

Original Article

An Evaluation of Research Colloquium Series in Senior High School Using Kirkpatrick's Four-Level Model

Gerald Ello ^{1,*}, Lemuel Ramos ²

Received: 29 May 2026; Revised: 12 July 2026;

Accepted: 14 July 2026; Published: 15 July 2026

DOI: <https://doi.org/10.66074/P8W9C2T6>

Abstract

This study evaluated the effectiveness of the Research Colloquium Series conducted in a private school in the Philippines during academic year 2024–2025 using Kirkpatrick's Four-Level Evaluation Model. The series included seven events designed to strengthen research culture and improve student competencies in statistical analysis, research ethics, and research dissemination. Using a mixed-methods design, the study gathered data through post-session surveys, assessments, document analysis, and interviews. Quantitative data were analyzed using descriptive statistics, while qualitative data were examined through thematic analysis. Findings showed positive outcomes across all four levels. Under Level 1 (Reaction), students expressed high satisfaction with the sessions, although more technical topics such as AI and Arduino were perceived as challenging. Under Level 2 (Learning), students attained a high post-training knowledge level based on post-session assessments, although the absence of a pre-test baseline limits causal interpretation. At Level 3 (Behavior), many student groups were able to conduct data analysis independently, while Grade 11 students reported increased motivation to participate in research. Under Level 4 (Results), several groups presented at conferences and submitted research for publication. Overall, the findings suggest that the colloquium series contributed positively to strengthening the research culture among senior high school students; however, the evaluation design, which lacked baseline and comparison measures—warrants cautious interpretation of these outcomes.

¹ City College of Tagaytay, Philippines

² Unida Christian Colleges, Philippines
* sircurts.educ@gmail.com

Keywords: program evaluation, research colloquium, senior high school research

1. Introduction

Research education in senior high school (SHS) is a cornerstone of the Philippine K to 12 curriculum, aimed at cultivating critical thinking, problem-solving, and inquiry-based learning among students. In addition to the core research subjects Practical Research 1 and Practical Research 2, several other subjects, including Inquiry, Investigation and Immersion, Work Immersion, and various Research Projects in Specialized Tracks—integrate research processes to prepare students for higher education and real-world challenges (DepEd, 2016). Despite this strong curricular emphasis, many SHS learners still struggle to translate theoretical knowledge into applied research competencies due to limited exposure, insufficient mentoring, and underdeveloped academic culture within schools.

Several studies have documented this gap. For instance, Pagulayan et al. (2022) found that while students exhibit positive attitudes toward research, they face significant challenges in understanding methodologies, statistical analysis, and academic writing—largely due to lack of structured mentoring and limited research immersion. Similarly, Cruz and Rafael (2021) noted that senior high school students often encounter difficulties completing their research projects independently, attributing these to gaps in teacher preparedness and insufficient institutional support for research integration. In the case of STEM tracks, Basilio and Ramos (2020) emphasized the need for continuous professional development and collaborative activities to bridge students' cognitive limitations with hands-on research exposure. These studies highlight the importance of sustained initiatives, such as colloquium series, research mentoring programs, and skills training workshops, to build a meaningful and sustainable research culture in SHS settings.

To bridge these gaps, one of the private schools located in Metro Manila, Philippines implemented a Research Colloquium Series during the first semester of School Year 2024–2025 and continuing into the second semester. This initiative comprised seven key events, each focusing on the development of research skills, values, and scholarly engagement among students. The series began with sessions on the application of Arduino and Artificial Intelligence (AI) in STEM research and prototyping, followed by a training workshop on statistical reliability analysis using JAMOWI software. A three-part statistics series focused on organizing data and conducting both correlational and comparative tests. Subsequently, two additional episodes were introduced: one on Research Ethics and Plagiarism, emphasizing academic integrity and responsible research conduct; and another on Research Conferences and Publications, highlighting dissemination pathways and scholarly communication. However, these last episodes were not conducted in a plenary format; instead, they were discussed by each research teacher in their respective classes. These two episodes were retained as part of the colloquium series because they were

designed, scheduled, and delivered as components of the same institution-wide research capacity-building program and shared its learning objectives in research ethics and dissemination; the shift to a class-based format was an implementation adjustment rather than a change in program scope. For transparency, they are nonetheless treated separately in the evaluation: because they were not delivered in plenary format, they were not covered by the post-session satisfaction surveys reported under Level 1, and their corresponding outcomes are instead reflected in the Level 3 and Level 4 indicators.

A robust research culture is characterized not only by the conduct of research activities but also by the shared values, expectations, and institutional practices that support inquiry and innovation (Macfarlane & Hughes, 2009). Building this culture at the SHS level is essential for fostering lifelong learning, academic resilience, and readiness for university-level scholarship (Brew & Mantai, 2017). This colloquium series was thus envisioned as a capacity-building intervention to reinforce this culture by deepening student engagement in research and promoting high-quality outputs through expert-led sessions and collaborative learning.

This study evaluates the effectiveness of the research colloquium series using Kirkpatrick's Four-Level Evaluation Model (Kirkpatrick & Kirkpatrick, 2006), a well-established framework for assessing educational programs. The model investigates outcomes at four levels: participants' reactions, their learning gains, observed behavioral changes, and the results or impact of the intervention. Employing this model allows for a comprehensive assessment of the initiative's short- and long-term influence on student research competencies and the evolving research culture within the senior high school.

The following research questions were formulated to guide the investigation, drawing from student survey responses, assessment scores, research outputs, and institutional records:

Level 1: Reaction

(1) What is the level of student satisfaction with the content of each research colloquium session? (2) What is the level of student satisfaction with the delivery of each research colloquium session? (3) What initial feedback did the students give for each research colloquium?

Level 2: Learning

(4) What are the students' assessment scores related to the colloquium sessions?

Level 3: Behavior

(5) How many students performed data analysis independently after attending the colloquium series? (6) How do Grade 11 students perceive the influence of Grade 12 students' research activities on their own interest in doing research?

Level 4: Results

(7) How many students presented their research in academic conferences after the colloquium series? (8) How many student research papers were submitted for publication after the colloquium series?

2. Methodology

2.1 Evaluation Design

This study employed a mixed-methods evaluation design grounded in Kirkpatrick's Four-Level Evaluation Model (Kirkpatrick & Kirkpatrick, 2006), which provides a structured framework for assessing the effectiveness of training and educational interventions. The model evaluates programs across four levels: (1) Reaction – participants' satisfaction and immediate responses to the program; (2) Learning – the extent of knowledge or skill acquisition; (3) Behavior – the application or transfer of learning to practice; and (4) Results – the broader impact on the organization or system.

Kirkpatrick's model remains one of the most widely used frameworks in training evaluation, emphasizing both outcomes and behavioral change (Bates, 2004; Kirkpatrick & Kirkpatrick, 2006). To evaluate the research colloquium series, the study employed a quantitative descriptive design and qualitative thematic analysis. The quantitative component involved post-colloquium surveys and assessment tools that measured satisfaction, knowledge gains, and performance outcomes. Descriptive research, as defined by Creswell and Creswell (2017), is appropriate for identifying observable trends and determining program effectiveness based on feedback and measurable results. In parallel, qualitative data were collected through open-ended responses and reflective prompts, which were analyzed using thematic analysis, a method for identifying and interpreting patterns within data (Braun & Clarke, 2006). By integrating both quantitative and qualitative approaches, the study provided a comprehensive evaluation of the Research Colloquium Series, enabling a multi-dimensional understanding of its effectiveness across all levels of Kirkpatrick's model.

2.2 Participants

The primary participants in this study were Grade 12 students enrolled, who attended one or more of the seven research colloquium sessions conducted throughout the academic year 2024–2025. Attendance per session ranged from approximately 120 to 150 students. A total of 494 Grade 12 students were officially enrolled, of which 214 students completed the entire colloquium series, representing 43.32% of the cohort. Among those who completed the series, 171 were from the STEM strand (53.11%) and 43 from the non-STEM strands (25%). For clarity, the two participation figures used in this study index different forms of involvement and are not interchangeable. The 494 students refer to the total enrolled Grade 12 cohort who sat

the post-session knowledge assessment reported under Level 2, which was administered to all enrolled students regardless of how many sessions they had attended. The 214 students refer specifically to those who attended the complete sequence of colloquium sessions. This distinction is made explicit because it affects how the Level 1 and Level 2 results should be interpreted. The completion rate of 43.32% indicates that fewer than half of the enrolled students attended the full series. This attrition is attributable primarily to the voluntary nature of participation and to scheduling overlaps with other academic and co-curricular commitments. Because students who persisted to the end of the series may have been more motivated or research-oriented than those who discontinued, the possibility of self-selection (survivor) bias is acknowledged. This bias may have contributed to the high satisfaction ratings reported under Level 1 and is discussed further in the Conclusion and Limitations section.

Table 1. Number of Grade 12 student participants.

Strand	Total Number of Grade 12 Students	Number Who Completed the Colloquium Series (Percent)
STEM	322	171 (53.11%)
Non-STEM	172	43 (25.00%)
Total	494	214 (43.32%)

In addition to the Grade 12 participants, Grade 11 students were also involved in the evaluation at Kirkpatrick's Level 3 (Behavior). A subset of Grade 11 students was purposively selected to participate in semi-structured interviews, focusing on how the research activities of their senior peers influenced their motivation and intention to engage in research in the future. These responses contributed to the qualitative component of the study and provided further insight into the broader behavioral and cultural impact of the colloquium series across grade levels.



Figure 1. Kirkpatrick's four-level evaluation model – illustrating reaction, learning, behavior, and results.

2.3 Instruments

To evaluate the effectiveness of the research colloquium series, the study employed both quantitative and qualitative instruments aligned with Kirkpatrick's Four-Level Evaluation Model. For Level 1 (Reaction), students completed post-session surveys that included Likert-scale items measuring satisfaction with content and delivery, as well as open-ended prompts for qualitative feedback. These responses were thematically analyzed to gain insights into students' perceptions and immediate reactions. For Level 2 (Learning), post-session tests were used to assess students' comprehension of topics such as data organization, statistical tools, ethical research practices, and dissemination. While no pre-tests were conducted, test scores were analyzed descriptively to evaluate mastery and learning gains. The satisfaction survey and the post-session knowledge assessment were developed by the research team on the basis of the content and learning objectives of each colloquium session. Prior to administration, both instruments were reviewed by the research teachers for content validity and clarity, and items were refined accordingly. The knowledge assessment was scored out of a maximum of 50 points, with all items weighted equally.

At Level 3 (Behavior), behavioral transfer was operationalized as the proportion of student research groups that completed their data analysis independently—that is, without direct intervention from teachers or facilitators. Independent completion was determined through the research teachers' judgment rather than any scoring rubric: a group was counted as having completed its data analysis independently when its members performed the analysis themselves instead of engaging an external statistician. Only this binary determination was recorded and analyzed for the present study. Interviews with Grade 11 students were conducted to explore how observing Grade 12 research influenced their own motivation and readiness to conduct research. Finally, Level 4 (Results) involved analyzing institutional records to determine how many student groups participated in research conferences or had their work accepted for publication, serving as indicators of long-term engagement and research capacity at the institutional level.

2.4 Data Analysis

Descriptive statistics such as means, frequencies, and percentages were used to analyze quantitative data from surveys and post-tests. Thematic analysis was applied to open-ended survey responses to identify themes in student feedback. The thematic analysis followed the six-phase procedure described by Braun and Clarke (2006): familiarization with the open-ended survey responses and interview transcripts, generation of initial codes, searching for themes, reviewing themes, defining and naming themes, and reporting. The semi-structured interviews involved a total of 25 purposively selected Grade 11 students. Coding was carried out by a single coder. Illustrative codes for each theme are presented in the Results section to support the

trustworthiness of the analysis. This multi-source, multi-method approach supports a comprehensive and replicable evaluation of the colloquium series.

2.5 Ethical Considerations

Ethical considerations were strictly observed throughout the conduct of the study. Participation was voluntary, and informed consent was obtained from all student participants through a digital agreement form embedded in the post-session survey. Data collected were anonymized to protect student identity and confidentiality. No personally identifiable information was used in the analysis or reporting of results. The study also adhered to institutional guidelines for ethical conduct in educational research, and necessary permissions were secured from the school administration prior to data collection. These ethical safeguards ensured that the evaluation respected the rights and well-being of all participants.

3. Results

This section presents the findings of the evaluation conducted on the Research Colloquium Series guided by Kirkpatrick's Four-Level Evaluation Model. The results are organized according to the four levels: Reaction, Learning, Behavior, and Results. Data were collected through post-colloquium surveys, post-session assessments, evaluation of student research outputs, and institutional records of conference participation and publication submissions.

3.1 Level 1: Reaction

Level 1 of Kirkpatrick's Evaluation Model focuses on participants' immediate reactions, particularly their level of satisfaction with the content and delivery of the training sessions. In this study, student feedback was collected after each research colloquium session to assess how well the content addressed their academic needs and how effectively the sessions were delivered. It should be noted that the satisfaction ratings reported in Tables 2 and 3 cover the five sessions that were delivered in plenary format and for which post-session surveys were administered. The two remaining episodes—Research Ethics and Plagiarism, and Research Conferences and Publications—were delivered within regular classes rather than as plenary sessions and were therefore not covered by the plenary satisfaction surveys; their contribution to the program is instead reflected in the Level 3 and Level 4 indicators. This distinction is made explicit to avoid conflating the seven program events with the five sessions on which the reaction-level data are based.

As reflected in Table 2, the overall mean rating for content across all sessions was 4.33, verbally interpreted as "Highly Satisfied." This indicates that students perceived the content of the colloquia to be relevant, informative, and applicable to their research tasks. The highest content satisfaction was observed in the Training Workshop on Reliability Analysis using JAMOV (mean = 4.62) and Episode 1: Data

Organization (mean = 4.61), which suggests strong alignment with students' current learning needs in research methodology. The consistently high ratings across the statistical sessions (Episodes 1 to 3) point to the importance of foundational quantitative research skills in the SHS research curriculum.

Table 2. Level of student satisfaction with the content of each research colloquium session.

Colloquium Session	Mean	Verbal Interpretation
STEM Research Colloquium: Application of AI and Arduino Robotics Platform in STEM Research and Prototyping	3.45	Satisfied
Training Workshop on Reliability Analysis using JAMOVİ	4.62	Highly Satisfied
Research Colloquium Series – Episode 1: Data Organization	4.61	Highly Satisfied
Research Colloquium Series – Episode 2: Statistical Tests of Relationship	4.53	Highly Satisfied
Research Colloquium Series – Episode 3: Statistical Tests of Difference	4.42	Highly Satisfied
Overall Mean	4.33	Highly Satisfied

Meanwhile, Table 3 shows that students were also highly satisfied with the delivery, averaging a mean of 4.33. This reflects strong facilitation, organization, and communication during the sessions. The highest delivery rating was given to Episode 1: Data Organization (mean = 4.54), further confirming the effectiveness of this session in both substance and delivery. Notably, the STEM Research Colloquium on AI and Arduino received the lowest ratings for both content (mean = 3.45) and delivery (mean = 3.37), though still within the "Satisfied" range. This suggests that while students appreciated the session, its technical nature may have posed comprehension challenges, or the delivery methods may have lacked engagement strategies suitable for the audience.

These findings demonstrate that the Research Colloquium Series was well received by the students. The high satisfaction ratings confirm the program's success in meeting its objective of strengthening the research culture among SHS learners through meaningful and well-facilitated academic engagements. However, the lower ratings for the STEM-focused session highlight areas for improvement, particularly in making complex, technology-driven topics more accessible and engaging a broader group of learners.

Table 4 presents a thematic summary of students' initial feedback from the research colloquium series, aligned with Kirkpatrick's first level, which evaluates participants' immediate responses in terms of satisfaction, engagement, and perceived relevance of the training sessions.

The first theme, Effective Delivery, reflects students' satisfaction with how the colloquium sessions were conducted. Students valued clarity, engagement, and the professional conduct of the sessions. The second theme, Relevant, highlights that students found the colloquium topics timely and applicable to their ongoing Practical

Table 3. Level of student satisfaction with the delivery of each research colloquium session

Colloquium Session	Mean	Verbal Interpretation
STEM Research Colloquium: Application of AI and Arduino Robotics Platform in STEM Research and Prototyping	3.37	Satisfied
Training Workshop on Reliability Analysis Using JAMOV1	4.51	Highly Satisfied
Research Colloquium Series – Episode 1: Data Organization	4.54	Highly Satisfied
Research Colloquium Series – Episode 2: Statistical Tests of Relationship	4.41	Highly Satisfied
Research Colloquium Series – Episode 3: Statistical Tests of Difference	4.39	Highly Satisfied
Overall Mean	4.33	Highly Satisfied

Research 2 work, particularly in data analysis and methodology. These ratings confirm that clear, organized, and engaging delivery styles contributed to student satisfaction. However, sessions with more technical content may require additional support strategies to achieve the same level of student engagement and understanding.

The third theme, Complexity of Technical Topics, identifies challenges students encountered in more advanced or STEM-focused sessions, such as those involving AI, Arduino, and statistical tools. Many students felt cognitively unprepared due to a lack of prior exposure to engineering, programming, or robotics concepts—especially those not enrolled in STEM tracks. The final theme, Limited Time and Practical Application, underscores student feedback regarding the pacing of sessions and the need for more illustrative examples.

Table 4. Thematic analysis result on students' initial feedback for each research colloquium.

Themes	Sample Codes
Effective Delivery	"The speaker was very engaging." "The discussion was clear and concise." "The speaker is good at explaining the topic." "It was a great session; I learned a lot." "The resource speaker gave clear instructions and examples."
Relevance	"The topic is very relevant to our ongoing research." "I finally understood how to use statistical tools in our data analysis." "This is useful for our research output." "Applicable ito sa Chapter 4 namin." (<i>This is applicable to our Chapter 4.</i>)
Complexity of Technical Topics	"Mahirap yung Arduino part, di ko masundan." (<i>The Arduino part was difficult; I could not follow it.</i>) "The AI and Arduino part was very confusing for those of us who are not in STEM." "It was too technical for me." "I think this topic is too advanced for senior high school students."
Limited Time and Practical Application	"Bitin yung session." (<i>The session felt too short.</i>) "More time sana para mas maraming matutunan." (<i>I wish there had been more time so we could learn more.</i>) "Sana mas marami pang examples." (<i>I wish there had been more examples.</i>) "Mas okay siguro kung may actual sample ng computation." (<i>It would have been better if an actual sample computation had been provided.</i>)

These qualitative data indicate that students generally responded positively to the research colloquium series, especially in terms of speaker effectiveness and topic relevance. However, areas for improvement include the need for more time, clearer

examples, and differentiated support for technical content. Addressing these points in future implementations can enhance both comprehension and engagement, further strengthening the program's impact at the reaction level.

3.2 Level 2: Learning

Table 5 presents the descriptive statistics of the students' assessment scores collected after the research colloquium sessions. A total of 494 students participated in the post-assessment, which was designed to measure their post-training knowledge level in relation to the topics discussed during the sessions. Because no pre-test was administered, these scores represent only post-training performance; in the absence of a baseline measure, they should be read as indicators of the knowledge level demonstrated after the sessions rather than as evidence of knowledge gain attributable to the colloquium series.

Table 5. Descriptive statistics of students' assessment scores

N	Min	Max	Mean	SD
494 students	10	50	40.63	8.3143

The scores ranged from 10 to 50, with 50 as the highest possible score. The mean score was 40.63, indicating that on average, students correctly responded to approximately 81% of the assessment items, which reflects a high post-training knowledge level with respect to the content delivered during the colloquium series. Because no pre-test was administered, however, these results cannot establish that the colloquium series caused the observed knowledge level; students may, for instance, have acquired part of this knowledge through their regular Practical Research 2 classes prior to the sessions. The findings are therefore interpreted as evidence of a high post-training performance under Level 2 of Kirkpatrick's Model, rather than as proof of knowledge acquisition attributable solely to the program.

3.3 Level 3: Behavior

Table 6. Number of groups that performed data analysis independently after attending the colloquium series.

Total Number of Groups	Groups Performed Data Analysis Independently	Percent
75 groups	45 groups	60.0%

Table 6 presents the number and percentage of student research groups who independently performed data analysis after attending the research colloquium series. Out of a total of 75 student groups, 45 groups (60.0%) successfully conducted data analysis on their own, without direct intervention or assistance from teachers or facilitators. This outcome demonstrates a moderate-to-high level of behavioral transfer, as defined in Kirkpatrick's Level 3. The act of independently performing data analysis signifies that the participating students not only understood the statistical

concepts introduced in the colloquia but were also able to apply them meaningfully in the conduct of their own research.

Table 7 presents the thematic analysis of Grade 11 students' perceptions on the influence of Grade 12 students' research activities, interpreted through the lens of Kirkpatrick Level 3. The first theme, *Admiration*, highlights how Grade 11 students were inspired by the confidence, creativity, and professionalism displayed by their senior peers. The second theme, *Real-World Relevance Awareness*, indicates that students have developed a clearer understanding of the practical value of research. The third theme, *Role Model*, underscores the influence of peer mentorship and observational learning, with students expressing that seeing their older peers successfully present research built their own confidence and interest in pursuing similar work.

Table 7. Thematic analysis result on Grade 11 students' perception on the influence of Grade 12 students' research activities.

Themes	Sample Codes
Admiration	"The part that inspired me the most was their confidence and creativity." "You could see how much effort they put into their work." "The way they speak and present gave a sense of professionalism." "How they presented their research is something I admire."
Awareness of Real-World Relevance	"It made me believe that there are alternative ways to ensure environmental sustainability." "Their projects showed how STEM can solve real problems." "Research can help solve a problem."
Role Modeling	"Seeing someone close to your age succeed makes it feel more achievable." "It builds confidence that we can do it too." "Nakakagana manood ng kapwa student na nagreresearch." (<i>Seeing fellow students conduct research is motivating.</i>)

3.4 Level 4: Results

In Kirkpatrick's evaluation model, Level 4: Results focuses on the final outcomes and institutional impact of a training or educational initiative. This includes changes in performance, productivity, or external recognition attributable to the intervention. As shown in Table 8, out of 75 total research groups, 17 groups (22.67%) successfully participated in external research conferences. This level of engagement reflects the students' ability to translate knowledge into practice, particularly in preparing and presenting scholarly work for external academic audiences.

Table 8. Number of groups accepted to and participated in research conferences.

Total Number of Groups	Groups Participated in Conferences	Percent
75 groups	17 groups	22.67%

Table 9 provides a breakdown of the research conferences in which students participated. The data shows that groups were accepted into five distinct research conferences, including both national and regional events hosted by reputable institutions. Notably, the De La Salle University-Dasmariñas High School Research

Conference had the highest participation with 9 groups, suggesting a strong institutional linkage and relevance of student research outputs to high-quality academic forums. This level of research dissemination illustrates the broader impact of the colloquium series beyond classroom learning, aligning with Brew and Mantai's (2017) assertion that active participation in research dissemination activities is a key marker of an emerging research culture.

Table 9. List of research conferences participated in by the students.

Research Conference	Institution / Location	No. of Groups
2025 Conference on Innovative Education, Research and Creative Works	iACADEMY Nexus Campus, Makati City	2
ONWARD 2025: 2nd High School Research Conference	FEU Cavite, Silang, Cavite	2
17th UPLB-CAS Student-Faculty Research Conference	UPLB, Los Baños, Laguna	1
2nd Dasmariñas Research Congress	KLD, Dasmariñas City, Cavite	3
5th De La Salle University-Dasmariñas High School Research Conference	DLSU-D, Dasmariñas City, Cavite	9
Total		17

Table 10 shows that out of 75 research groups, 7 groups (9.33%) successfully had their research papers accepted for publication. These papers were accepted by external, locally indexed journals, rather than by the school's own internal outlets. While the reach of these publications may be limited, the act of getting published reflects the students' ability to produce scholarly work that met the standards of formal academic dissemination. The acceptance into even peer-reviewed locally indexed illustrates the early stages of research culture development within the institution and suggests that the colloquium series was effective in contributing to long-term institutional goals such as promoting research productivity and public recognition of student scholarship.

Table 10. Number of groups accepted to research publications.

Total Number of Groups	Groups Accepted to Publications	Percent
75 groups	7 groups	9.33%

4. Discussion

The findings of this study reveal that the Research Colloquium Series produced meaningful and consistent outcomes across all four levels of Kirkpatrick's Evaluation Model, underscoring its value as a structured academic intervention in the senior high school setting.

4.1 Level 1: Reaction

The high satisfaction ratings across content (mean = 4.33) and delivery (mean = 4.33) confirm that the colloquium sessions were well-received overall. Sessions centered on practical statistical skills—such as data organization and reliability analysis using JAMOV—garnered the highest ratings, reflecting their direct applicability to students' ongoing Practical Research 2 requirements. This aligns with Kirkpatrick and Kirkpatrick's (2006) premise that positive reactions are most likely when participants perceive the training as immediately relevant to their tasks.

In contrast, the lower ratings for the STEM-focused session on AI and Arduino suggest that technical complexity and insufficient background preparation can diminish participant satisfaction even when content is intrinsically valuable. The qualitative themes of Complexity of Technical Topics and Limited Time and Practical Application further support this interpretation, pointing to the need for differentiated instruction and extended session formats when delivering advanced content to a mixed-strand audience.

4.2 Level 2: Learning

A mean post-assessment score of 40.63 out of 50 (approximately 81%) across 494 students indicates a high post-training knowledge level. This result is consistent with Bates' (2004) argument that well-designed training programs grounded in clear learning objectives are capable of producing measurable cognitive gains. The absence of pre-tests, however, limits the ability to attribute this knowledge level to the colloquium series, as students may have had varying degrees of prior knowledge. Future implementations should incorporate a pre- and post-test design to better isolate the instructional effect.

4.3 Level 3: Behavior

The finding that 60% of student research groups (45 out of 75) conducted data analysis independently represents a meaningful transfer of learning to practice—the core concern of Kirkpatrick's Level 3. This behavioral shift suggests that the colloquium series succeeded in building not only declarative knowledge but also procedural competence, enabling students to apply statistical tools without direct teacher facilitation. This is particularly significant in the Philippine SHS context, where Cruz and Rafael (2021) identified independent research completion as a persistent challenge.

The thematic analysis of Grade 11 students' perceptions further enriches the Level 3 findings. The emergence of themes such as *Admiration*, *Real-World Relevance Awareness*, and *Role Model* reveals that the behavioral impact of the colloquium series extended beyond the immediate participants. The observational learning experienced by Grade 11 students aligns with Bandura's (1977) social

learning theory, wherein vicarious reinforcement through peer role models fosters self-efficacy and behavioral intention. This cross-grade influence suggests the colloquium series functions as a culture-building mechanism that transcends its direct audience.

4.4 Level 4: Results

The institutional outcomes recorded at Level 4 provide encouraging evidence consistent with the colloquium series' intended long-term impact. The participation of 17 research groups (22.67%) in external conferences, including reputable venues such as UPLB, FEU Cavite, and DLSU-D, indicates that students produced work of sufficient quality to be accepted into peer-reviewed forums. This level of scholarly engagement is consistent with Brew and Mantai's (2017) assertion that active participation in dissemination activities is a hallmark of an emerging research culture. At the same time, these conference and publication outcomes should be interpreted with caution, as they may also have been shaped by factors beyond the colloquium series itself—such as individual teacher mentoring, the prior ability and motivation of the participating groups, the level of school support, and the likelihood that stronger groups self-selected into competitive submissions. These results are therefore best read as outcomes associated with the program rather than as effects attributable to it alone.

The acceptance of 7 research groups (9.33%) for publication, while modest in scope due to the local indexed nature of the outlets, nonetheless represents a meaningful institutional milestone. In a senior high school setting where publication is not yet a standard expectation, this outcome signals a positive trajectory toward research productivity. It also reinforces the argument of Pagulayan et al. (2022) and Basilio and Ramos (2020) that sustained, structured interventions are necessary to convert student research interest into tangible scholarly output.

Taken together, the results across all four levels indicate that the Research Colloquium Series effectively bridged the gap between theoretical research knowledge and applied academic practice, contributing meaningfully to the development of a research culture within the institution. The consistent alignment of findings with existing literature lends further credence to the program's design and implementation.

5. Conclusion and Limitations

The evaluation of the Research Colloquium Series implemented in senior high school showed meaningful outcomes across all levels of Kirkpatrick's Evaluation Model. Students reported high levels of satisfaction with the overall content and delivery of the sessions, particularly those focused on practical areas such as statistical testing, data organization, and reliability analysis. However, sessions involving more technically advanced topics, such as AI and Arduino, were perceived as more difficult by some students due to limited background knowledge. Evidence of learning was demonstrated through post-colloquium assessments, and many student research groups were able to apply their learning by conducting data analysis independently.

Furthermore, the qualitative feedback from Grade 11 students highlighted the influence of observing their Grade 12 peers, revealing themes of admiration, relevance to real-world issues, and the motivational effect of peer role modeling.

Several research groups went on to present their work in academic conferences and submit their outputs for publication, indicating that the colloquium series had a broader institutional impact on promoting a culture of research and scholarly engagement. Despite these positive outcomes, the study faced several limitations that temper the strength of its conclusions. First, the absence of a pre-test, a comparison group, and any baseline measure means that the design cannot establish that the observed outcomes were caused by the colloquium series; the findings are therefore descriptive and associational rather than causal. Second, participation was voluntary and fewer than half of the enrolled students (43.32%) completed the full series, which introduces the risk of self-selection or survivor bias—students who persisted to the end may have been more motivated or research-oriented than those who did not, potentially inflating the satisfaction and performance results. Third, most publication outputs were accepted by locally indexed journals, which may not fully reflect the scholarly potential of the students' work. Finally, several of the Level 3 and Level 4 outcomes may also have been shaped by factors outside the program, including teacher mentoring, prior student ability, and school support.

Considering these findings, several recommendations are proposed. Future sessions that involve complex or highly technical content should be supported with preparatory materials or prerequisite orientations. Implementing a pre- and post-test assessment structure is also advised to more accurately measure knowledge acquisition. To encourage broader participation, the colloquium series could be integrated more formally into the research curriculum. Moreover, support for research dissemination can be enhanced by guiding students toward indexed journals and competitive academic conferences. It is also recommended to sustain opportunities for junior students to observe and reflect on senior research activities, fostering early engagement and building a pipeline for research readiness.

Overall, the findings suggest that the Research Colloquium Series was a worthwhile academic initiative that was associated with positive outcomes in research competence, motivation, and participation in scholarly work among senior high school students. While the evaluation design does not permit causal claims, the consistent pattern of results across the four levels points to a promising contribution to the school's research culture. With continued refinement—particularly the addition of baseline measures—and sustained institutional support, the program has good potential to further strengthen this culture.

Acknowledgment

The authors acknowledge the support of the school administration, research teachers, and all student participants who made this evaluation possible. No external funding was received for this study. The corresponding author also expresses sincere

gratitude to Dr. Alvin Tenorio for his guidance and mentorship in the course *Theory and Methods of Evaluation*, under which this study was developed and completed as a course requirement and output.

Conflict of Interest Statement

The authors declare that they have no relevant financial or non-financial interests to disclose.

References

- Bandura, A. (1977). *Social learning theory*. Prentice Hall.
http://www.ascib.ase.ro/mps/Bandura_SocialLearningTheory.pdf
- Basilio, R. J., & Ramos, A. G. (2020). Enhancing research competence among STEM senior high school students: Issues and interventions. *Asia Pacific Journal of Multidisciplinary Research*, 8(1), 34–41.
- Bates, R. (2004). A critical analysis of evaluation practice: The Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning*, 27(3), 341–347.
<https://doi.org/10.1016/j.evalprogplan.2004.04.011>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brew, A., & Mantai, L. (2017). Academics' perceptions of the challenges and barriers to building research capacity in a teaching-focused institution. *Higher Education Research & Development*, 36(3), 453–467. https://www.researchgate.net/profile/Lilia-Mantai/publication/312097617_Academics%27_perceptions_of_the_challenges_and_barriers_to_implementing_research-based_experiences_for_undergraduates/links/5c7a1e37299b1268d30b5cc/Academics-perceptions-of-the-challenges-and-barriers-to-implementing-research-based-experiences-for-undergraduates.pdf
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
[https://books.google.com/books?hl=en&lr=&id=335ZDwAAQBAJ&oi=fnd&pg=PT16&dq=+Creswell,+J.+W.,+%26+Creswell,+J.+D.+\(2018\).+Research+design:+Qualitative,+quantitative,+and+mixed+methods+approaches+\(5th+ed.\).+SAGE+Publications.&ots=YEASOOrnsJ&sig=8aWU_usyETVvGc9l3klr4zXk4QQ](https://books.google.com/books?hl=en&lr=&id=335ZDwAAQBAJ&oi=fnd&pg=PT16&dq=+Creswell,+J.+W.,+%26+Creswell,+J.+D.+(2018).+Research+design:+Qualitative,+quantitative,+and+mixed+methods+approaches+(5th+ed.).+SAGE+Publications.&ots=YEASOOrnsJ&sig=8aWU_usyETVvGc9l3klr4zXk4QQ)
- Cruz, R. M., & Rafael, R. M. (2021). Challenges encountered by senior high school students in completing research outputs. *International Journal of Research Studies in Education*, 10(4), 52–63.
<https://doi.org/10.5861/ijrse.2021.202042>
- Department of Education (DepEd). (2016). *Senior High School Curriculum Guides: Practical Research 1 and 2; Inquiry, Investigation and Immersion*.
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.). Berrett-Koehler.
https://www.academia.edu/download/36183859/Evaluating_Training_Programs_EXCERPT.pdf

Macfarlane, B., & Hughes, G. (2009). Turning teachers into academics? The role of educational development in fostering synergy between teaching and research. *Innovations in education and teaching international*, 46(1), 5-14.
<https://www.tandfonline.com/doi/abs/10.1080/14703290802646214>

Pagulayan, J. M., Salazar, M. P., & De Vera, J. B. (2022). Exploring the difficulties of research writing among senior high school students in the Philippines. *Journal of Education and Practice*, 13(17), 15–24.

Author Contributions: Ello, G.; Study design, method conception, data analysis, and manuscript writing. Ramos, L.; Implementation of colloquia, data collection, critical review, and editing.