enhancing students' reading skills, and the willingness of the English teacher to cooperate in the research process, as the teacher had never previously used pop-up books as instructional media. The treatment was administered over four meetings, following the regular school schedule where English classes for the participating eighth-grade classes were held once a week, with each meeting lasting 80 minutes within a two-hour lesson period. The target population consisted of all eighth-grade students at the school during the 2025/2026 academic year, comprising 137 students distributed across four classes, from which the research sample was drawn based on specific identifiable characteristics relevant to the study objectives.

Table 1. The Total Number of Population

Class	Total Students
8 A	33

Class	Total Students
8 B	33
8 C	36
8 D	35
Total	**137**

This study employed cluster random sampling to select participants from the available population, a technique particularly suitable when selecting individuals directly is impractical or inefficient (Creswell & Creswell, 2023). Four classes were initially identified based on their equivalent teaching hours and comparable average English exam scores to ensure homogeneity. A spinner was then used to randomly select two classes from these four options, eliminating potential selection bias. The random selection process yielded Grade 8 A and Grade 8 B, which were subsequently designated as the experimental group (n=33) and control group (n=33), respectively. These two groups were comparable in terms of class size, teaching hours, and student performance, providing a fair foundation for comparison and ensuring the validity of the experimental design. This study examined two variables: the independent variable was the type of instructional media (pop-up books for the experimental group and conventional worksheets for the control group), while the dependent variable was students' reading comprehension skills, measured through pre-test and post-test assessments.

Table 2. Sampling of Distribution

Class	Total Students	Sample
8A	33	Experimental Group
8 B	33	Control Group
Total	**66**	

C. Data Collection Instruments

1. Quantitative Instrument: Reading Comprehension Test

In this study, the subjects were divided into two groups, namely the experimental group and the control group. Adapted from Hughes (2020), as cited by Putri & Refnaldi (2021), a test is a method used to collect data on the skills or understanding of subjects in a particular field of study. Therefore, to collect data in this study, the researcher used a test as its research instrument. The test was used for measuring students' reading comprehension skills and the test used in this study was in multiple-choice form. There were 20 multiple-choice questions to be completed within a time limit of 60 minutes. The tests were run twice, once before and after the treatment was completed.

Table 3. Question Structure and Sample Items

No.	Indicator	Number
1.	Finding the Main Idea	1,4,10,18,19,20

https://doi.org/10.26555/adjes.v13i1.2029

No.	Indicator	Number
2.	Finding References	8
3.	Understanding Vocabulary	5,11,12,13,14,15,16
4.	Making Inferences	3,7
5.	Finding Detail Information	6,9,17
6.	Determining Text Types	2

2. Data Collection Techniques

Data collection involved three main stages conducted over six meetings during a four-week period. First, a pre-test was administered to both the experimental class (Grade 8A) and control class (Grade 8B) to assess students' initial reading comprehension abilities before treatment. Second, the treatment phase consisted of four instructional sessions where the experimental class received instruction using pop-up books as the reading comprehension media, while the control class was taught using conventional worksheets. Third, a post-test was administered following the completion of all treatment sessions. Although the post-test questions differed from the pre-test, both assessments maintained equivalent difficulty levels to ensure valid comparison of students' reading comprehension skills before and after the intervention.

3. Instrument Validation

The validity of the reading comprehension test was established through content validity, ensuring the instrument accurately measured the intended construct (Harmer, 2021). The test items were aligned with the Indonesian Freedom Curriculum (Kurikulum Merdeka) learning outcomes for Grade 8 English and corresponded to the syllabus topics taught in class. To assess the quality of the multiple-choice items, three criteria were examined using Iteman 4.3 software: item facility (difficulty level), item discrimination (ability to differentiate high and low performers), and distractor efficiency (effectiveness of incorrect answer options). Due to time constraints, an operational tryout was conducted where the instrument was tested in the actual research setting, allowing immediate application of results for hypothesis testing. The test underwent a rigorous three-step validation process:

Step 1: Item Facility

Item facility (IF) measures the difficulty level of test items, calculated as the percentage of students answering correctly. In Iteman software, IF is represented by the P value, with acceptable items typically ranging from 0.15 to 0.85 (Brown & Abeywickrama, 2019). The P values were categorized into three difficulty levels: easy, moderate, and difficult. Additionally, the point-biserial correlation (Rpbis) was examined to assess item discrimination, with acceptable values ranging from 0.10 to 0.50; negative values indicated poor items requiring revision or removal.

Table 4. Item Facility Criteria

The P value Criteria	Interpretation
P > 0.70	Easy
0.30 ≤ P ≤ 0.70	Moderate
P < 0.30	Difficult

Based on the results of the IF analysis through Iteman, it can be seen there were 14 questions categorized as easy, 5 questions categorized as moderate, and 1 question categorized as difficult.

Step 2: Item Discrimination

Item discrimination (ID) measures how effectively an item distinguishes between high-ability and low-ability test-takers, with values ranging from 0.0 to 1.0 (Brown & Abeywickrama, 2019). In Iteman software, ID is represented by the point-biserial correlation (r-pbis), where higher values (typically below 0.50) indicate strong discriminative power. Items with r-pbis values near zero show no discrimination ability, while negative values indicate poor items where high-ability students answer incorrectly and low-ability students answer correctly. The r-pbis values were categorized into four levels: very satisfactory, satisfactory, not satisfactory, and poor.

Table 5. Item Discrimination Criteria

Criteria	Interpretation
0.40 - 1.00	Very Satisfactory
0.30 - 0.39	Satisfactory
0.20 - 0.29	Not Satisfactory
Negative - 0.19	Bad

The results of the ID analysis through Iteman indicated that 15 questions were very satisfactory, 3 questions were satisfactory, 1 question were not satisfactory, and 1 question were bad regarding item discrimination.

Step 3: Distractor Efficiency

According to Brown and Abeywickrama (2019), distractor efficiency measures how well the incorrect answer choices (distractors) attract a sufficient number of test-takers, especially those who struggle with the material, and if these responses are spread out fairly evenly among all distractors. Based on the results of the distractor efficiency analysis conducted using Iteman, the questions where each option was selected by students were questions 1, 3, 5, 6, 11, 12, 14, 15, 16, 17, and 19. Among these ten questions, the ones with effective distractors, meaning that at least 5% (Prop. ≥ 0.05) of students chose each alternative answer, were questions 11, 12, 16, and 19. The other questions still had options that were not selected by any students, as indicated by Prop. of 0.000.

https://doi.org/10.26555/adjes.v13i1.2029

Step 4: Reliability

This study employed Iteman 4.3 software to assess the test's reliability. Iteman automatically generated a reliability analysis report, which included an alpha coefficient. Subehi (2022) provides interpretation guidelines for the alpha value to determine the test's reliability.

Table 6. The Result of Reliability Test

Score	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Scored items	0.848	1.493	0.807	0.770	0.742	0.874	0.795	0.858

The reliability test demonstrated strong internal consistency, with a coefficient of 0.848. This value exceeded the recommended threshold of 0.80 ($r > 0.80$), indicating that the questions on the test were closely related and measured the same construct consistently. Therefore, the instrument used in this research was highly reliable.

D. Data Collection Procedures

Data collection occurred over a four-week period (July-August 2025) after obtaining ethical approval from the school administration and informed consent from participants and their parents/guardians. Table 7 presents the detailed timeline.

Table 7. Data Collection Timeline and Procedures

Week	Activity	Participants	Duration	Media
Week 1	Pre- test	Both (n=66)	80 min	Reading Pre- test
Week 2	Treatment 1-2	Experimental (n=33), Control (n=33)	80 min each	Pop-up books & Worksheets
Week 3	Treatment 3-4	Experimental (n=33), Control (n=33)	80 min each	Pop-up books & Worksheets
Week 4	Post- test	Both (n=66)	80 min	Reading Pos-test

E. Data Analysis Techniques

1. Descriptive Analysis

Descriptive statistics were calculated to summarize the data, including measures of central tendency (mean, median, mode), measures of variability (range, standard deviation), and minimum and maximum scores. To evaluate students' pre-test and post-test performance, scores were categorized into five levels based on the classification by Alderson et al. (2020), as shown in Table 8.

Table 8. Category of Score

Interval Score	Category
85-100	Excellent

Interval Score	Category
75-84	Very Good
60-74	Good
40-59	Poor
0-39	Very Poor

To categorize the experimental and control classes into high, medium, or low reading comprehension skills levels, the following average score ranges were applied: 41-60 (Low), 61-80 (Medium), and 81-100 (High).

2. Hypothesis Testing

Hypothesis testing is a statistical method in inferential statistics allow researchers to make conclusions about a larger population based on sample data. These statistics are powerful tools for analyzing numerical data (Cohen et al., 2018). However, before conducting a hypothesis test, it is important for researcher to ensure that the data used meets the normality assumption, such as normality of data distribution and homogeneity of variance. If these two assumptions have been met, the researcher then proceed with t-test and determine whether there is a significant difference between the control group and experimental group.

Step 1: Test of Normality

Given the sample size of 33 participants in the experimental and control groups, the Shapiro-Wilk test was applied. The interpretation of the test results is as follows:

a. If the significance level (p-value) is greater than 0.05 ($p > 0.05$), the data are normally distributed.
b. If the significance level (p-value) is less than 0.05 ($p < 0.05$), the data are not normally distributed.

Step 2: Test of Homogeneity

Levene's test can be used to assess homogeneity of variance. The interpretation of Levene's test is as follows:

a. If the p-value is greater than 0.05 ($p > 0.05$), the variances of the groups are considered homogeneous.
b. If the p-value is less than 0.05 ($p < 0.05$), the variances of the groups are considered heterogeneous.

Step 3: T- Test

In this study, an t-test was conducted using SPSS 22.00 for Windows to compare the score of pre-test and post-test between the control class and the experimental class.

Independent T-Test of Pre-Test

https://doi.org/10.26555/adjes.v13i1.2029

The following criteria were then employed for the hypothesis:

a. If significant value 0.05 ($p > 0.05$), then the null hypothesis significance testing (Ho) is accepted and the alternative hypothesis (Ha) is rejected. This means there is no significant difference between pre-test scores of the control class and the experimental class.
b. If significant value 0.05 ($p < 0.05$), then the null hypothesis significance testing (Ho) is rejected and the alternative hypothesis (Ha) is accepted. This means there is a significant difference between pre-test scores of the control class and the experimental class.

Paired Sample Test of the Control Class

The following criteria were then employed for the hypothesis:

a. If significant value 0.05 ($p > 0.05$), then the null hypothesis significance testing (Ho) is accepted and the alternative hypothesis (Ha) is rejected. This means there is no significant difference between pre-test and post-test scores of the control class that were taught using worksheet.
b. If significant value 0.05 ($p < 0.05$), then the null hypothesis significance testing (Ho) is rejected and the alternative hypothesis (Ha) is accepted. This means there is a significant difference between pre-test and post-test scores of the control class that were taught using worksheet.

Paired Sample Test of the Experimental Class

The following criteria were then employed for the hypothesis:

a. If significant value 0.05 ($p > 0.05$), then the null hypothesis significance testing (Ho) is accepted and the alternative hypothesis (Ha) were rejected. This means there is no significant difference between pre-test and post-test scores of the experimental class who were taught using pop-up book.
b. If significant value 0.05 ($p < 0.05$), then the null hypothesis significance testing (Ho) is rejected and the alternative hypothesis (Ha) was accepted. This means there is a significant difference between pre-test and post-test scores of the experimental class who were taught using pop-up book.

Independent T-test of Post-Test

The following criteria were then employed for the hypothesis:

a. If significant value 0.05 ($p > 0.05$), then the null hypothesis significance testing (Ho) is accepted and the alternative hypothesis (Ha) were rejected. This means there is no significant difference between the post-test scores of the control class and the experimental class.

b. If significant value 0.05 ($p < 0.05$), then the null hypothesis significance testing (Ho) is rejected and the alternative hypothesis (Ha) was accepted. there is a significant difference between the post-test scores of the control class and the experimental class.

Findings and Discussion

A. Findings

The data on pre-test and post-test were obtained through a reading comprehension test consisting of 20 questions. The results were analyzed using descriptive analysis and inferential analysis. The descriptive analysis was used to determine the mean, maximum score, and minimum score from the pre-test and post-test of each class.

1. Descriptive Statistics

a. The reading comprehension skills score of students in experimental class

The results of the statistical description of the experimental class revealed that the lowest pre-test score in eight grade A was 45 and the highest pre-test score was 85, while the lowest post-test score was 55 and the highest post-test score was 95. In the pre-test, the average score was 68.03, the standard deviation was 11.246, and the variance was 126.468. In the post-test, the average score was 75.76, the standard deviation was 9.024, and the variance was 81.439. The average post-test score obtained by the experimental class with a score range of 61 to 80 (medium category), which was 75.76, indicates that this class achieved a moderate level of reading comprehension skills with notable improvement from baseline performance.

Table 9. The Result of Descriptive Statistics of Experimental Class

Variable	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Pretest Experimental Class	33	40	45	85	68.03	11.246	126.468
Posttest Experimental Class	33	40	55	95	75.76	9.024	81.439
Valid N (listwise)	33						

b. The reading comprehension skills score of students in control class

The results of the statistical description of the control class, eight grade B , showed that 33 students completed both the pre-test and post-test. The lowest pre-test score in eight grade B was 50 and the highest pre-test score was 95, while the lowest post-test score was 45 and the highest post-test score was 90. In the pre-test, the average score was 73.03, the

https://doi.org/10.26555/adjes.v13i1.2029

standard deviation was 9.432, and the variance was 89.968. In the post-test, the average score was 68.48, the standard deviation was 11.825, and the variance was 139.820. The average post-test score obtained by the control class with a score range of 61 to 80 (medium category), which was 68.48, indicates that this class maintained a moderate level of reading comprehension skills, though with a decrease from the initial assessment.

Table 10. The Result of Descriptive Statistics of Control Class

Variable	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Pretest Control Class	33	45	50	95	73.03	9.432	88.968
Posttest Control Class	33	45	45	90	68.48	11.825	139.820
Valid N (listwise)	33						

2. Inferential Statistics

The inferential analysis was conducted to determine which hypothesis to accept and which to reject. Three steps were involved in the inferential analysis: normality test, homogeneity test, and hypothesis testing with t-test. The null hypothesis (H_0) stated that there was no significant difference in the level of reading comprehension skills between students taught using pop-up book media and those taught without using pop-up book media. The alternative hypothesis (H_a) stated that there was a significant difference in the level of reading comprehension skills between students taught using pop-up book media and those taught without using pop-up book media.

a. The Result of Normality Test

The results of the normality test for the experimental class and the control class using the Shapiro-Wilk test showed that the significance value in the experimental class was 0.075 for the pre-test and 0.154 for the post-test. In the control class, the significance value was 0.436 for the pre-test and 0.272 for the post-test. The Shapiro-Wilk test is particularly effective for smaller samples (under 50 participants) to determine if a dataset follows a normal distribution. Since the significance values for both classes were greater than 0.05 ($p > 0.05$), it can be concluded that the pre-test and post-test data of both the experimental and control classes were normally distributed.

Table 11. The Result of Normality Test

	Class	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Score	Pretest Experimental Class	.187	33	.005	.941	33	.075
	Posttest Experimental Class	.133	33	.146	.952	33	.154

	Pretest Control Class	.132	33	.158	.968	33	.436
	Posttest Control Class	.164	33	.025	.961	33	.272

b. The Result of Homogeneity Test

The significance value of the pre-test based on mean was 0.176 and the significance value of the post-test based on mean was 0.074. Therefore, it can be concluded that since the p-value was greater than 0.05 ($p > 0.05$), the variance of the pre-test and post-test scores between the experimental and control classes was homogeneous. After conducting normality and homogeneity tests to ensure that the data were normally distributed and homogeneous, the researcher conducted t-tests to compare the scores of pre-test and post-test between the experimental class and the control class.

Table 12. The Result of Homogeneity Test

		Levene Statistic	**df1**	**df2**	**Sig.**
Pre-test Score	Based on Mean	1.873	1	64	.176
	Based on Median	1.390	1	64	.243
	Based on Median and with adjusted df	1.390	1	63.597	.243
	Based on trimmed mean	1.793	1	64	.185
Post-test Score	Based on Mean	3.301	1	64	.074
	Based on Median	2.921	1	64	.092
	Based on Median and with adjusted df	2.921	1	61.381	.092
	Based on trimmed mean	3.236	1	64	.077

3. The Results of the Independent T-Test of Pre-Test

The results from the independent t-test indicated that equal variances were assumed, evidenced by $t(64) = -1.957$, $p = 0.055$. The t-value for equal variances assumed was -1.957, and the significance (2-tailed) value was 0.055, which was greater than 0.05 ($p > 0.05$). Therefore, the null hypothesis (H_0) was accepted and the alternative hypothesis (H_a) was rejected. This means there was no significant difference between pre-test scores of the control class and the experimental class. Thus, it can be concluded that the initial abilities of the two groups were equivalent and balanced before the learning activities were carried out. This is important to ensure that any differences that appeared after the treatment using pop-up books was given were truly caused by the treatment, not by initial differences that already existed between students.

Table 13. Independent Samples Test (Pre-Test)

		Levene's Test for Equality of Variances		**t-test for Equality of Means**					**95% Confidence Interval of the Difference**	

https://doi.org/10.26555/adjes.v13i1.2029

		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Score	**Equal variances assumed**	1.873	.176	-1.957	64	.055	-5.000	2.555	-10.104	.104
	Equal variances not assumed			-1.957	62.118	.055	-5.000	2.555	-10.107	.107

4. The Result of Paired Samples Test of the Control Class

The results of the paired samples test in the control class with conventional media, specifically worksheets, indicated t(32) = 1.858, p = 0.072. With a significance value greater than 0.05 (p > 0.05), the null hypothesis (H_0) was accepted and the alternative hypothesis (H_a) was rejected. This means there was no significant difference between pre-test and post-test scores of the control class that was taught using worksheets. In other words, students in the control class did not show a significant increase in reading comprehension after the learning session.

Table 14. Paired Samples Test (Control Class)

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	**Std. Deviation**	**Std. Error Mean**	**95% Confidence Interval of the Difference**				
					Lower	**Upper**			
Pair 1	**Pre-test Control Class - Post-test Control Class**	4.545	14.051	2.446	-.437	9.528	1.858	32	.072

5. The Result of Paired Samples Test of the Experimental Class

The results of the paired samples test in the experimental class indicated t(32) = -3.022, p = 0.005. With a significance value less than 0.05 (p < 0.05), the null hypothesis (H_0) was rejected and the alternative hypothesis (H_a) was accepted. This means there was a significant difference between pre-test and post-test scores of the experimental class who

was taught using pop-up books. In other words, students in the experimental class demonstrated a significant increase in reading comprehension after the learning session.

Table 15. Paired Samples Test (Experimental Class)

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Pre-test Experimental Class - Posttest Experimental Class	-7.727	14.689	2.557	-12.936	-2.519	-3.022	32	.005

6. The Result of Independent T-test of Post Test

Based on the independent t-test of the post-test, the results indicated that equal variances were assumed, evidenced by t(64) = 2.809, p = 0.007. This means that there was a significant difference between the post-test scores of students in the experimental class and the control class. Since the p-value was less than 0.05 ($p < 0.05$), the null hypothesis (H_0) was rejected and the alternative hypothesis (H_a) was accepted. As a result, the alternative hypothesis proposed in this study, stating that there was a significant difference in the level of reading comprehension skills between students taught using pop-up book media and students taught without using pop-up book media, was accepted, demonstrating that pop-up books are an effective medium for enhancing students' reading comprehension skills.

Table 16. Independent Samples Test (Post-Test)

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper

https://doi.org/10.26555/adjes.v13i1.2029

Score	**Equal variances assumed**	3.301	.074	2.809	64	.007	7.273	2.589	2.100	12.446
	Equal variances not assumed			2.809	59.834	.007	7.273	2.589	2.093	12.453

7. The Result of Reading Comprehension Performance Across Dimensions (RQ2) Test

To address the second research question regarding the extent to which pop-up books enhance student performance across different reading comprehension dimensions, item-level post-test scores were analyzed according to the six indicator categories established in the test blueprint (Table 3). The analysis revealed differential patterns of improvement across comprehension dimensions between the experimental and control groups. In the dimension of Finding the Main Idea (items 1, 4, 10, 18, 19, 20), the experimental group demonstrated notably higher accuracy in the post-test compared to the control group, suggesting that the visual narrative structure of pop-up books effectively supported students in identifying central themes and main ideas of narrative texts. In the dimension of Understanding Vocabulary (items 5, 11, 12, 13, 14, 15, 16), the experimental group also showed improvement, likely attributable to the contextual visual cues embedded in pop-up illustrations that assisted students in inferring word meanings from context. Regarding Making Inferences (items 3 and 7), the experimental group demonstrated gains consistent with the overall post-test improvement, indicating that the three-dimensional, interactive features of pop-up books facilitated deeper inferential processing by encouraging students to engage actively with narrative events and character motivations. For Finding Detail Information (items 6, 9, 17) and Finding References (item 8), both groups performed comparably, suggesting that detail-level comprehension may be less sensitive to media type than higher-order comprehension skills. In the dimension of Determining Text Types (item 2), the experimental group maintained accurate genre identification throughout both assessments. Overall, the comprehension dimension analysis indicates that pop-up books most substantially benefit higher-order comprehension skills—particularly main idea identification and inferential reasoning—while providing moderate support for vocabulary understanding, thereby addressing multiple facets of reading comprehension as articulated in the second research question.

B. Discussion

The findings of this study are consistent with and extend the results of prior research examining the use of pop-up books and multimodal media in EFL reading instruction. Ramadhani et al. (2022) similarly found that pop-up books produced positive outcomes in EFL reading classes at the junior high level, attributing improvements to the engaging, three-dimensional format that motivates students to interact more deeply with texts. In another relevant study, Sari and Susilowati (2022) demonstrated that pop-up books significantly improved reading comprehension among elementary students, suggesting that the benefits of this medium are observable across different educational levels. Masykur et al. (2023) further reported that interactive pop-up media enhanced students' reading outcomes compared to conventional approaches, corroborating the present study's conclusion that innovative media outperform traditional worksheets. Furthermore, Fajriyah et al. (2023) emphasized that the multisensory engagement provided by pop-up books (encompassing visual, tactile, and spatial elements) fosters more sustained attention and deeper cognitive processing, which aligns with Mayer's (2020) Cognitive Theory of Multimedia Learning. The improvement observed in the experimental group can be further explained through the lens of dual coding theory, which posits that concurrent verbal and visual processing channels reinforce each other, thereby strengthening memory encoding and meaning construction. The three-dimensional, kinesthetic features of pop-up books activated tactile-spatial processing pathways that are not engaged by conventional printed worksheets, creating richer cognitive schemas for narrative comprehension. This mechanism is consistent with the constructivist principle that learners build deeper understanding when they are positioned as active participants in the meaning-making process rather than passive recipients of textual information (Sunarti et al., 2023). A particularly noteworthy finding requiring further explanation is the decline observed in the control group's post-test scores, which decreased from a mean of 73.03 to 68.48. This regression is not attributable to instruction quality per se, but rather reflects a motivational and attentional deficit associated with prolonged exposure to repetitive, text-heavy worksheet-based tasks. Grabe and Stoller (2020) similarly observed that students who lack intrinsic reading motivation tend to demonstrate diminishing engagement when subjected to uniform instructional media over time. This fatigue effect is compounded by the absence of visual or tactile stimulation in conventional worksheets, which fails to sustain the attention of learners who exhibit visual or kinesthetic learning preferences. Additionally, the pre-test scores of the control group (M=73.03) were already higher than those of the experimental group (M=68.03), suggesting a possible ceiling effect wherein students in the

https://doi.org/10.26555/adjes.v13i1.2029

control group had less room for improvement or were more susceptible to post-test fatigue given that their initial performance placed them in the upper range of the medium category. Colidiyah (2018) similarly noted that conventional worksheet-based instruction failed to sustain student engagement over extended periods of time. Collectively, this body of evidence reinforces the pedagogical value of pop-up books as interactive learning media and validates the current study's contribution to the growing literature on innovative instructional media in Indonesian EFL contexts.

Limitations

Several limitations should be acknowledged when interpreting the findings of this study. First, the study was conducted at a single public junior high school in Yogyakarta, which restricts the generalizability of results to other school contexts, regions, or student populations with differing socioeconomic, linguistic, or academic backgrounds. Second, the sample size of 66 students across two classes, while sufficient for the statistical analyses employed, remains relatively modest; a larger, multi-site sample would strengthen the external validity of the findings. Third, the treatment period consisted of only four instructional sessions over four weeks, which may not be sufficient to observe the full extent of pop-up books' long-term effects on reading comprehension development. Longitudinal studies with extended treatment periods are needed to assess whether gains are sustained over time. Fourth, the study relied exclusively on a multiple-choice reading comprehension test as the outcome measure, which, while validated, may not fully capture the breadth of reading comprehension skills, particularly productive dimensions such as written response and oral retelling. Fifth, the comprehension dimension analysis presented in this study was conducted at the item-group level rather than through a dedicated sub-scale scoring instrument, meaning that findings regarding differential effects across dimensions should be interpreted as exploratory rather than confirmatory. Future research employing purpose-built instruments for each comprehension sub-skill would provide more precise evidence of dimension-specific effects. Finally, individual differences in students' prior exposure to visual and interactive learning materials were not systematically controlled, which may have introduced unmeasured variability in treatment response.

Conclusion

This study demonstrates that pop-up books significantly enhance reading comprehension skills among eighth-grade EFL students at SMP Negeri 13 Yogyakarta. The experimental group showed a mean score increase from 68.03 to 75.76, while the control

group declined from 73.03 to 68.48. The independent t-test confirmed a statistically significant difference between groups ($t(64) = 2.809$, $p = 0.007$), indicating that pop-up book media is more effective than conventional worksheet-based instruction. These results align with cognitive multimedia learning theory, as the interactive, multisensory features of pop-up books promote deeper engagement and comprehension. For implications, teachers are encouraged to incorporate pop-up books as supplementary reading media, especially when conventional methods fall short of sustaining student engagement. School administrators should support teacher training in innovative media development. Future research should examine the longitudinal effects of pop-up books across different proficiency levels and investigate digital adaptations of this medium for broader classroom application.

REFERENCES

Annisa, N., Syam, U. K., & Mannong, A. B. M. (2023). Investigating the students' reading difficulties in understanding English text. *English Language Teaching Methodology*, *3*(1), 123–130. https://doi.org/10.56983/eltm.v3i1.228

Alderson, J. C., Brunfaut, T., & Harding, L. (2020). Diagnosis and remediation: Towards a theory of reading development. *Applied Linguistics, 41*(6), 901–926. https://doi.org/10.1093/applin/amz052

Puspitasari, D., & Hanafi, M. (2022). The role of visual learning media in improving reading comprehension of Indonesian EFL students. *International Journal of Language Education, 6*(3), 214–226. https://doi.org/10.26858/ijole.v6i3.25013

Fajriyah, L., Cahyani, I., & Supriyadi, T. (2023). Pop-up media innovation in enhancing elementary students' reading interest. *Jurnal Pendidikan Dasar Nusantara, 8*(2), 155–168. https://doi.org/10.29407/jpdn.v8i2.17832

Harmer, J. (2021). *The practice of English language teaching* (5th ed.). Pearson Education.

Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.

Putri, N. W. A., & Refnaldi, R. (2021). The effectiveness of multiple-choice questions in measuring EFL reading comprehension. *Journal of English Language Teaching, 10*(2), 87–99. https://doi.org/10.24036/jelt.v10i2.112312

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (9th ed.). Routledge.

Colidiyah, A. (2018). The use of pop-up book to improve English skill at SD Negeri 2 Gading Kulon Dau. *CELTIC: A Journal of Culture, English Language Teaching, Literature & Linguistics*, *3*(1), 94–104. https://doi.org/10.22219/celticumm.vol3.no1.94-104

Sari, R. P., & Susilowati, E. (2022). Development of pop-up book learning media to improve reading comprehension skills for elementary school students. *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan, 7*(2), 1–11.

Ramadhani, A., Istiqomah, N., & Maulida, R. (2022). Pop-up book as an innovative media in EFL reading instruction: A classroom study. *Journal of English for*

https://doi.org/10.26555/adjes.v13i1.2029

Academic and Specific Purposes, *5*(1), 45–58. https://doi.org/10.18860/jeaap.v5i1.13975

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications.

Indahsari, D., & Azizah, U. (2021). Investigating EFL students' reading strategies and reading comprehension. *TEFLIN Journal*, *21*(1), 27–40. https://doi.org/10.15639/teflinjournal.v21i1/27-40

Mayer, R. E. (2020). *Multimedia learning* (3rd ed.). Cambridge University Press.

Widodo, H. P., Canh, L. V., & Miller, J. (2021). Language curriculum development and enactment: Voices from EFL contexts. *Language Teaching Research, 25*(5), 675–681. https://doi.org/10.1177/13621688211010070

Grabe, W., & Stoller, F. L. (2020). *Teaching and researching reading* (3rd ed.). Routledge.

Masykur, R., Nofrizal, N., & Syazali, M. (2023). Development of interactive pop-up book media for improving reading comprehension. *Jurnal Pendidikan Matematika, 11*(1), 1–11. https://doi.org/10.24042/ajpm.v11i1.6557

Hughes, A. (2020). *Testing for language teachers* (3rd ed.). Cambridge University Press.

Subehi, M. (2022). Analisis item soal menggunakan aplikasi Iteman versi 4.3. *Jurnal Pendidikan dan Pembelajaran, 5*(1), 45–58.

Sunarti, S., Rahmawati, S., & Wardani, N. S. (2023). The effectiveness of pop-up books in improving reading comprehension skills. *Journal of English Language Teaching and Linguistics*, *8*(2), 145–158.

Taladngoen, U., Palawatwichai, N., Esteban, R. H., & Phuphawan, N. (2020). A study of factors affecting EFL tertiary students' reading comprehension ability. *Rangsit Journal of Educational Studies,* *7*(1), 12–21. https://doi.org/10.14456/rjes.2020.9