# Effect of End-of-Topic Tests on Pupils' Performance in Integrated Science: A Study Using Control and Experimental Groups

**Peter Chirwa**

**Department of Natural Sciences, Ministry of Education, Zambia**

## ABSRACT

Integrated Science, a core subject in Zambia's junior secondary curriculum, serves as a foundation for senior secondary subjects like Physics, Chemistry, and Biology. Despite its significance, student performance often falls short due to inadequate assessment practices and feedback mechanisms. Continuous assessment, especially end-of-topic tests, is a critical tool for evaluating learning progress, but its effectiveness in improving academic outcomes in Integrated Science has not been rigorously evaluated. This study investigated the effect of end-of-topic tests on pupils' performance and attitudes toward learning Integrated Science.

The study employed a quasi-experimental design at Mejocama Secondary School, Lusaka, involving 60 Grade 8 pupils divided into experimental and control groups. The experimental group underwent regular end-of-topic tests over one academic term, while the control group followed traditional instruction without such tests. Data were collected using pre-tests, post-tests, and questionnaires. Quantitative data were analyzed using descriptive statistics, Shapiro-Wilk normality tests, and independent samples t-tests. Qualitative data on pupils' attitudes were analyzed thematically.

Pre-test scores showed no significant difference between the experimental and control groups ($p = 0.229$). However, post-test results revealed significant performance improvements in the experimental group. The mean score for post-test one was 52.5% for the experimental group compared to 31.1% for the control group ($p = 0.001$). Similarly, post-test two indicated a mean score of 62.3% for the experimental group versus 43.9% for the control group ($p = 0.001$). Qualitative findings showed that end-of-topic tests enhanced students' study habits, conceptual understanding, and confidence, although some reported test-related anxiety.

End-of-topic tests significantly improve academic performance in Integrated Science and positively influence pupils' attitudes toward learning. These findings highlight the importance of integrating well-structured assessments into teaching practices to enhance learning outcomes.

**Key Words:** End-of-Topic tests, Integrated Science, Continuous Assessment, Formative Assessment

## INTRODUCTION

Integrated science is a core subject in junior secondary education, essential for developing foundational knowledge for advanced studies in senior secondary subjects such as Physics, Chemistry and Biology. However, performance in this subject often varies widely due to differences in teaching methodologies, assessment strategies, and student engagement.

In today's competitive educational society, accurate assessment plays a crucial role (Tariq & Ali, 2019). An assessment that is fair, valid, and reliable serves its purpose, while an inadequate assessment fails to do so. Berry (2008) highlights the significance of assessment in education by revealing both students' achievements and areas for improvement, thereby enabling subsequent actions to be implemented. Nitko (2004) defines assessment as a process of collecting information about student learning and performance to make decisions about students, curricula, programs, and educational policies.

# BACKGROUND

Assessment serves as a means of measuring students' abilities in acquiring specific knowledge or skills, allowing for the evaluation of education quality at all levels. McAlpine (2002) views learner assessment as a two-way communication process that provides feedback on the educational process or product to key stakeholders. Tariq & Ali (2019) adds that assessment allows schools to maintain complete records of students' growth and progress, enabling unbiased judgments across cognitive, affective, and psychomotor domains. Educational practitioners and stakeholders utilize assessment results to evaluate the entire educational system, motivate students, improve instructional planning and content, and certify students' achievements at specific levels.

Regular end-of-topic tests are a common formative assessment approach used to reinforce learning and provide feedback to students and teachers (Ministry of General Education, 2017). However, their effectiveness compared to other learning approaches remains an open question. This study introduces an experimental design to assess the effect of end-of-topic tests on pupils' academic performance in Integrated Science by comparing an experimental group that received regular end-of-topic tests and a control group that did not.

## Problem Statement

Integrated Science is a critical subject for building foundational scientific knowledge and skills, yet students' performance often remains below expectations. Traditional teaching methods often lack the structured feedback mechanisms necessary for continuous improvement. While end-of-topic tests are widely used to reinforce learning, their effectiveness in improving pupil performance has not been systematically evaluated in the context of Zambian schools.

This study seeks to address this gap by investigating the impact of end-of-topic tests on pupils' academic performance in Integrated Science. Specifically, it will determine whether regular testing enhances understanding, retention, and overall academic outcomes compared to traditional instruction without such tests.

## Role of Assessment in Education

Assessment, in a broad sense, refers to the various methods and tools used by educators to evaluate, measure, and document students' academic readiness, learning progress, skill acquisition, and educational needs. Assessments serve as indicators for students, clarifying what they need to learn and giving tangible meaning to valued learning objectives (Stiggins, 2007). Assessment has different effects at different times or the stages of the course (Tariq & Ali, 2019). According to Black and Wiliam (1998), assessments serve two main purposes: formative assessment to support learning and summative assessment to evaluate learning outcomes. End-of-topic tests primarily function as formative assessments as they assist teachers to continuously gauge learners` understanding and tailoring their teaching to ensure challenges to learning are controlled (Ministry of Education, 2013). And studies such as those by Brookhart (2017) have shown that well-structured assessments promote student motivation, self-regulation, and deeper understanding of concepts.

## End-of-Topic Tests: Definition and Purpose

End-of-topic tests are a kind of assessments administered at the conclusion of a specific topic or chapter. Their primary purpose is to evaluate the retention and comprehension of key concepts or gauge whether learning has taken place (Stiggins, 2007). Research by Brown et al. (2014) highlights that periodic testing encourages retrieval practice, which strengthens memory and enhances long-term retention. Continuous assessment helps teachers to interpret and synthesis information about the learners and create a shared academic culture dedicated to assuring and improving the quality of education (Airasian, 1991). Continuous assessment (CA) is a critical component of student evaluation in Zambia. However, to ensure the validity of tests and the assessment process, it is crucial to align end-of-course test scores with preceding continuous assessment scores from teacher-made tests (Olufemi, 2014).

In the Zambian context, continuous assessment is an ongoing diagnostic and school-based process that utilizes various assessment tools to measure learner performance (Kapambwe, 2006). Nitko (2004) describes continuous assessment as an ongoing process of gathering and interpreting information about student learning, which informs instructional decisions. Quansah (2005) explains that continuous assessment serves two purposes: enhancing the validity and reliability of students' results and fostering effective learning and work habits (Quansah, 2005). Continuous assessment is a form of evaluation that occurs throughout the learning process, providing continuous feedback and opportunities for improvement. Continuous assessment is systematic, formative, guidance-oriented, and diagnostic in nature (Nitko, 2004). It promotes frequent interaction between learners and teachers, allowing teachers to identify strengths and weaknesses and provide targeted feedback to help learners focus on areas that need improvement.

## Impact of End-of-Topic Tests on Academic Performance

A research study by Roediger and Karpicke (2006) found that frequent testing improved long-term knowledge retention and its application. This "testing effect" indicates that assessments do more than measure learning; they also enhance it. Research on the impact of testing on academic performance has yielded mixed results.

However, the design and implementation of the end-of-topic tests determine their effectiveness. The improved outcome does not only depend on the design and implementation but also the curriculum should be learner-centred and outcome-based (Ministry of Education, 2013). The outcome of quality tests aligned with curriculum objectives have shown to improve pupils` performance in science subjects (Harlen W. , 2013).

## Control and Experimental Group Designs in Educational Research

In order to establish causality in a scientific study or educational research, a control group is used by isolating the effect of an independent variable. By comparing outcomes between groups exposed to different interventions, researchers can isolate the effects of specific variables, such as end-of-topic tests (Campbell & Stanely, 1963).

In the context of this study, the experimental group underwent regular end-of-topic testing, while the control group followed traditional teaching methods without such assessments. Previous research, including those by (Hamtini, Albasha, & Varoca, 2015; Kubsch, et al., 2022; Harlen & James, 1997), have validated the reliability of this approach in assessing educational interventions.

## Integrated Science Education and Assessment

Integrated Science education aims to provide students with a cohesive understanding of scientific principles across disciplines. According to UNESCO (2017), effective teaching and assessment practices in Integrated Science are critical for preparing students for real-world problem-solving.

End-of-topic tests in Integrated Science can address specific challenges, such as students' difficulties in connecting theoretical knowledge to practical applications (Luft & Tiene, 2001; Roediger III & Karpicke, 2006). When designed to include application-based and interdisciplinary questions, these tests have been shown to enhance students' critical thinking and analytical skills.

## Gaps in the Literature

Despite the growing body of research on assessments, there is limited focus on the specific impact of end-of-topic tests in Integrated Science education. Additionally, few studies have utilized control and experimental group designs to rigorously evaluate these tests' effectiveness. This study seeks to address these gaps by exploring how regular end-of-topic testing influences pupils' performance in Integrated Science.

## Hypothesis

➢ Null Hypothesis ($H_0$): There is no significant difference in performance between the control and experimental groups.

# METHODOLOGY

The research design that was used in this research was a quasi-experimental design and the participants were not randomly assigned to the control and the experimental groups (Cresswell, 2009). A quasi-experimental research design is an empirical study used to estimate the casual impact of an intervention on its target population. This research design was more appropriate for this study as it objectively evaluated the impact of end-of-topic tests on pupils' performance. A mixed method approach was used, in that both qualitative and quantitative methods were utilized in analyzing data to ensure a comprehensive evaluation. This study was conducted at Mejocama Secondary School in Lusaka province of Lusaka district in Zambia. It was conducted among all the two (2) grade eight (8) pupils taking Integrated Science as the target population which was sixty (60). The sample had 60 pupils from two (2) classes for pre-test, 52 pupils for post-test one and 57 pupils for post-test two, who were selected purposively since the school has only two classes of grade eight and assigned one class to the experimental group and the other group to the control group using coin flip.

Data was collected using the following instruments, achievement tests; pre-test, post-test and end of topic tests and questionnaires in order to answer the research questions. Pupils' achievement from pre-test, post-test and end of topic tests made up quantitative data while the questionnaire captured the qualitative data (pupils` attitudes and perceptions). The data collection procedures were divided into three phases; Pre-test (Baseline Assessment), Intervention and Post-test (Final Assessment).

## Phase 1: Pre-Test (Baseline Assessment)

A standardized test administered to both groups to assess their prior knowledge of Integrated Science topics in order to help establish the homogeneity of the groups.

## Phase 2: Intervention

## Experimental Group:

To the experimental group end-of-topic tests after teaching each topic over one academic term were administered

## Control Group:

The control group was taught the same topics but without administering end-of-topic tests.

## Phase 3: Post-Test (Final Assessment)

At the end of the term, a standardized test was administered covering all topics to both groups in order to determine the group which achieved higher than the other.

The data collected was analyzed using statistical Package for Social Science (SPSS) programme version 20, where the descriptive statistics were computed for both pre-test and post-test as well as end of topic tests. The mean, standard deviation and frequencies were generated under descriptive statistics. Descriptive statistics provides simple summaries about the sample and made no predictions (Trochim, 2006). Before an independent sample t-test was performed, the data from the pre-test, end of topic tests and post-test was first tested for normality using Shapiro-Wilk test. Testing for normality is one of the assumptions data must meet in order for an independent sample t-test to give valid results. The null hypothesis ($H_0$) was that the data is approximately normal and its alternative hypothesis ($H_1$) was that data is not approximately normal. It follows that if the p-value is greater than alpha (α) = 0.05 (P>0.05), then do not reject the null hypothesis and conclude that data is normally distributed. If the p-value is less than alpha (α) = 0.05(P<0.05), then the null hypothesis would be rejected and conclude that data was not normally distributed. An independent sample t-test analysis was done on the pre-test, post-test and end of topic tests, in order to test the level of degree of significance between the two group's means (experimental and control) being compared at alpha (α) = 0.05 level of significance. According to the institute for digital research and education (2014) an independent t-test can be designed to compare means of the same variable between two groups.

For the qualitative data collected, thematic analysis was used to identify common themes and insights from open-ended questionnaire responses. With this approach, research is effective in obtaining culturally specific information based on the subjective assessment of attitude, opinions, behaviours and accuracy of findings and conclusion are the function of the researchers` insight and impression. Moreover, qualitative approach seeks to understand a given research problem or topic from the perspectives of the population involved in the research, it is also a form of social inquiry that seeks to understand the way people interpret and make sense of different experience they are encountering in their life (Bryman, 2013).

## Ethical Considerations

The researcher obtained informed consent from pupils and school administration. The names of all pupils whose results were used as data for this study were kept anonymous to ensure participants' privacy and confidentiality. The researcher also ensure that all pupils received adequate teaching and support, regardless of group allocation.

# RESEARCH FINDINGS

## Findings from the Pre-test Results

Table 1, 2 and 3 shows the SPSS output before treatment. The treatment in this case was referred to the use of end of topic test in assessing the pupils, which was only administered to the experimental group.

Table 1:Shapiro-Wilk Normality Test For Pre-Test Results

|  | Group Name | Shapiro-Wilk | | |
|---|---|---|---|---|
|  |  | Statistic | df | Sig. |
| Pre-test | Control | .939 | 32 | .070 |
|  | Experimental | .975 | 28 | .732 |

In both cases the p-value was greater than alpha ($\alpha$) = 0.05 (P=0.259 > 0.070 and P=0.732 > 0.05). This indicated that the test scores from the pre-test were normally distributed and it implied that the independent sample t-test could be used on this data.

Table 2: Descriptive Statistics for Pre-Test Results

|  | Group Name | N | Mean | Std. Deviation |
|---|---|---|---|---|
| Pre-test | Control | 32 | 42.88 | 17.437 |
|  | Experimental | 28 | 47.61 | 11.717 |

Table 2 shows the descriptive statistics, which describes the main features of the pre-test results which includes the sample size N=32 for the control group and N=28 experimental group, the mean for the experimental group, M=42.88%, the mean for the control group, M=47.61%, standard deviation for the control group Std=17.437 and standard deviation for the experimental group Std=11.717. The difference in the pre-test mean scores between the experimental and the control group was 4.73% and the variations in the data (i.e. Spread of test scores) was wider for the experimental group (Std deviation=11.717) than the control group (Std deviation=17.437).

Table 3: Independent Sample T-Test for the Pre-Test Results

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | | |
| | | | | | | | | | **Lower** | **Upper** |
| Pre-test | **Equal Variances assumed** | 2.608 | .112 | -1.215 | 58 | **.229** | -4.732 | 3.894 | **-**12.526 | 3.062 |

Table 3 presents the independent sample t-test results for the pre-test scores of the experimental and control groups. The first section highlights Levene's test for equality of variances and F=2.608, the p-value was 0.112 (p= 0.112 > 0.05). The second section shows the t-test for equality of means, where the p-value was 0.229 (p= 0.229 > 0.05, t=−1.215).

## FINDINGS FROM THE POST-TEST 1 RESULTS

Table 4: Shapiro-Wilk Normality Test for Post-Test 1 Results

| | | Shapiro-Wilk | | |
|---|---|---|---|---|
| | Group Name | Statistic | df | Sig. |
| Post-test 1 | Control | .885 | 27 | .006 |
| | Experimental | .908 | 25 | .028 |

The Shapiro-Wilk test evaluates whether the data follows a normal distribution. For the control group (Statistic=0.885, df=27, p=0.006), the p-value is less than 0.05, indicating that the post-test 1 scores are not normally distributed. Similarly, for the experimental group (Statistic=0.908, df=25, p=0.028), the p-value is also less than 0.05, confirming that the post-test 1 scores for this group are not normally distributed. Hence, the researcher proceeded with Mann-Whitney U test, a non-parametric test.

Table 5: Descriptive Statistics for Post-Test 1 Results

| | N | Mean | Std. Deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| Post-test 1 | 52 | 41.38 | 24.897 | 0 | 92 |
| Group | 52 | 1.48 | .505 | 1 | 2 |

The descriptive statistics for post-test 1 show that the 52 participants had an average score of 41.38, with a high variability in performance (standard deviation = 24.897), ranging from 0 to 92. For the grouping variable, which distinguishes the experimental group (1) from the control group (2), the mean value of 1.48 reflects a nearly equal distribution between the two groups, with a standard deviation of 0.505.

Table 6: Ranks Table for Post-Test 1 Results

|  | Group Post-test 1 | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| Post-test 1 | 1 | 27 | 19.59 | 529.00 |
|  | 2 | 25 | 33.96 | 849.00 |
|  | Total | 52 |  |  |

The ranks table provides a summary of the distribution of scores for the two groups. The control group (Group 1), with 27 participants, has a mean rank of 19.59 and a sum of ranks of 529.00. On the other hand, the experimental group (Group 2), with 25 participants, has a mean rank of 33.96 and a sum of ranks of 849.00. The higher mean rank of the experimental group suggests that the experimental group performed better on average in the post-test 1 compared to the control group. This preliminary observation indicates a potential difference in the performance distributions between the two groups, which is further tested using the Mann-Whitney U test.

Table 7: Test Statistics Table for Post-Test 1

|  | Post-test 1 |
|---|---|
| **Mann-Whitney U** | **151.000** |
| Wilcoxon W | 529.000 |
| Z | -3.438 |
| **Asymp. Sig. (2-tailed)** | **.001** |

The Mann-Whitney U test statistics show that the U-value is 151.000, with a corresponding Z-score of -3.438 and an asymptotic significance level (p) of 0.001.

**Findings from the Post-test 2 Results**

Table 8: Shapiro-Wilk Normality Test for Post-Test 2 Results

|  | Group Name | Shapiro-Wilk | | |
|---|---|---|---|---|
|  |  | Statistic | df | Sig. |
| Post-test 2 | Control | .965 | 29 | .441 |
|  | Experimental | .968 | 28 | .525 |

Table 8 above shows that the p-value for the control group (Statistic=0.965, df=29, p=0.441), is greater than 0.05, indicating that the post-test 2 scores are normally distributed. Similarly, for the experimental group (Statistic=0.968, df=28, p=0.525), the p-value is also greater than 0.05, confirming that the post-test 1 scores for this group are were normally distributed.

Table 9: Descriptive Statistics for Post-Test 2 Results

|  | Group Name | N | Mean | Std. Deviation |
|---|---|---|---|---|
| Post-test 2 | Control | 29 | 43.93 | 21.020 |
|  | Experimental | 28 | 62.32 | 16.475 |

The Table 9 above shows the descriptive statistics for post-test 2 and the control group had 29 participants with a mean score of 43.93 and a standard deviation of 21.020, indicating moderate variability in scores. The experimental group had 28 participants with a higher mean score of 62.32 and a lower standard deviation of 16.475, reflecting relatively less variability in performance compared to the experimental group.

Table 10: Independent Sample T-Test for the Post-Test 2 Results

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| | | | | | | | | | **Lower** | **Upper** |
| Post-test 2 | **Equal Variances assumed** | 1.565 | .216 | -3.668 | 55 | **.001** | -18.390 | 5.014 | **-28.439** | -8.342 |

In the Table 10 above, Levene's Test for Equality of Variances for post-test 2 results produced an F-value of 1.565 and a significance level (p=0.216), indicating that the assumption of equal variances is satisfied. Thus, the "Equal Variance Assumed" row is used for interpretation. The t-test for equality of means showed a t-value of -3.668 with 55 degrees of freedom and a significance level (p=0.001), which is less than 0.05. This indicates a statistically significant difference in the post-test 2 mean scores between the two groups. The mean difference was -18.390, suggesting that one group scored, on average, 18.390 points lower than the other group. The 95% confidence interval for the mean difference ranges from -28.439 to -8.342, further supporting the statistical significance of the result, as the interval does not include zero.

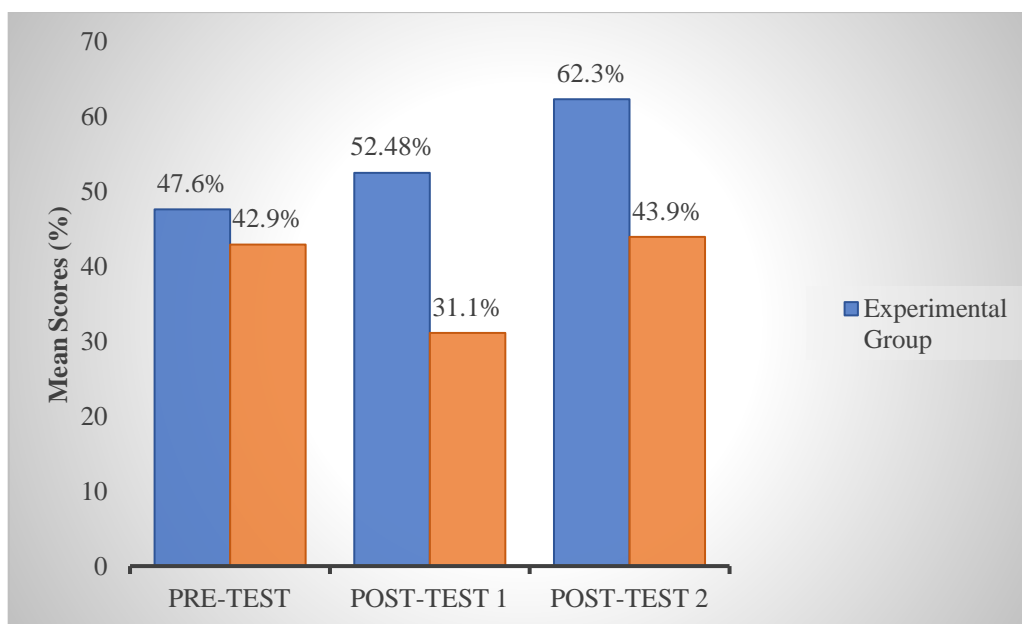**The Relationship of Test Results between Experimental Group and Control Group**



Figure 1: The Relationship of Test Results between Experimental Group and Control Group

In this study, it has been shown from figure 1 that the mean score from a post-test and end topic test results that the experimental group achieved significantly higher performance than the control group. The significant

difference in mean scores between the two groups is an indication that End of topic test is an effective assessment tool in Integrated Science at secondary school level.

## The Attitude of Pupils towards End-of-Topic Tests

The thematic analysis of pupils' responses revealed several key themes regarding their attitudes toward end-of-topic tests. These themes highlight the perceived benefits, challenges, and overall impact of this assessment approach on pupils' academic behaviours and attitudes.

### Theme 1: Motivation to Study and Continuous Revision

Many pupils expressed that end-of-topic tests motivated them to engage in regular study and revision. Statements such as *"It helps me to be continuously studying because tests are given at the end of each topic"* and *"It helps me study hard, especially when I fail"* demonstrate how these tests encourage consistent academic effort.

This finding aligns with research that emphasizes the role of formative assessments in promoting sustained engagement with the subject matter. The regularity of end-of-topic tests creates a structured environment for study, reducing the likelihood of last-minute revision and fostering deeper understanding.

### Theme 2: Enhanced Understanding and Mastery of Concepts

Pupils indicated that the tests contributed to a better understanding of the topics, helping them identify and address gaps in their knowledge. For instance, one pupil stated, *"End-of-topic tests made me confident to answer questions and help my friends when revising."* Another added, *"It gives me an opportunity to revise and know things and concepts before writing the mid-term test and end-of-term test."*

This suggests that the tests do not only aid individual learning but also facilitate collaborative learning through peer discussions and revisions. Such practices are vital for mastering complex topics in Integrated Science, as they allow pupils to revisit and reinforce their knowledge systematically.

### Theme 3: Development of Confidence and Resilience

The analysis revealed that end-of-topic tests fostered confidence in pupils. For example, one pupil remarked, *"I have improved,"* while another shared, *"End-of-topic tests made me confident to answer questions."* These responses suggest that regular testing builds resilience and self-assurance in handling assessments.

However, some pupils also reported feelings of nervousness and fear, as one stated, *"I sometimes feel scared and nervous to write end-of-topic tests, but they have been helpful."* This duality highlights the need to balance the pressure of frequent testing with supportive feedback mechanisms to ensure that pupils view tests as constructive rather than punitive.

# DISCUSION OF THE FINDINGS

## Effect of End-of-Topic tests on Pupils` Performance in Integrated Science

To test for the research objective as to whether the use of end of topic test had significant impact on the performance of pupils in Integrated Science, an independent sample t-test was set in two tailed normal distribution at significance level, α =0.05 for the pre-test scores and post-test two while post-test one was not normally distributed, hence, Mann-Whitney U test was conducted. According to Table 2, the descriptive statistics of the pre-test revealed that the mean of the control group was 42.88% and 47.61% for the experimental group with a mean score difference of 4.73%. The first section of Table 3 highlights Levene's test for equality of variances, which assesses whether the variances of the two groups are equal. According to Levene (1960), levene's test is an inferential statistic used to assess the equality of variances for a variable, calculated for two or more groups. With a null hypothesis ($H_0$) stating that the variances are equal ($\sigma_1^2 = \sigma_2^2$ ) and a significance level (alpha) of 0.05, the p-value was 0.112 (p= 0.112 > 0.05). Since the p-value exceeded

0.05, the null hypothesis was not rejected, indicating no significant difference in variances between the groups. Thus, the analysis relied on the row assuming equal variances.

The second section shows the t-test for equality of means, where the p-value was 0.229 (p= 0.229 > 0.05, t=−1.215). Again, the null hypothesis ($H_0$) was not rejected, indicating no statistically significant difference in pre-test scores between the groups. The 4.73% difference in mean scores was not significant and likely occurred by chance, confirming that the experimental and control groups were equivalent before the treatment.

The post-test one was a mid-term test (first post test) given to both experimental group and control group which the researcher wanted to know short term impact of end of topic test on pupil`s performance in Integrated Science. From the Figure 1, the experimental group outperformed the control group in the post-test one. The experimental group`s mean score was 52.5% while the mean score for the control group was 31.1%. The post-test one scores were not normally distributed, thus, the researcher proceeded with non parametric test (Mann-Whitney U) and the U-value was 151.000, with a corresponding Z-score of -3.438 and an asymptotic significance level (p) of 0.001. Since p=0.001 was less than the threshold of 0.05, we rejected the null hypothesis, concluding that there was a statistically significant difference in the post-test 1 scores between the experimental and control groups. The negative Z-score indicates that the control group scored lower on average compared to the experimental group. These results confirm that the difference in performance between the two groups is unlikely to have occurred by chance, highlighting the impact of the intervention introduced to the experimental group.

The findings from post-test two was the end of term test administered to both control and experimental groups. Table 9 shows the descriptive statistics of which the mean score for control group was 43.9% while the mean score of the experimental group was 62.3% implying that the experimental group performed better than the control group with a difference of 18.4%. After comparing the two means from the post-test two results using an independent sample t-test at significant level with alpha (α) =0.05, table 4.3.3 showed that the P-value (sig) was 0.001 which was less than 0.05, (P=0.001 < α=0.05, t=-3.668). Since P < 0.05, then the null hypothesis was rejected and concluded that there was a statistically significant difference between the experimental group and the control group and the mean difference (18.4%) was not by chance. This implied that using end-of-topic tests to assess Integrated Science has a long-term positive impact on pupils' academic performance than assessing Integrated Science using the conversional forms of assessment.

These findings support the findings of a research study by (Roediger III & Karpicke, 2006) which found that frequent testing improves not only long-term knowledge but also the application of the very knowledge. The improved results of the experimental group in integrated science confirms the quality of tests and their alignment with the curriculum objectives as observed by Harlen (2013).

**The Attitude of Pupils towards End-of-Topic tests**

The findings indicate that end-of-topic tests have a largely positive impact on pupils' attitudes and academic behaviours. By promoting consistent study habits, enhancing conceptual understanding, and building confidence, these tests contribute significantly to academic performance in Integrated Science.

However, the emotional responses, including fear and nervousness, underscore the need for a supportive testing environment. Educators should ensure that the feedback provided after tests is constructive and emphasizes growth rather than failure. Moreover, integrating stress management strategies, such as test preparation workshops, could help alleviate pupils' anxiety.

The use of Bloom's taxonomy in structuring test questions was a notable strength, ensuring that pupils engage with the material at multiple cognitive levels. This practice not only prepares pupils for high-stakes exams but also nurtures critical thinking and problem-solving skills, which are essential for academic success in science.

The research findings are aligned with studies done by many researchers such as (Roediger III & Karpicke, 2006; Harlen W. , 2013; Ministry of Education, 2013; Brown, Roediger III, & McDaniel, 2014; Mulenga-Hangane, Daka, Msango, Mwelwa, & Kakupa, 2019).

# CONCLUSION

The study explored the effect of end-of-topic tests on pupils' performance in Integrated Science and their attitudes toward this assessment method. The findings revealed that end-of-topic tests significantly enhanced pupils' academic performance, with the experimental group consistently outperforming the control group in both short-term (mid-term) and long-term (end-of-term) assessments. The statistically significant differences in scores underscore the effectiveness of regular and structured testing in improving knowledge retention and application. Additionally, pupils expressed largely positive attitudes toward end-of-topic tests, noting benefits such as improved study habits, better conceptual understanding, and increased confidence. However, some emotional challenges, including fear and nervousness, were highlighted, indicating the need for a supportive testing environment. The study further demonstrated the importance of aligning test design with curriculum objectives and incorporating Bloom's taxonomy to address multiple cognitive levels, fostering critical thinking and problem-solving skills. Overall, the research confirmed that end-of-topic tests are a valuable tool for enhancing both academic performance and positive learning behaviors when implemented thoughtfully and supported by constructive feedback and stress management strategies.

# RECOMMENDATIONS

Based on the findings in the study, the following recommendations were proposed:

- Schools should consider incorporating end-of-topic tests as a regular assessment method to improve academic performance and encourage consistent study habits.

- Teachers should ensure that test feedback emphasizes growth and areas of improvement rather than failure, fostering a positive learning experience.

- Tests should be aligned with curriculum objectives and designed using Bloom's taxonomy to assess various cognitive levels, ensuring comprehensive evaluation and critical thinking development.

- Conduct similar studies in other subjects or educational contexts to confirm the generalizability of the findings and refine best practices for implementing formative assessments.

# ACKNOWLEDGEMENT

# REFERENCES

1. Airasian, P. W. (1991). Classroom Assessment. New York: McGraw-Hill.
2. Berry, R. (2009). Assessment for Learning.Hong Kong: Hong Kong University Press. https:doi.org/10.5790/hongkong/9789622099579.001.0001
3. Brookhart, S. M. (2017). How to use grading to improve learning.
4. Brown, P. C., Roediger III, H. L., & McDaniel, M. A. (2014). Make it Stick: The Science of Successful Learning.
5. Bryman, A. (2013). Social Research Methods (4th ed.). United Kingdom: Oxford University Press.
6. Campbell, D. T., & Stanely, J. (1963). Experimental and quasi-experimental designs for research.
7. Cresswell, J. (2009). Research Design:Qualitative, Quantitative and Mixed Approaches. Sage Publication.
8. Hamtini, T., Albasha, S., & Varoca, M. (2015, February 10). Towards Designing an Intelligent Educational Assessment tool. Journal of Software Engineering and Applications, 8(2).
9. Harlen, W. (2013). Assessment & Inquiry-based science education: Issue in Policy and practice. Global Network of Science Academics.

10. Harlen, W., & James, M. (1997). Assessment and Learning: Differences and Relationships between Formative and Summative Assessment. Assessment in Education, 4, 365-379. doi:10.1080/0969594970040304

11. Kapambwe, W. M. (2006). Formative evaluation of the implementation of Continuos Assessment Pilot Program (CAPP) at Basic School Level in Zambia.

12. Kubsch, M., Czinczel, B., Lossjew, J., Wyrwich, T., Bednorz, D., Bernholt, S., . . . Rummel, N. (2022, 08 22). Toward learning progression analytics-Developing learning environments for the automated analysis of learning using evidence centered design. Frontiers in Education, 7. doi:10.3389/feduc.2022.981910

13. Luft, P., & Tiene, D. (2001). Teaching in a technology-rich classroom. Educational Technology, 23-31.

14. McAlpine, M. (2002). Principles of Assessment.

15. Ministry of Education. (2013). Zambia Education Curriculum Framework 2013. Lusaka: Ministry of Education.

16. Ministry of General Education. (2017). National Learning Assessment Framework (NLAF). Lusaka: Ministry of General Education.

17. Ministry of General Education. (2017). National Learning Assessment Framework(NLAF). Lusaka: Ministry of General Education.

18. Mulenga-Hangane, M., Daka, H., Msango, H. J., Mwelwa, K., & Kakupa, P. (2019, June 10). Formative Assessment as a Means of Improving Learners` Achievement: Lessons from Selected Primary Schools of Lusaka, Zambia. Journal of Lexicography and Terminology, 3(1).

19. Nitko, A. J. (2004). Continuous Assessment and Performance Assessment. Retrieved January 3, 2025, from http://www.moec.gov.jm.pdf

20. Olufemi, A. S. (2014). Relationship between continuous assessment and junior school certificate examination mathematics scores in Ekiti state. International Journal of Liberal Arts and Social science, 2(6).

21. Osborne, J., & Dillon, J. (2010). Good peractice in science teaching: What research hasto say.

22. Quansah, K. B. (2005). Continuous Assessment Handbook. Accra: BECAS.

23. Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. Psychological Science, 17(3), 249-255.

24. Stiggins, R. (2007). Assessments through the students` eye. Educational Leadership, 8(64), 22-26. Retrieved December 20, 2024

25. Tariq, A. A., & Ali, A. (2019, June 31). Correlation between Internal and External Assessment at University level: A Case Study of I.E.R, University of Peshawar. Al-Idah Bi-Annual Research Journal, 37(I).

26. Tariq, A., & Ali, A. (2019). Correlation between Internal and External Assessment at University Level: A Case Study of I.E.R, University of Peshawar.

27. Trochim, W. M. (2006). The Qualitative Debate. Research Methods Knowledge Base. Retrieved from http://www.socialresearchmethods.net/kb/qualmeth.php

28. UNESCO. (2017). Education for sustainable development goals:Learning Objectives.