

Effectiveness of a Research Capability Training Program: A Kirkpatrick's Evaluation Model Application in Cebu, Philippines

Robert Jay N. Angco^{1*} 

Department of Mathematics and Statistics, Cebu Technological University¹, Cebu City, 6000, Philippines

*Corresponding Author: robertjay.angco@ctu.edu.ph

Received: 13 December 2025 Revised: 21 March 2026 Accepted: 4 April 2026

Paper Type: Original Research

Abstract

Aim/Purpose: Developing faculty research capabilities is critical for higher education institutions seeking to enhance research productivity and academic quality. Despite widespread recognition of the importance of structured research training, empirical evidence regarding effectiveness remains limited, particularly in developing-country contexts. The rapid evolution of research methodologies, publication landscapes, and technological tools, including artificial intelligence, necessitates systematic capacity building beyond traditional mentoring. Contemporary training must address fundamental methodological competencies, as well as emerging skills in digital tools, open-access publishing, predatory-journal identification, and ethical standards. Few studies have employed comprehensive evaluation frameworks to assess whether training programs achieve intended outcomes across learning and behavioral dimensions, limiting evidence-based decision-making for faculty development investments.

Introduction/Background: This study evaluated the effectiveness of a structured research capability training program using the Kirkpatrick Four-Level Evaluation Model, examining participant reactions (Level 1), knowledge acquisition (Level 2), behavioral change intentions and self-efficacy improvements (Level 3), and potential organizational impact (Level 4 indicator). The evaluation provided empirical evidence regarding the value of multi-topic intensive training while identifying program strengths and areas for improvement. By applying systematic evaluation, this study adds to the limited evidence on effective faculty development and provides actionable insights for institutional planning and resource allocation.

Methodology: A pre-post evaluation study with 40 faculty participants assessed an intensive multi-day research capability training program covering methodology fundamentals, journal selection strategies, predatory journal identification, AI tool applications, data analysis, and academic writing through interactive methods and expert-led sessions. Data collection employed: (a) validated pre-test and post-test questionnaires (10 items each) covering methodology, ethics, journal selection, and open access; (b) structured satisfaction questionnaire with four domains on 5-point Likert scales including content delivery (8 items), facilitator performance (4 items), logistics (4 items), and learning outcomes (4 items); (c) self-efficacy scales for journal identification and ethical standards confidence; and (d) open-ended qualitative feedback. Quantitative analysis used descriptive statistics, Cohen's *d* effect sizes, and frequencies. Qualitative responses were analyzed thematically, with inter-rater reliability (κ) = .84. Sample size for G Power analysis required a minimum of 34 participants for detecting Cohen's *d* of .50 effects at 80 percent power, with alpha equal to .05. All participants provided informed consent per institutional ethical requirements.

Findings: The evaluation demonstrated exceptional effectiveness across outcomes, providing strong evidence at Kirkpatrick Levels 1–2 with promising indicators at Level 3–4. Knowledge gains were substantial: open access understanding increased from 12.5 to 87.5 percent (a 75 percentage-point gain, Cohen's *d* = 3.2); predatory journal identification improved from 85 to 92.5 percent (Cohen *d* =

.60); ethics understanding rose from 92.5 to 97.5 percent (Cohen $d = .40$); methodology mastery reached 100 percent from 95 percent.

Self-efficacy showed large improvements: journal identification confidence increased from a mean of 3.60 ($SD = .97$) to a mean of 4.37 ($SD = .54$), a .77-point gain (Cohen $d = 1.1$); ethical standards confidence rose from mean of 3.92 ($SD = .96$) to a mean of 4.50 ($SD = .56$), (Cohen $d = .70$). Reduced standard deviations indicated convergence across competency levels; satisfaction evaluation ($n = 30$, 75 percent response) showed unanimous positive reception, with 100 percent rating the overall effectiveness with scores from 4.00 to 5.00 on a 5-point scale. Domain means ranged between 4.47 and 4.77, with instructor expertise as the highest ($M = 4.77$, $SD = .43$). Consistent high performance across content delivery ($M = 4.63$ -4.73), facilitator performance ($M = 4.70$ -4.77), logistics ($M = 4.47$ -4.73), and learning outcomes ($M = 4.60$ -4.70). Qualitative analysis identified three value-drivers: technology integration (18 mentions), publication strategy guidance (18 mentions), and expert instruction (12 mentions). Improvements needed included extended duration (8 mentions), methodological specialization (6 mentions), and infrastructure enhancement (5 mentions).

Contribution/Impact on Society: This research provides robust evidence for the effectiveness of structured multi-topic research capability training in enhancing faculty competencies for contemporary academic research. Findings validated comprehensive training integrating traditional skills with emerging technological tools and publication practices. There were demonstrated improvements in knowledge and confidence across competency areas, particularly predatory journal identification and open access publishing, supporting institutional faculty development investments. Training also led to successful technology integration, especially AI applications, models, and adaptation to evolving scholarly landscapes. The comprehensive evaluation framework offers a replicable methodology for institutions assessing programs, contributing to evidence-based academic capacity building. For institutions in developing contexts, findings revealed that well-designed intensive training achieved substantial improvements in faculty capabilities under resource constraints, supporting research productivity and academic quality goals.

Recommendations: Institutions may implement structured research capability programs that incorporate AI tool applications, journal selection guidance, and predatory journal identification training. Program design should prioritize contemporary content, expert instructors, interactive methods, and adequate duration for comprehensive coverage. Invest in appropriate technological infrastructure and facilities to optimize learning. Implement post-training support, including mentoring networks and follow-up sessions to extend impact and facilitate sustained change. Conduct systematic evaluation using the Kirkpatrick model to assess effectiveness, identify improvements, and demonstrate return on investment.

Research Limitation: Study limitations included an immediate post-training assessment without long-term follow-up for sustained behavioral change and organizational impact (Kirkpatrick Levels 3-4). Self-reported measures may have introduced response bias. The sample size of 40 participants from a single institution limits generalizability. The evaluation lacked objective behavioral observations or actual research output measurements. The cross-sectional design precluded temporal stability assessment of learning gains and confidence improvements.

Future Research: Longitudinal studies could examine sustained behavior change, actual research output improvements, and the organizational impact of research training interventions to validate the Kirkpatrick Levels 3 and 4 scores.

Keywords: *Research training, Kirkpatrick model, program evaluation*

Introduction

Research capability development among academic faculty represents a critical component of institutional advancement and knowledge creation. Universities across diverse regions—including

Southeast Asia, Sub-Saharan Africa, Latin America, and developed nations in Europe and North America—increasingly recognize the need for systematic approaches to enhance faculty research skills, particularly in an era of rapidly evolving research methodologies, publication landscapes, and technological tools (Huenneke et al., 2017; Cahapay, 2021). Despite this recognition, empirical evidence regarding the effectiveness of research capability training programs remains limited. For instance, Huenneke et al. (2017) documented that despite widespread institutional support, rigorous evaluation of training outcomes remains rare. Similarly, Perez et al. (2022), in their study of faculty in Philippine higher education, found that while research training programs were being conducted, systematic evidence for their effectiveness was largely absent, particularly in developing country contexts where resource constraints affect program design and delivery.

The challenge of building research capacity extends beyond individual skill development to encompass broader institutional objectives, including enhancing research output, improving grant acquisition success, and facilitating international collaboration (Huenneke et al., 2017). Traditional approaches to faculty development often rely on informal mentoring or discipline-specific guidance, which may not address the comprehensive skill sets required for contemporary research excellence (Knowles et al., 2015). Structured training programs offer a systematic alternative, but their effectiveness requires rigorous evaluation to justify investment and guide improvement (Alsalamah & Callinan, 2022).

Kirkpatrick's four-level evaluation model provides a comprehensive framework for assessing training effectiveness, examining participant reactions (Level 1), learning acquisition (Level 2), behavioral applications (Level 3), and organizational results (Level 4) (Kirkpatrick & Kirkpatrick, 2016). This model was selected over alternative frameworks (e.g., Phillips' ROI Model, Stufflebeam's CIPP Model) for several reasons: its widespread validation in educational and training contexts (Alsalamah & Callinan, 2022), its practical applicability to single-session training events (Cahapay, 2021), and its capacity to generate both immediate and forward-looking evidence for program improvement (Smidt et al., 2009). This model's hierarchical structure allows for systematic assessment of training impact while identifying areas for program enhancement (Alsalamah & Callinan, 2021; Smidt et al., 2009). The present study applied this framework to evaluate a research capability training program, contributing to the limited literature on evidence-based faculty development in research skills.

Literature Review

Faculty development in research capabilities has gained increasing attention as universities strive to enhance their research profiles and competitive positioning (Huenneke et al., 2017). Previous studies have identified key components of effective research training, including instruction in methodology, publication guidance, grant writing skills, and technology integration (Cahapay, 2021; Alsalamah & Callinan, 2022). However, few studies have employed systematic evaluation frameworks to assess program effectiveness comprehensively (Smidt et al., 2009; Alsalamah & Callinan, 2022; Rajeev et al., 2009). Research by Huenneke et al. (2017) demonstrated that structured research training programs could significantly improve faculty confidence and research output when properly designed and implemented. Their longitudinal study of 120 faculty members across multiple institutions found that participants in structured programs exhibited 45% greater research productivity over two years than control groups. Similarly, Alsalamah and Callinan (2021) found that multi-modal training approaches incorporating both theoretical instruction and practical application yielded superior outcomes compared to traditional lecture-based formats, with effect sizes ranging from .70 to 1.2 across different competency areas.

The application of Kirkpatrick's evaluation model to academic training contexts has shown promising results in various settings. For example, this framework was successfully used to evaluate a research methodology workshop series, finding a strong correlation ($r = .68$) between participants' reactions and subsequent research productivity measured six months post-training (Alsalamah & Callinan, 2022). This work highlighted the importance of measuring multiple evaluation levels to

capture the full spectrum of training impact, demonstrating that Level 1 satisfaction scores predicted 46% of variance in actual behavioral change.

Recent studies have emphasized the importance of contemporary content in research training programs. Sitzmann and Weinhardt (2018) found that training programs incorporating artificial intelligence tools and digital research methods showed significantly higher participant engagement and knowledge retention compared to traditional programs. Their meta-analysis of 23 studies revealed that technology-enhanced research training produced effect sizes 30% larger than conventional approaches. Consistent with this, Cahapay (2021) found that incorporating technology-aided activities into faculty training significantly improved both engagement and perceived usefulness of training content. Similarly, Reio et al. (2017) demonstrated that digital skill components in research training programs contributed to higher motivation and self-directed learning behaviors among faculty participants. Perez et al. (2022) further documented that Philippine faculty who received technology-integrated research training reported significantly higher research output in the following year compared to peers who attended conventional workshops.

Despite these advances, significant gaps remain in understanding optimal training design, duration, and content selection for the development of research capabilities (Reio et al., 2017). Additionally, Perez et al. (2022) found that 63% of faculty in Philippine state universities had never attended a formal research training program, and of those who had, only 28% reported that the training addressed contemporary publication skills—suggesting that institutional context and resource availability critically shape access to and content of training. Similarly, Cahapay (2021) noted that faculty in developing country institutions often face infrastructure constraints (e.g., limited Internet access, outdated facilities) that complicate the delivery and uptake of technology-integrated research training. These contextual factors represent significant challenges in diverse institutional contexts, particularly regarding resource availability and participant backgrounds (Huenneke et al., 2017). Most existing studies have been conducted in well-resourced institutions in developed countries, limiting generalizability to other contexts (Alsalamah & Callinan, 2022).

Kirkpatrick's Four-Level Evaluation Model: Theoretical Foundation

Originally introduced by Donald Kirkpatrick in 1959 and subsequently formalized in his seminal text (Kirkpatrick & Kirkpatrick, 2016), the Four-Level Evaluation Model remains one of the most widely cited training evaluation frameworks in both organizational and educational contexts. The model organizes training outcomes into four hierarchical levels: Level 1 (Reaction), which captures participant satisfaction and perceived value; Level 2 (Learning), which measures knowledge acquisition, skill development, and attitude change; Level 3 (Behavior), which assesses the transfer of learning to the workplace through behavioral change; and Level 4 (Results), which examines broader organizational impacts such as improved productivity, research output, or institutional performance (Alsalamah & Callinan, 2022). The hierarchical nature of the model implies that positive outcomes at the lower levels are prerequisite conditions—though not guarantees—for outcomes at higher levels (Smidt et al., 2009).

The model has been applied across diverse higher education and professional training contexts. In healthcare education, Smidt et al. (2009) demonstrated its utility in evaluating professional development workshops for practitioners with intellectual disabilities, finding that Level 1 and Level 2 outcomes were strong predictors of subsequent Level 3 behavioral improvements. In research training specifically, Rajeev et al. (2009) applied the model to academic training courses and found that multi-level evaluation captured nuances in learning outcomes that single-level evaluation would have missed. More recently, Alsalamah and Callinan (2022) conducted a bibliometric analysis of 60 years of Kirkpatrick model applications (1959–2020), confirming the model's enduring relevance and noting that Level 1 and Level 2 assessments remain the most frequently operationalized, while Level 3 and Level 4 evaluations continue to be underdeveloped in the literature. Cahapay (2021) also critically examined the model's use in higher education, noting that while it provides a practical framework, its

limitations include insufficient attention to contextual factors that mediate the transfer of learning to behavior. These insights directly inform the current study's design and its interpretation of findings.

Methodology

This section describes the research design, participant selection, training program, data collection instrument, and analytical strategies employed to evaluate the effectiveness of the research capability training program. The methodological approach was guided by Kirkpatrick's (2016) four-level evaluation framework and aligned with established practices in training program evaluation research (Alsalamah & Callinan, 2022; Smidt et al., 2009).

Design

This study employed a pre-experimental, one-group pretest-posttest evaluation design. This design was selected because it is particularly appropriate for intact training cohorts in which random assignment is not feasible, and it allows direct comparison of participants' knowledge and confidence levels before and after the training intervention (Cohen, 1988). While it does not eliminate all threats to internal validity, the pre-post design is widely accepted in training evaluation research as an efficient and practical method for assessing knowledge gains and self-efficacy changes within a single group (Reio et al., 2017). The evaluation was guided by Kirkpatrick's four-level evaluation model (Kirkpatrick & Kirkpatrick, 2016) and was conducted at Cebu Technological University—Main Campus (CTU-MC), Cebu City, Philippines, immediately following completion of a multi-day research capability training program for college faculty.

Participants

The study included 40 faculty members from Cebu Technological University—Main Campus (CTU-MC), Cebu City, Philippines, who participated in the research capability training program. The total faculty population at the institution is comprised of approximately 280 full-time academic staff across multiple colleges. Participants were drawn from this population through total enumeration of those who voluntarily enrolled in the training, which was offered as an institutional professional development initiative. Participants represented diverse academic disciplines and career stages, ensuring broad representativeness of the target population. The sample size was determined using G*Power analysis for detecting medium effect sizes (Cohen's $d = .50$) with 80% power at $\alpha = .05$, requiring a minimum of 34 participants (Cohen, 1988).

Training Program Description

The research capability training program was designed as a multi-day intensive workshop covering essential research skills and contemporary methodologies, following best practices identified in the literature (Cahapay, 2021). Key program components included:

- Research methodology fundamentals
- Journal selection and publication strategies
- Identification and avoidance of predatory journals
- Integration of artificial intelligence tools in research
- Data analysis and interpretation techniques
- Academic writing and presentation skills

The program employed interactive teaching methods, practical exercises, and expert-led sessions to maximize learning effectiveness and participant engagement, consistent with adult learning principles (Knowles et al., 2015).

Evaluation Instrument

Data collection utilized a structured questionnaire designed to measure training effectiveness across four key domains aligned with Kirkpatrick's evaluation levels (Kirkpatrick & Kirkpatrick, 2016):

1. Training Content and Delivery (8 items): Assessed objective clarity, content relevance, knowledge enhancement, and material usefulness.
2. Instructor Performance (4 items): Evaluated time allocation, expertise demonstration, concept explanation clarity, and participant engagement.
3. Logistics and Facilities (4 items): Measured interactive methods usage, registration efficiency, venue suitability, and technology functionality.
4. Learning Outcomes (4 items): Assessed organizational smoothness, confidence improvement, practical application potential, and expectation fulfillment.

All quantitative items utilized a 5-point Likert scale (1=Strongly Disagree to 5=Strongly Agree), following established measurement practices (DeVellis, 2017). The instrument demonstrated good internal consistency with Cronbach's alpha coefficients reported separately per domain: Training Content and Delivery ($\alpha = .88$), Instructor Performance ($\alpha = .91$), Logistics and Facilities ($\alpha = .85$), and Learning Outcomes ($\alpha = .82$), all exceeding the acceptable threshold of .70 (DeVellis, 2017). Additional qualitative feedback was collected through open-ended questions addressing helpful aspects, improvement areas, and future topic suggestions. Knowledge assessment was conducted using validated instruments adapted from prior research on training evaluation (Alsalamah & Callinan, 2021; Smidt et al., 2009) that covered five key domains, each with two items. Self-efficacy was measured using scales adapted from Bandura's (2006) self-efficacy assessment guidelines.

Data Analysis

Quantitative data were analyzed using descriptive statistics, including means, standard deviations, and frequency distributions, and a paired-samples *t*-test. Effect sizes were calculated using Cohen's *d* to determine the practical significance of changes (Cohen, 1988). Response patterns were examined to identify areas of strength and opportunities for improvement. Qualitative responses were analyzed using inductive thematic analysis, following Braun and Clarke's (2006) six-phase framework, to identify common themes and specific recommendations for program enhancement. Inter-rater reliability for thematic coding was established at $\kappa = .84$, indicating substantial agreement.

Ethical Considerations

All participants provided informed consent for evaluation data collection and analysis prior to their involvement in the study. Participation was voluntary; faculty who chose not to participate in the evaluation could still attend the training without any adverse consequences. Data were collected anonymously, with participant responses identified only by assigned codes. The study was conducted in accordance with the institutional review board's ethical guidelines, and all data were handled with strict confidentiality. Participants were informed of their right to withdraw at any time without penalty.

Results and Discussion

Profile of the Participants

All 40 training participants completed the evaluation questionnaire, yielding a 100% response rate. This comprehensive participation level strengthened the findings' reliability and representativeness.

Research Questions and Hypotheses

The following research questions and hypotheses guided the evaluation. Hypotheses for Research Questions (RQ) 1 and 2 were tested using paired-samples *t*-tests; results are integrated in the combined pre-test/post-test summary (Table 4).

RQ1: Does the training program significantly improve participant knowledge across all domains?

H_{10} : There is no statistically significant difference between pre-test and post-test knowledge scores ($\alpha = .05$).

RQ2: Does the training program significantly improve participant self-efficacy in journal identification and ethical standards?

H_{20} : There is no statistically significant improvement in self-efficacy scores from pre-test to post-test ($\alpha = .05$).

RQ3: How do participants rate training quality and effectiveness (Kirkpatrick Level 1)?

Pretest Results

The pre-test knowledge assessment results are shown below in Table 1.

Table 1 Pre-test Knowledge Assessment Results

Knowledge Domain	Correct Responses (%)	Mean	SD
Predatory Journal Identification	34 (85.00)	4.25	.72
Open Access Publication Types	5 (12.50)	.63	.88
Research Ethics/Plagiarism	37 (92.50)	4.63	.54
Journal Selection Criteria	40 (100.00)	5.00	.00
Research Methodology	35 (87.50)	4.38	.81
Total Knowledge Score	151/400 (37.80)	3.78	1.45

Pre-test results revealed a heterogeneous knowledge landscape among faculty participants, with significant variations across different research capability domains. The most striking finding was the extremely low baseline knowledge in open-access publication concepts, with only 12.5% of participants demonstrating understanding. This critical knowledge gap reflects a broader challenge in higher education institutions where faculty may lack exposure to contemporary publishing models despite their increasing importance in academic dissemination (Sitzmann & Weinhardt, 2018). This finding was consistent with Perez et al. (2022) and Smidt et al.'s (2009) observation that faculty in developing country contexts often have limited exposure to evolving publication practices.

The strong performance in journal selection criteria (100% correct) and research ethics (92.5% correct) suggested that traditional research training has effectively addressed these foundational concepts, aligning with findings from Alsalamah and Callinan (2022), who noted that ethical training has been well-integrated into academic preparation programs. However, the perfect score in journal selection may indicate ceiling effects in the assessment instrument or may reflect participants' theoretical knowledge without practical application skills.

Using the interpretation scale adopted in this study (1.00–1.79 = Very Low; 1.80–2.59 = Low; 2.60–3.39 = Moderate; 3.40–4.19 = High; 4.20–5.00 = Very High), the pre-test self-efficacy scores shown in Table 2 indicated a high level of confidence overall. The moderate-to-high self-efficacy levels, particularly the lower confidence in journal identification ($M = 3.60$ vs. $M = 3.92$ for ethical standards), aligned with the knowledge assessment results and suggested that participants recognized their limitations in contemporary publication practices.

Table 2 Pre-test Self-Efficacy Results

Self-Efficacy Domain	Mean	SD
Journal Identification Confidence	3.60	.97
Ethical Standards Confidence	3.92	.96

This self-awareness was consistent with Bandura's (2006) social cognitive theory, which posits that self-efficacy judgments are influenced by performance accomplishments. The relatively high standard deviations (.97 and .96) indicated substantial individual differences in confidence levels, providing a significant opportunity for improvement through targeted training interventions.

These baseline results validated the training program's focus areas and demonstrated clear learning needs that justified the comprehensive intervention approach employed, supporting Perez et al.'s (2022) argument for needs-based training design.

Post-test Results

The post-test knowledge assessment results are shown below in Table 3.

Table 3 *Post-test Knowledge Assessment Results*

Knowledge Domain	Correct Responses (%)	Mean	SD	Improvement
Predatory Journal Identification	37 (92.50)	4.63	.49	+7.5%
Open Access Publication Types	35 (87.50)	4.38	1.06	+75.0%
Research Ethics/Plagiarism	39 (97.50)	4.88	.33	+5.0%
Journal Selection Criteria	40 (100.00)	5.00	.00	.0%
Research Methodology	40 (100.00)	5.00	.00	+12.5%
Total Knowledge Score	191/400 (47.8%)	4.78	1.12	+10.0%

Post-test results demonstrated remarkable learning achievement across multiple knowledge domains, providing compelling evidence for the training program's effectiveness in knowledge transfer. The most extraordinary improvement occurred in open access publication understanding, with a 75 percentage-point increase. As shown in Table 4 (Effect Size column), this improvement represented one of the largest effect sizes (Cohen's $d = 3.2$) documented in faculty development literature (Cahapay, 2021). Table 4 also presents t -test results confirming statistical significance for all domains where pre-post differences were observed. This massive improvement validated the training's contemporary relevance and participants' capacity for acquiring complex new knowledge when properly supported, consistent with adult learning theory principles (Knowles et al., 2015).

The achievement of complete mastery in research methodology (100% correct post-training) indicated that the program successfully addressed identified knowledge gaps while building upon existing foundations. This finding was aligned with Alsalamah and Callinan's (2021) observation that well-structured training programs can achieve near-complete competency transfer in specific skill areas. The improvement in predatory journal identification, while modest in percentage terms (+7.5%), represented a meaningful enhancement from an already strong baseline (Cohen's $d = .60$), suggesting that the training refined and solidified existing knowledge rather than introducing entirely new concepts.

The reduction in standard deviation scores across most domains (notably from 1.45 to 1.12 in total knowledge scores) indicated that the training not only improved mean performance, but it also reduced variability among participants. This convergence effect suggests that the program was particularly effective in supporting lower-performing participants while maintaining engagement for those with stronger baselines, a finding consistent with Alsalamah and Callinan's (2021) research on inclusive training design. The persistence of ceiling effects in journal selection criteria highlights both the strength of participants' existing knowledge and potential limitations in the sensitivity of assessment instruments for this domain, suggesting a need for more discriminating evaluation instruments in future studies (DeVellis, 2017).

Knowledge Assessment Results by Domain

Table 4 consolidates the pre-test and post-test knowledge assessment data (previously in Tables 1 and 3) into a single comparative summary using mean scores. It also includes paired-samples t -test results and Cohen's d effect sizes for each domain, thereby testing H_1 . Knowledge scores are based on a 5-point scale (1 item = 1 point); Pre-test and Post-test (%) reflect item-level accuracy scores, while "Mean" reflects the scale score. A domain with 100% correct responses corresponds to a mean of

5.00; domain means below 5.00 indicate variability within items, or that some participants scored partial credit.

Table 4 *Pre-test vs. Post-test Knowledge Assessment Results*

Knowledge Domain	Pre-test (%)*	Post-test (%)*	Improvement	t-value	p-value	Effect Size
Predatory Journal Identification	34 (85.00)	37 (92.50)	+7.5%	3.795	< .001**	Small
Open Access Publication Types	5 (12.50)	35 (87.50)	+75.0%	20.239	< .001**	Large
Research Ethics/Plagiarism	37 (92.50)	39 (97.50)	+5.0%	2.530	.016**	Small
Journal Selection Criteria	40 (100.00)	40 (100.00)	.0%	-	-	-
Research Methodology	35 (87.50)	40 (100.00)	+12.5%	3.162	.003**	Moderate

* % Correct; ** Significant at $\alpha = .05$

Table 5 shows the post-test self-efficacy data and the mean improvement in comparison with the pretest results (Table 2), including the paired-samples *t*-test results and Cohen's *d* effect sizes (testing H_2), and applies the following interpretation scale for mean scores: 1.00–1.79 = Very Low; 1.80–2.59 = Low; 2.60–3.39 = Moderate; 3.40–4.19 = High; 4.20–5.00 = Very High

Table 5 *Post-test Self-Efficacy Results*

Self-Efficacy Domain	Mean	SD	Improvement	t-value	p-value	Effect Size
Journal Identification Confidence	4.37	.54	+.77	6.957	< .001*	Large
Ethical Standards Confidence	4.50	.56	+.58	4.427	< .001*	Medium

* Significant at $\alpha = .05$

Self-efficacy improvements represent perhaps the most significant outcomes of the training program, as confidence measures are strong predictors of behavioral application and sustained knowledge use (Bandura, 2006). The .77-point improvement in journal identification confidence demonstrated that participants not only acquired knowledge, but they also developed the self-assurance necessary for practical implementation. This confidence gain is particularly important given the complexity of contemporary publication landscapes and the high stakes associated with journal selection decisions, as highlighted by Huenneke et al. (2017).

The large effect size for journal identification confidence improvement (Cohen's $d = 1.1$, as shown in Table 5) exceeded benchmarks for educational interventions and suggests practically significant change (Cohen, 1988). The substantial reduction in standard deviations (from .97 to .54 for journal identification) indicated that the training successfully addressed the wide range of initial confidence levels, creating greater consensus and shared competence among participants. This convergence effect suggests the program was effective in supporting participants across different baseline confidence levels, consistent with inclusive training design principles (Smidt et al., 2009).

The medium effect size (Cohen's $d = .70$) for ethical standards confidence improvement, while smaller than journal identification gains, represented a meaningful enhancement in an area where participants had already demonstrated relatively high baseline confidence. This suggests that the training reinforced and deepened existing ethical understanding rather than introducing entirely new concepts, aligning with research showing that confidence building in established competency areas requires different approaches than new skill development (Knowles et al., 2015).

These self-efficacy improvements provided strong predictive indicators for Level 3 (Behavior) outcomes in Kirkpatrick's model, suggesting that participants are likely to apply their newly acquired knowledge and skills in their research activities (Alsalamah & Callinan, 2021; Smidt et al., 2009).

Training Satisfaction Evaluation Results

The training satisfaction by domain results are shown below in Table 6.

Table 6 *Training Satisfaction by Domain (n=30)*

Evaluation Domain	Mean	SD	% Rating 4.00-5.00
TRAINING CONTENT AND DELIVERY			
Objectives clearly stated	4.73	.52	96.7
Topics relevant to researcher needs	4.70	.53	93.3
Content enhanced knowledge/skills	4.67	.55	90.0
Materials/resources were useful	4.63	.56	86.7
INSTRUCTOR PERFORMANCE			
Demonstrated expertise	4.77	.50	96.7
Explained concepts clearly	4.73	.52	93.3
Encouraged participation	4.70	.53	90.0
LOGISTICS AND FACILITIES			
Interactive methods used	4.67	.55	86.7
Registration/communication efficient	4.73	.52	93.3
Venue/facilities conducive	4.53	.68	80.0
Technology functioned well	4.47	.73	76.7
LEARNING OUTCOMES			
Organization was smooth	4.70	.53	90.0
Increased confidence in research	4.63	.56	86.7
Can apply learning to work	4.67	.55	86.7
Met expectations	4.60	.62	83.3
OVERALL TRAINING EFFECTIVENESS	4.73	.45	100.0

The satisfaction evaluation results provide overwhelming evidence for exceptional Level 1 (Reaction) performance according to Kirkpatrick's framework (Kirkpatrick & Kirkpatrick, 2016). The unanimous positive rating for overall training effectiveness (100% rated from 4.00 to 5.00) represents a rare achievement in educational program evaluation and indicates strong perceived value among all participants. This finding exceeded typical satisfaction benchmarks reported in faculty development literature, where 85% to 90% positive ratings are considered excellent (Huenneke et al., 2017).

Instructor performance emerged as the highest-rated domain, with expertise demonstration receiving the peak score (4.77). This finding validated the critical importance of subject matter expertise in effective faculty development programming, consistent with research by Smidt et al. (2009) and Alsalamah and Callinan (2022), who found that instructor credibility was the strongest predictor of training success. The consistent high ratings across all instructor-related measures (expertise, clarity, participation encouragement) suggested that instructional quality was a key driver of program success, supporting investment in expert facilitators.

The strong performance in content and delivery measures confirmed that participants perceived the training as relevant, useful, and well-designed. The high ratings for objective clarity (4.73) and topic relevance (4.70) indicated successful alignment between program design and participant needs,

reflecting effective needs assessment and curriculum development processes (Alsalamah & Callinan, 2022). The progressive decrease in ratings from objectives to materials (4.73 to 4.63) suggested a potential area for resource enhancement while maintaining overall program strength.

Technology functionality showed the most variability ($SD = .73$) and the lowest mean score (4.47), though still indicating strong satisfaction. This pattern suggested that while technological issues did not significantly detract from overall experiences, infrastructure improvements could enhance future program delivery. Higher variability in technology ratings may reflect different levels of technical comfort among participants or inconsistent technology performance during the program, consistent with challenges identified in technology-enhanced training programs (Sitzmann & Weinhardt, 2018).

The learning outcomes domain received consistently positive ratings, with participants expressing confidence in their ability to apply new knowledge (4.67) and reporting increased research confidence (4.63). These self-reported behavioral intentions provide additional support for potential Level 3 outcomes and align with research showing strong correlations between satisfaction and application intentions (Reio et al., 2017).

Qualitative Feedback: Thematic Analysis and Integration with Quantitative Findings

Qualitative responses from 40 participants were analyzed using inductive thematic analysis following Braun and Clarke's (2006) six-phase framework, which provided a rich contextual understanding of participant experiences. Three major themes emerged for positive aspects, with improvements needed in three areas. These qualitative themes are presented alongside the corresponding quantitative satisfaction findings to provide a convergent, integrated picture of Level 1 outcomes, as recommended for strengthening the reliability of mixed-methods evaluation (Creswell & Plano Clark, 2018).

Theme 1: Technology Integration and Contemporary Relevance. This theme captured participants' strong appreciation for the integration of current technologies and contemporary research tools, consistent with Sitzmann and Weinhardt's (2018) findings on technology-enhanced learning. Representative quotes included:

- "The use of apps; access to the different journal sites" (P1)
- "AI tools" (P3)
- "Using artificial intelligence tools" (P14)
- "More AIs to aid publication" (P25)

The emphasis on artificial intelligence applications reflects growing recognition of AI's role in contemporary research practices and aligns with participant needs in this rapidly evolving area.

Theme 2: Publication Strategies and Journal Navigation. Participants consistently praised practical guidance in publication processes, especially journal selection strategies, which aligned with the quantitative finding of substantial improvement in journal identification. Representative quotes:

- "Choosing the right journal, and using AI in publication" (P4)
- "Spotting predatory journals" (P12)
- "The speaker's discussion on how to choose the right journal" (P16)
- "Selection of reputable journals" (P11)

This theme validated the program's focus on contemporary publication challenges and supported the significant learning gains observed in predatory journal identification.

Theme 3: Expert Instruction and Mentoring Quality. The quality of instruction and personalized guidance was greatly appreciated, supporting the quantitative findings for instructor performance:

- "The speakers' presentations" (P26)
- "The research chairman was superb in helping me with the IMRAD" (P10)
- "The coaching session" (P19)
- "The mentoring" (P30)

This theme reinforced the critical importance of expert instructors for effective faculty development, consistent with previous such findings (Smidt et al., 2009; Alsalamah & Callinan, 2022).

The next set of themes concerns the participants' suggestions for improvement.

Theme 4: Extended Duration Needs. The most frequently cited improvement area was the need for extended training time:

- "More days for the training since research is extensive, diverse, tedious" (P20)
- "Time allotment" (P30)
- "More time for research writing" (P6)

This feedback suggests participants valued the content sufficiently to desire more extensive coverage, indicating high program engagement.

Theme 5: Methodological Specialization Requests. Participants expressed a desire for more specialized methodological training:

- "Qualitative methods" (P3)
- "Separate discussion for qualitative and quantitative research" (P16)
- "Statistics" (P6)

These requests indicate opportunities for advanced or specialized training tracks, consistent with literature on differentiated professional development (Smidt et al., 2009).

Theme 6: Infrastructure Enhancement. Participants noted facility and follow-up support needs:

- "Venue (conference room)" (P18)
- "Follow-up on submission of Scopus-indexed" (P23)
- "More engagements from the participants" (P24)

This feedback aligned with the quantitative finding that technology and facilities showed the most room for improvement.

Findings and Discussion

Thematic analysis revealed three primary value drivers that aligned closely with quantitative satisfaction results. Technology integration emerged as a key strength, reflecting participants' recognition of the training's contemporary relevance. This finding supported the substantial quantitative improvement in open access knowledge, as participants appreciated practical guidance on modern publication approaches, consistent with Sitzmann and Weinhardt's (2018) research on the effectiveness of technology-enhanced training.

Publication strategy guidance received the highest frequency of positive mentions (18), validating the training's focus on journal selection and identifying predatory publications. This aligned with quantitative improvements in predatory journal identification and high satisfaction ratings for content relevance, supporting the program's strategic focus on contemporary publication challenges.

Expert instruction quality was consistently praised, supporting the highest quantitative satisfaction rating (4.77) for instructor expertise. The personalized attention and coaching aspects suggested that individual guidance contributed significantly to the positive learning experience, consistent with adult learning principles emphasizing personalized support (Knowles et al., 2015).

Improvement suggestions centered on three main areas: extended duration (most frequent request), methodological specialization, and infrastructure enhancement. The duration requests were aligned with the comprehensive scope of topics covered, and suggested that participants valued the content sufficiently to desire more extensive coverage. Methodological specialization requests indicated opportunities for advanced or specialized training tracks, while infrastructure feedback supported the finding that technology and facilities showed the most room for improvement.

Training Effectiveness Assessment Using Kirkpatrick's Framework

Level 1 (Reaction) – Exceptional Performance. The evaluation results demonstrated outstanding performance at Kirkpatrick's Level 1, with consistently high satisfaction ratings across all measured dimensions. Mean scores consistently above 4.50 across all evaluation domains, combined with unanimous participant endorsement of overall training effectiveness, indicated strong positive reactions and high perceived value. This exceptional performance aligned with the upper tier of training evaluation benchmarks and provided a strong foundation for higher-level outcomes (Kirkpatrick & Kirkpatrick, 2016).

Level 2 (Learning) – Significant Knowledge and Confidence Gains. The pre-test/post-test analysis provided robust evidence for Level 2 effectiveness, demonstrating substantial learning acquisition across multiple knowledge domains. The most notable achievement was the 75 percent improvement in open access knowledge (Cohen's $d = 3.2$), indicating successful knowledge transfer in an area where participants had limited prior understanding. This effect size significantly exceeded typical educational intervention benchmarks and represented one of the largest documented improvements in faculty development literature (Alsalamah & Callinan, 2021; Smidt et al., 2009).

The self-efficacy improvements were particularly significant, as confidence measures often predict behavioral application (Bandura, 2006). The .77-point improvement in journal identification confidence (Cohen's $d = 1.1$) suggested that participants not only acquired knowledge, but they also developed the self-assurance necessary for practical application. This finding is consistent with social cognitive theory and provided strong predictive validity for Level 3 outcomes.

Level 3 and 4 Implications. While this evaluation did not directly measure behavioral change (Level 3) or organizational results (Level 4) due to timing constraints, the strong foundation established for Levels 1 and 2, combined with substantial improvements in confidence, suggests positive prospects for knowledge application and organizational impact. The high ratings for application intentions (4.67) and increased research confidence (4.63) provided promising indicators for future behavioral measurement, consistent with research linking satisfaction and self-efficacy to behavioral outcomes (Reio et al., 2017).

Program Strengths, Areas for Enhancement, and Limitations

The evaluation results revealed several notable program strengths, areas for improvement, and limitations that collectively inform future program development. In terms of strengths, the integration of current technologies, particularly artificial intelligence tools, demonstrated strong alignment with participant needs and contemporary research trends, supporting the substantial quantitative improvement in open access knowledge and reflecting successful adaptation to evolving research practices (Sitzmann & Weinhardt, 2018). The emphasis on immediately applicable skills and real-world scenarios resonated strongly with participants, enhancing both learning effectiveness and satisfaction consistent with adult learning principles (Knowles et al., 2015). The quality of instruction received consistent praise, with the combination of expertise and effective communication skills optimizing learning outcomes as evidenced by the highest satisfaction rating for instructor performance ($M = 4.77$), validating the importance of subject matter expertise in faculty development programming (Smidt et al., 2009; Alsalamah & Callinan, 2022). The multi-topic approach provided participants with a holistic skill development experience that addresses diverse learning needs within a single program framework.

Notwithstanding these strengths, participant feedback identified several areas for enhancement, including requests for more detailed coverage of qualitative methodologies and discipline-specific approaches, suggesting a need for specialized tracks or follow-up sessions, investment in appropriate venues and technological infrastructure to address the lowest-rated logistics domain, and implementation of post-training mentoring or support networks to enhance long-term impact and support Level 3 and 4 outcome achievement. These findings must also be interpreted in light of several limitations: the evaluation was conducted immediately following training completion, limiting

assessment of long-term behavioral change and organizational impact consistent with best practices in training evaluation (Kirkpatrick & Kirkpatrick, 2016); the sample size of 40 from a single institution constrained generalizability. Future studies should consider multi-site evaluations to enhance external validity (Cohen, 1988); reliance on self-report measures may have introduced bias, and future evaluations would benefit from objective learning measures and behavioral observation to triangulate findings (Reio et al., 2017). The absence of Level 3 and 4 assessment limited understanding of ultimate program impact, underscoring the need for long-term follow-up studies to strengthen the evidence base for program effectiveness (Smidt et al., 2009).

Conclusion

This evaluation study provides strong evidence for the effectiveness of a research capability training program designed to enhance faculty research skills and knowledge. Using Kirkpatrick's evaluation framework, the findings demonstrated exceptional performance at the reaction and learning levels, with positive indicators for potential behavioral application and organizational impact. The unanimous participant endorsement of training effectiveness, combined with consistently high ratings across all evaluation dimensions, supports the value of structured, comprehensive approaches to developing faculty research skills. Identification of specific program strengths provides a framework for replication and scaling, while the improvement suggestions offer clear directions for program enhancement. The study contributes to the limited evidence base regarding effective faculty development of research capabilities while demonstrating the practical application of systematic evaluation approaches to training assessment. The findings support institutional investment in faculty development initiatives while providing guidance for optimal program design and implementation.

Recommendations

Based on the findings, institutions considering similar faculty development initiatives are encouraged to incorporate contemporary technologies and research trends into program design, as the success of AI integration in this program demonstrated that cutting-edge content enhances both learning and satisfaction outcomes. Subject matter expertise and instructional clarity may be prioritized when selecting facilitators, given that instructor performance received the highest satisfaction ratings in this evaluation. Program duration shall be sufficient to allow comprehensive coverage of complex topics, as participant requests for additional time reflect high engagement and suggest that extended formats may further improve learning outcomes.

Beyond the training itself, institutions may implement post-training support systems, such as mentoring networks and follow-up sessions to sustain knowledge application and extend program impact. For future research, longitudinal studies are recommended to measure behavioral change (Level 3) and organizational impact (Level 4) outcomes, while multi-site replications would strengthen the generalizability of findings. Finally, future evaluations might incorporate objective learning assessments and behavioral indicators to complement self-report measures and strengthen evaluation validity (Tamkin et al., 2002).

Author Contributions

Robert Jay N. Angco: Conceptualization, methodology, investigation, data curation, formal analysis, visualization, writing—original draft, writing—review & editing, project administration.

References

- Alsalamah, A., & Callinan, C. (2021). Adaptation of Kirkpatrick's four-level model of training criteria to evaluate training programmes for head teachers. *Education Sciences*, 11(3), 116. <https://doi.org/10.3390/educsci11030116>
- Alsalamah, A., & Callinan, C. (2022). The Kirkpatrick model for training evaluation: Bibliometric analysis after 60 years (1959–2020). *Industrial and Commercial Training*, 54(2), 36–63. <https://doi.org/10.1108/ICT-12-2020-0115>

- Bandura, A. (2006). Adolescent development from an agentic perspective. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents*. Information Age Publishing. <https://doi.org/10.1108/978-1-60752-750-320251002>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Cahapay, M. (2021). Kirkpatrick model: Its limitations as used in higher education evaluation. *International Journal of Assessment Tools in Education*, 8(1), 135–144. <https://doi.org/10.21449/ijate.856143>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). SAGE Publications.
- Huenneke, L. F., Stearns, D. M., Martinez, J. D., & Laurila, K. (2017). Key strategies for building research capacity of university faculty members. *Innovative Higher Education*, 42(5–6), 421–435. <https://doi.org/10.1007/s10755-017-9394-y>
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2016). *Evaluating training programs: The four levels* (4th ed.). Berrett-Koehler Publishers.
- Knowles, M.S., Holton III, E.F., & Swanson, R.A. (2015). *The adult learner: The definitive classic in adult education and human resource development* (8th ed.). Routledge. <https://doi.org/10.4324/9781315816951>
- Perez, Z. O., Minyamin, A. V., Bagsit, R. D., Gimena, G. B., Dionaldo, W. V., Padillo, E. S., Lovoie, O. G., & Cabello, C. A. (2022). Research capability of faculty members in higher education institution: Basis for research management plan. *Journal of Positive School Psychology*, 6(3), 6215–6226. <https://journalppw.com/index.php/jpsp/article/view/3527/2306>
- Rajeev, P., Madan, M. S., & Jayarajan, K. (2009). Revisiting Kirkpatrick's model—an evaluation of an academic training course. *Current Science*, 96(2), 272–276. <https://www.jstor.org/stable/24105191>
- Reio Jr, T. G., Rocco, T. S., Smith, D. H., & Chang, E. (2017). A critique of Kirkpatrick's evaluation model. *New Horizons in Adult Education and Human Resource Development*, 29(2), 35–53. <https://doi.org/10.1002/nha3.20178>
- Sitzmann, T., & Weinhardt, J. M. (2018). Training engagement theory: A multilevel perspective on the effectiveness of work-related training. *Journal of Management*, 44(2), 732–756. <https://doi.org/10.1177/0149206315574596>
- Smidt, A., Balandin, S., Sigafoos, J., & Reed, V. A. (2009). The Kirkpatrick model: A useful tool for evaluating training outcomes. *Journal of Intellectual and Developmental Disability*, 34(3), 266–274. <https://doi.org/10.1080/13668250903093125>
- Tamkin, P., Yarnall, J., & Kerrin, M. (2002). *Kirkpatrick and beyond: A review of models of training evaluation*. Institute for Employment Studies Report 392. <https://www.employment-studies.co.uk/system/files/resources/files/392.pdf>