

**MULTI-LEVEL POST-TRAINING EVALUATION FRAMEWORK FOR
MANAGEMENT DEVELOPMENT PROGRAMS AT PT GARUDA
MAINTENANCE FACILITY AERO ASIA TBK**

Muhammad Rafly Zinedine Zhilaal^{1*}
Universitas Islam Indonesia, Yogyakarta, Indonesia
22311488@students.uii.ac.id *(Corresponding)

Majang Palupi²
Universitas Islam Indonesia, Yogyakarta, Indonesia
majang_palupi@uii.ac.id

Abstract

Leadership development programs in safety-critical industries require robust evaluation mechanisms to demonstrate organizational value, yet most organizations confine assessment to Level 1 satisfaction surveys. This study investigates post-training evaluation practices for management development programs at PT Garuda Maintenance Facility (GMF) Aero Asia Tbk, Indonesia's premier aviation Maintenance, Repair, and Overhaul (MRO) provider, and develops a comprehensive evaluation framework grounded in the Kirkpatrick-Phillips model. Employing a qualitative descriptive-analytical approach within a design and development research paradigm, the study analyzed 638 module-level evaluation responses and 584 qualitative participant comments across three programs: the Operational Leadership Program (OLP), MRO Management, and MRO Finance. Thematic analysis, content analysis of training syllabi, and descriptive quantitative analysis were integrated through methodological triangulation. Findings reveal that GMF's evaluation is uniformly confined to Level 1 reaction measurement, producing consistently high satisfaction scores while Levels 2 through 5 remain entirely absent, a pattern termed the evaluation ceiling paradox. The study proposes an integrated ADDIE-Kirkpatrick-Phillips framework encompassing enhanced reaction surveys, pre-test and post-test instruments, 360-degree behavioral assessments, results tracking templates, and ROI calculation guidelines, differentiated by program type and contextualized for the MRO industry and Indonesian BUMN governance setting. The framework is designed for phased 24-month implementation leveraging existing technology infrastructure.

Keywords: Post-Training Evaluation, Kirkpatrick-Phillips Model, Leadership Development, Aviation MRO, Training Effectiveness



INTRODUCTION

Human resources constitute one of the most critical assets in any organization. Despite rapid technological advancement, people remain the cornerstone of organizational success, making continuous development of employee knowledge and skills an enduring strategic imperative (Millmore et al., 2007). This imperative is particularly pronounced in technical and safety-critical industries such as aviation maintenance, repair, and overhaul (MRO), where competent personnel directly influence operational safety, service quality, and business competitiveness. Mello (2014) emphasizes that strategic human resource management (including systematic training and development programs) serves as a key driver of organizational performance and competitive advantage.

Training investment patterns reveal significant disparities across nations and industries. According to Indonesia's Minister of Manpower, a World Bank survey found that fewer than 10% of Indonesian companies provide formal training to their employees, contrasting sharply with China (80%), the Philippines (60%), and Vietnam (20%) (Aziz, 2019). This low investment correlates with Indonesia's persistently lower labor productivity relative to regional peers. Consistent with this, Prahara et al. (2025) demonstrated that education and skills development represent the most significant determinants of labor productivity among Indonesian workers, underscoring that structured human capital investment, including systematic training programs, is a critical lever for organizational performance improvement. In the aviation MRO sector, workforce development challenges are intensifying globally due to aging workforces, surging post-pandemic demand, and rapid technological change including digitalization and automation (Aircraft-parts, 2025).

PT Garuda Maintenance Facility (GMF) Aero Asia Tbk, as Indonesia's premier MRO provider and a significant player in the Asia-Pacific region, has recognized training as a strategic priority. GMF's Leader & Talent Development (L&TD) unit administers three principal management development programs: the Operational Leadership Program (OLP) targeting senior managers and high-potential talent; the MRO Management program providing cross-functional business knowledge; and the MRO Finance program equipping department heads with financial literacy for decision-making. Despite substantial investment in these programs, current evaluation practices are limited exclusively to Level 1 satisfaction surveys, leaving fundamental questions about learning acquisition, behavioral transfer, and organizational impact entirely unanswered.

The limitation of satisfaction-based evaluation is well-documented in training literature. Kirkpatrick (2016), whose four-level model has served as the field's gold standard for over six decades, emphasizes that Level 1 (Reaction) provides no evidence of actual learning, behavior change, or organizational impact. Phillips and Phillips (2016) subsequently added Level 5 (Return on Investment), enabling financial justification of development investments. Research by Twitchell et al. (2000) found that while 86-100% of organizations conduct Level 1 evaluation, only 21-49% implement Level 4, reflecting the reality that higher evaluation levels demand greater complexity, time, and resources. Ahmad et al. (2024) demonstrated that comprehensive multi-level evaluation produces statistically significant relationships between

reaction, learning, behavior, and results, enabling targeted program improvements and enhanced organizational performance.

This study addresses three research questions: (1) What are GMF's current training evaluation practices, and what gaps exist relative to the Kirkpatrick-Phillips framework? (2) How can a comprehensive post-training evaluation framework be designed and adapted for GMF's MRO management development programs? (3) What implementation guidelines and practical instruments are needed to operationalize the framework within GMF's organizational constraints? The study aims to develop a theoretically grounded, contextually calibrated evaluation framework that bridges identified gaps and enables evidence-based management of leadership development investments in the MRO context.

REVIEW OF LITERATURE

The Kirkpatrick-Phillips model constitutes the primary theoretical foundation for this study. Kirkpatrick's (2016) four-level framework, comprising Level 1 (Reaction), Level 2 (Learning), Level 3 (Behavior), and Level 4 (Results), has dominated organizational training evaluation for over six decades. Level 1 assesses participant satisfaction and perceived relevance. Level 2 measures knowledge, skill, and attitudinal acquisition through pre-tests, post-tests, and skill demonstrations. Level 3 evaluates behavioral transfer to the workplace, typically assessed 3-6 months post-training through multi-source feedback or supervisor observations. Level 4 examines training's impact on organizational performance metrics such as productivity, quality, safety, or financial outcomes. Phillips and Phillips (2016) added Level 5 (ROI), converting Level 4 results into monetary terms and comparing benefits against program costs using the formula: $ROI\% = [(Benefits - Costs) / Costs] \times 100$.

Despite the model's widespread adoption, empirical evidence reveals that higher-level evaluation remains rare. Twitchell et al. (2000) documented that 86-100% of organizations assess reactions, 71-90% measure learning, but only 43-83% evaluate behavior, and merely 21-49% implement results-level assessment. Fortenberry (2025) found that in healthcare leadership programs, 95% conducted Level 1 assessment but only 13% tracked Level 4 outcomes. Ahmad et al. (2024) structural equation modeling confirmed significant positive relationships across all four Kirkpatrick levels, with reaction influencing learning ($\beta = 0.43$, $p < 0.001$), learning influencing behavior ($\beta = 0.51$, $p < 0.001$), and behavior influencing results ($\beta = 0.38$, $p < 0.001$).

Transfer of learning theory constitutes a critical complementary framework. Baldwin and Ford (1988) identified three categories of factors influencing whether training translates into workplace performance: training design, trainee characteristics, and work environment. Environmental factors (supervisor support, peer reinforcement, and organizational climate) prove most influential. Hirv-Biene et al. (2025) documented that environmental support explained 41% of variance in transfer success, exceeding both trainee characteristics (18%) and training design (23%). De Jong et al. (2025) found that participants who set specific application goals post-training demonstrated 47% higher transfer rates at three-month follow-up. Blume et al. (2019) meta-analysis documented a persistent learning-to-behavior gap: training producing significant learning gains ($d = 1.2$) frequently failed to produce corresponding behavioral changes ($d = 0.3$), primarily attributable to environmental constraints rather than learning deficits.

Andragogical theory (Knowles et al., 2015) provides a third foundational framework, establishing that adult learners differ fundamentally from children in requiring problem-centered, experientially grounded, and autonomy-respecting learning environments. These principles directly inform evaluation design, emphasizing that assessment instruments must treat participants as collaborative partners in their own development rather than passive subjects of testing. Evaluation should emphasize application to authentic work challenges and frame findings developmentally to support continuous professional growth.

Critical perspectives on Kirkpatrick acknowledge that the model's hierarchical structure oversimplifies relationships between evaluation levels (Bates, 2004), and that it provides no account of organizational conditions necessary for transfer. Alternative frameworks such as CIRO (Warr et al., 1970) and CIPP (Stufflebeam, 1971) address some limitations but lack Kirkpatrick's practitioner familiarity and direct correspondence to evaluation gaps. The Kirkpatrick-Phillips model was retained for this study because: (1) it is the most widely validated and practitioner-accessible framework; (2) its sequential levels map directly onto the specific gaps documented in GMF's current practice; and (3) the Phillips ROI extension addresses GMF's specific accountability needs as a BUMN subsidiary. The integration of transfer theory and andragogy partially addresses Kirkpatrick's theoretical limitations.

Research on leadership development evaluation in safety-critical industries reveals a pronounced gap. Güneş et al. (2020) identified significant disparities between training content and actual competency requirements in aviation maintenance organizations. Aircraft-parts (2025) documented that the global MRO sector faces a critical skills gap as 60% of the maintenance workforce approaches retirement. Despite these challenges, existing evaluation scholarship originates predominantly from Western corporate, healthcare, or military settings, leaving aviation MRO environments underexplored. This study contributes to the literature by contextualizing evaluation theory within the dual constraints of MRO industry requirements and Indonesian BUMN governance norms.

RESEARCH METHOD

This study employed a qualitative descriptive-analytical approach within a design and development research paradigm. Design and development research is defined as systematic study of design, development, and evaluation processes with the aim of establishing an empirical basis for the creation of instructional and non-instructional products and tools (Richey & Klein, 2007). This paradigm was selected because the primary contribution of the study is a theoretically grounded evaluation framework and associated instruments, rather than hypothesis testing or causal inference.

The unit of analysis was GMF's L&TD unit and its three management development programs: the Operational Leadership Program (OLP), MRO Management, and MRO Finance. Primary data were obtained through three sources: (1) training evaluation documents including 638 module-level quantitative responses and 584 qualitative participant comments collected via Google Forms during training delivery; (2) training syllabi and materials for all three programs; and (3) participatory observation during a four-month internship period (April 2025–July 2025) embedded within the L&TD unit.

Three complementary analytical techniques were integrated through methodological triangulation. Thematic analysis following Braun and Clarke's (2006) six-stage procedure served as the primary technique for qualitative data, including participant comments and document analysis. Descriptive quantitative analysis (means, standard deviations, modes, inter-program comparisons) provided a supporting quantitative picture of Level 1 evaluation patterns. Content analysis of training syllabi and materials enabled learning objective inference for each program module, calibrated to Bloom's Taxonomy (2001), forming the basis for Level 2 instrument design.

Validity was ensured through data source triangulation across three programs and multiple data types, iterative practitioner review throughout the internship period, expert validation by academic supervisors, and thick description of context enabling transferability assessment. Reliability was maintained through systematic documentation of analytical procedures, theoretical grounding in validated frameworks, and peer debriefing. Ethical compliance required explicit approval from the L&TD unit supervisor for use of internal documents; all quantitative data are presented in aggregated form and qualitative comments reported thematically without individual attribution.

Researcher reflexivity was explicitly acknowledged, as the researcher occupied an insider position within the unit under study. Potential biases (including insider bias, confirmation bias, and social desirability effects) were mitigated through cross-program triangulation, active search for disconfirming evidence, peer debriefing, and theoretical anchoring rather than reliance on organizational perspectives alone.

RESULTS AND DISCUSSION

Current Training Evaluation Practices at GMF

GMF's evaluation system is uniformly confined to Kirkpatrick Level 1 (Reaction) assessment across all management development programs. Evaluation instruments are administered digitally via Google Forms immediately upon program completion, capturing four dimensions: instructor knowledge, instructor presentation ability, instructor punctuality, and instructor teaching methods; material completeness, job relevance, content currency, and adequacy of practice opportunities; and time sufficiency and session effectiveness. While these instruments generate systematic satisfaction data, no pre-training baseline assessment, post-training knowledge testing, behavioral follow-up, performance tracking, or ROI calculation exists.

Analysis of 638 module-level evaluation responses revealed consistently high satisfaction across all three programs. The OLP produced 143 module-level responses with an instructor aspect average of 4.91/5.00 (SD = 0.22–0.35) and a material aspect average of 4.80/5.00 (SD = 0.43–0.55). The MRO Management program generated 220 responses with instructor and material averages of 4.86/5.00 and 4.78/5.00 respectively. The MRO Finance program produced 275 responses with lower but still highly positive averages of 4.63/5.00 for instructor quality and 4.54/5.00 for material relevance, with higher variability (SD = 0.51–0.66) potentially reflecting the challenge of communicating technical financial concepts to participants with diverse educational backgrounds. Across all programs, the practice

opportunities dimension consistently received the lowest average rating and highest standard deviation, suggesting systematic participant desire for more hands-on application activities.

These quantitative patterns exhibit ceiling effects: modal responses cluster at scale maximum (5.00) with restricted ranges, raising questions about whether the instrument detects meaningful variation in training quality or merely reflects social desirability responding. More importantly, consistently high satisfaction scores create what this study terms an evaluation ceiling paradox, a self-reinforcing organizational mechanism whereby psychometrically limited data generate false confidence, suppressing motivation for deeper evaluation. High ratings signal that training is working well, reducing perceived urgency for more rigorous assessment, yet satisfaction scores neither verify learning acquisition nor predict behavioral transfer or organizational impact.

Table 1.
Cross-Program Quantitative Evaluation Summary

Program	N Responses	Instructor Avg	Material Avg	Time Avg
OLP	143	4.91/5.00	4.80/5.00	2.72/3.00
MRO Management	220	4.86/5.00	4.78/5.00	2.90/3.00
MRO Finance	275	4.63/5.00	4.54/5.00	2.72/3.00

Thematic Analysis of Qualitative Feedback

Thematic analysis of 584 qualitative comments across the three programs identified four recurring themes with direct implications for evaluation framework development. First, the dominant theme of general satisfaction and appreciation (approximately 38% of codable comments) confirms the quantitative pattern but reveals its limitation: participants express appreciation for training delivery without addressing its developmental impact. Second, requests for enhanced learning assessment (2.2% of explicit comments, with implicit resonance across multiple themes) demonstrate participant awareness that satisfaction differs from learning and that objective measurement would provide value. OLP participants explicitly requested 'pre-assessment and post-assessment for every material to better map participant knowledge before and after,' representing unsolicited demand for Level 2 evaluation from the very learners the framework would serve.

Third, the theme of content complexity and pace variation (4.0% in MRO Finance, distributed across programs) indicates that training content is calibrated to assumed rather than empirically established baseline knowledge, underscoring the need for pre-testing to enable differentiated instruction. Fourth, requests for continued engagement and post-training follow-up (including Indonesian-language comments requesting 'updates on company developments after training') signal participant desire for ongoing connection between L&TD and training alumni, a relationship that Level 3 behavior evaluation would formalize. Together, these qualitative patterns suggest that the evaluation gaps identified



through gap analysis are not merely academic limitations but align with participant-expressed developmental needs.

Gap Analysis: Current Practice Versus Kirkpatrick-Phillips Framework

The gap analysis applied the Kirkpatrick-Phillips five-level framework as an evaluative benchmark, revealing systematic deficiencies at Levels 2 through 5. At Level 2 (Learning), GMF conducts no pre-test baseline assessment and no post-training knowledge verification. Consequently, the organization cannot determine whether participants acquired intended knowledge, identify individuals requiring additional support, or demonstrate to stakeholders that training investment produced measurable competency development. At Level 3 (Behavior), evaluation concludes at program completion with no follow-up assessment of workplace application. This constitutes the most critical gap: training's ultimate purpose is changing managerial behavior to improve performance, yet GMF has no visibility into whether this objective is achieved. Research consistently documents that the majority of training fails to transfer to sustained workplace application (Ardondi et al., 2025), but GMF cannot determine whether its programs experience similar transfer challenges.

At Level 4 (Results), no system connects training participation to organizational performance metrics. GMF cannot demonstrate that leadership development contributes to MRO performance indicators such as quality, productivity, safety, or customer satisfaction, which is particularly problematic given the organization's 2024 strategic priorities emphasizing profitability enhancement and operational excellence (GMF Integrated Report, 2024). At Level 5 (ROI), the absence of learning, behavior, and results data renders cost-benefit analysis impossible. Without financial justification, training risks perception as an administrative expense rather than strategic investment, creating vulnerability during budget deliberations. Table 2 summarizes the documented evaluation gaps and their organizational consequences.

Table 2.
Evaluation Gap Analysis: Current Practice vs. Kirkpatrick-Phillips Standards

KP Level	Current Status	Key Gap	Organizational Consequence
Level 1 – Reaction	Implemented (GMF)	Narrow construct coverage; ceiling effects	False confidence; limited diagnostic value
Level 2 – Learning	Absent	No pre/post knowledge testing	Cannot verify learning acquisition
Level 3 – Behavior	Absent	No transfer follow-up	Training impact on performance invisible
Level 4 – Results	Absent	No KPI linkage	Strategic value undemonstrated



Level 5 – ROI	Absent	No cost-benefit analysis	Financial justification impossible
---------------	--------	--------------------------	------------------------------------

The Proposed Comprehensive Evaluation Framework

The proposed framework integrates the ADDIE instructional design model with the Kirkpatrick-Phillips evaluation hierarchy, transfer of learning theory, and andragogical principles within GMF's specific organizational context. Before presenting the evaluation framework in detail, it is necessary to position it within the broader training lifecycle proposed for GMF. Table 3 presents the overall ADDIE-based training framework that organizes GMF's existing training methods into an integrated development cycle, with comprehensive evaluation as its culminating and feedback-generating phase.

Table 3.

Overall Training Framework

Phase	Methods
Pre-Training (Analyze, Design, Develop)	> Training Needs Analysis (TNA) > 360-degree Assessment
Training (Implement)	> Classroom (Collaborative & Gamification) > E-learning > OJT
Post-Training (Evaluate)	> Kirkpatrick Evaluation > Phillips ROI

*Note. The ADDIE model is proposed as the integrative framework that GMF currently lacks. The Post-Training (Evaluate) phase represents the proposed contribution of this research. *Feedback Loop: Evaluation informs the next TNA cycle.*

The framework is organized around five evaluation levels, differentiated by program type: Levels 1-2 for the MRO Finance program (foundational financial literacy), Levels 1-3 for the MRO Management program (competency-oriented cross-functional knowledge), and Levels 1-5 for the flagship OLP (strategic leadership development with highest investment level). This differentiated application reflects Phillips and Phillips' (2016) recommendation for ROI selectivity, extended in this study by tying evaluation depth to program type, developmental objective, and organizational feasibility. Table 4 presents the complete framework structure.

Table 4.
Proposed Comprehensive Post-Training Evaluation Framework

Level	Primary Focus	Instrument	Timing	Gap Addressed	Key Question Answered
Level 1 Enhanced Reaction	Satisfaction & transfer readiness (intention, confidence, utility, support)	Enhanced digital survey (Google Forms)	Immediately post-training	Problem 1: Limited scope of current evaluation	Did participants find training relevant, and are they ready to apply it?
Level 2 Learning	Knowledge & competency gains (pre-post differential)	Pre-test and post-test aligned with learning objectives; Case-based application task	Pre-test: Day 1 before; Post-test: Final session	Problem 2: Absence of learning measurement	Did participants actually learn the intended knowledge and skills?
Level 3 Behavior	Leadership behavior change; Workplace application of competencies	360-degree feedback (self, supervisor, peers, direct reports); Competency-based assessment	3-6 months post-training	Problem 3: No behavior change tracking	Did participants apply what they learned in their actual work?
Level 4 Results	Organizational performance impact; departmental metrics; strategic outcomes	Departmental KPI tracking; Trained vs. untrained comparison; Stakeholder perception survey	6-12 months post-training	Problem 4: Lack of business results linkage	Did training contribute to measurable organizational performance improvement?
Level 5 ROI	Financial return on	Phillips ROI methodology	12+ months	Problem 5: Inability to	Did the training

	training investment; monetary cost-benefit analysis	(benefit isolation, monetary conversion, full cost analysis)	post-training; Applied selectively to key programs	calculate ROI	investment generate adequate financial returns?
--	---	--	--	---------------	---

Note. Each level addresses a specific evaluation gap documented in GMF's current practice. Differentiated application by program type reflects evaluation feasibility and investment level.

At Level 1, the enhanced reaction instrument retains GMF's current satisfaction dimensions while adding five transfer-readiness constructs: behavioral intention (participants' commitment to implement new practices), learning confidence (self-efficacy regarding content mastery), perceived utility (anticipated applicability to specific work contexts), environmental support perception (expected supervisor and peer reinforcement), and a dedicated open-response section employing structured qualitative prompts rather than spontaneous comment collection. These additions, requiring approximately two to three additional minutes of participant time, provide richer transfer readiness data while maintaining instrument feasibility.

Level 2 instruments introduce program-specific pre-test and post-test assessments aligned with learning objectives inferred through content analysis of training syllabi and calibrated to Bloom's Taxonomy cognitive levels. OLP assessments emphasize strategic thinking, operational leadership, and innovation competencies. MRO Management assessments evaluate cross-functional business knowledge acquisition across nine operational domains. MRO Finance assessments verify financial literacy development, with particular attention to the baseline variability documented in thematic analysis. Learning data enable identification of participants requiring additional support, curriculum refinement decisions grounded in empirical evidence, and stakeholder demonstration of competency development.

Level 3 instruments employ 360-degree behavioral feedback administered 90 days post-training, utilizing eight MRO-relevant leadership competencies derived from GMF's AKHLAK corporate values and industry-specific leadership requirements: strategic orientation, operational excellence, stakeholder communication, decision-making under pressure, continuous improvement facilitation, safety consciousness, collaborative problem-solving, and talent development. Multi-source ratings from self, immediate supervisor, peers, and direct reports enable pre-post comparison where Level 3 assessment uses the same competency framework as GMF's existing pre-training 360-degree assessment, enabling rigorous measurement of training-attributed behavioral change.

Level 4 tracking links OLP participation to organizational performance metrics through three convergent evidence strategies: comparative analysis of performance metrics in departments led by OLP graduates versus comparable untrained managers; longitudinal tracking of departmental KPIs (including safety incident rates, quality audit scores, employee engagement, and productivity indices) pre and post-training participation; and targeted application of an isolation methodology adapted from Phillips' ROI process model to

Multi-Level Post-Training Evaluation Framework...

estimate training's net contribution. Level 5 ROI calculation for OLP employs the Phillips formula: $ROI\% = [(Benefits - Costs) / Costs] \times 100$, applied selectively to quantifiable Level 4 outcomes, with conservative attribution assumptions and transparent reporting for executive communication.

Implementation is structured as a phased 24-month timeline organized into four stages as presented in Table 5. This phasing reflects three theoretical principles: training evaluation literature recommends incremental capability building before advancing to more complex levels (Phillips & Phillips, 2016); Kirkpatrick and Kirkpatrick (2016) suggest allowing 3-6 months post-training before Level 3 assessment and 6-12 months before Level 4; and organizational change research indicates that major process changes require 18-24 months for institutionalization (Kotter, 1996).

Table 5.

Phased Implementation of the Proposed Post-Training Evaluation Framework

Phase	Timeline	Priority Actions	Expected Outcomes	Rationale
Phase 1: Foundation	Months 1-6	Enhance Level 1 with transfer-readiness items; Pilot Level 2 for MRO Finance	Transfer readiness data; First learning evidence	Begin with lowest complexity levels; pilot approach manages risk (Phillips & Phillips, 2016)
Phase 2: Expansion	Months 7-12	Scale Level 2 to all programs; Pilot Level 3 for OLP with existing 360-degree data	Cross-program learning data; First behavioral evidence	6-month gap enables Level 3 behavioral assessment (Kirkpatrick & Kirkpatrick, 2016)
Phase 3: Integration	Months 13-18	Scale Level 3 to MRO Management; Implement Level 4 for OLP; Establish HR partnerships	Business impact evidence; Multilevel data	12-month period allows business results to materialize (Phillips & Phillips, 2016)

Phase 4: Optimization	Months 19-24	Conduct Level 5 ROI for OLP; Refine all instruments; Document SOPs	ROI evidence; Institutionalized system	Sufficient data for credible ROI; 18-24 months for system institutionalization (Kotter, 1996)
--------------------------	-----------------	--	--	---

Note. Each phase builds on the preceding one, with evaluation data from earlier phases informing instrument design in later phases. GMF should adapt phasing and duration based on actual implementation experience, resource availability, and strategic priorities.

Integrative ADDIE-Kirkpatrick-Phillips Conceptual Model

The comprehensive framework situates the Kirkpatrick-Phillips evaluation hierarchy as the evaluative component of GMF's complete ADDIE training development cycle. Figure 1 presents the integrated conceptual model connecting organizational strategic goals, the ADDIE process, the five Kirkpatrick-Phillips evaluation levels, and the continuous improvement feedback loop.

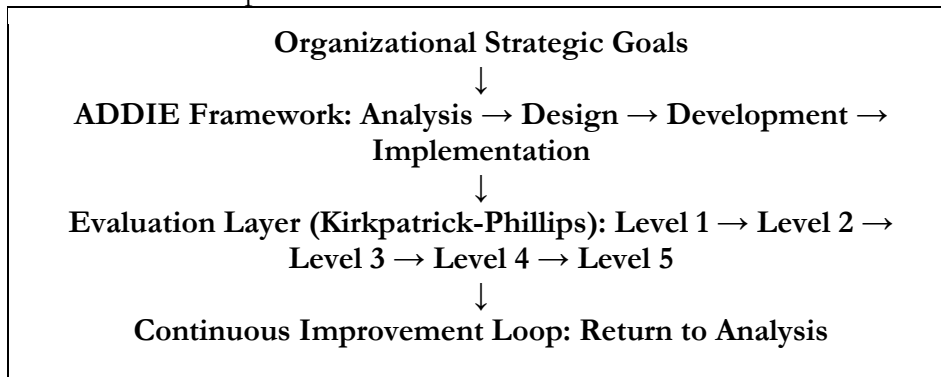


Figure 1.

Integrated ADDIE and Proposed Kirkpatrick-Phillips Evaluation Framework

In the Analyze stage, Training Needs Analysis identifies competency gaps aligned with organizational strategy while establishing baseline data for ROI calculation and defining expected Level 4 outcomes. The Design stage formulates measurable learning objectives aligned with Level 2 assessment requirements and Level 3 behavioral competencies. The Development stage creates enhanced Level 1 instruments, pre/post-test content, 360-degree tools, and KPI tracking frameworks concurrently with curriculum materials. The Implement stage delivers training while conducting Level 1 and Level 2 assessments and integrating transfer action planning. Post-implementation Evaluation measures Level 3 behavior change, monitors Level 4 performance impacts, calculates Level 5 ROI for flagship programs, and feeds findings back to needs analysis for continuous improvement. This integration transforms training from discrete events into a continuous organizational learning system, with evaluation generating knowledge that drives evidence-based program refinement and accountability at all levels.

Theoretical Contributions

This study makes three theoretical contributions to human resource development literature. First, it proposes the evaluation ceiling paradox as an analytical lens describing a self-reinforcing organizational mechanism whereby consistently high satisfaction scores may suppress institutional motivation for deeper evaluation, extending Bates' (2004) psychometric critique of Level 1 measures by hypothesizing a second-order organizational effect operating through institutional rationalization, cultural score compression, and availability heuristics. Second, it advances a principle of evaluation framework design as theory integration rather than single-model application, demonstrating that Kirkpatrick's practical strengths and theoretical gaps (particularly its absence of a transfer explanation) can be addressed by pairing it with Baldwin and Ford's (1988) transfer theory and Knowles et al.'s (2015) andragogical principles, producing richer predictive and explanatory power than any single framework alone. Third, the differentiated application of evaluation depth by program type, developmental objective, and organizational context (Levels 1-2 for foundational programs, 1-3 for competency programs, 1-5 for flagship investments) contributes a practically grounded typology addressing a gap in literature that tends to present Kirkpatrick-Phillips as a uniform prescription rather than an adaptable framework.

CONCLUSION

This study investigated post-training evaluation practices at PT GMF Aero Asia Tbk and developed a comprehensive evaluation framework addressing identified deficiencies. Through analysis of 638 quantitative evaluation responses, 584 qualitative participant comments, and training materials across three management development programs, the study documented that GMF's evaluation system is uniformly confined to Level 1 satisfaction measurement, creating an evaluation ceiling paradox in which psychometrically limited data produce organizational confidence while more consequential questions about learning, behavior, and impact remain entirely unexamined. Systematic absence of Levels 2 through 5 assessment constitutes strategic vulnerability: the organization cannot verify learning acquisition, measure behavioral transfer, demonstrate training's contribution to organizational performance, or justify development investments in financial terms.

The proposed integrated ADDIE-Kirkpatrick-Phillips framework addresses each gap through theoretically grounded, contextually calibrated instruments: enhanced reaction surveys capturing transfer-readiness constructs; pre/post knowledge assessments aligned to program-specific learning objectives; 360-degree behavioral feedback utilizing eight MRO-relevant leadership competencies; organizational performance tracking templates linked to GMF's strategic KPIs; and ROI calculation guidelines applicable selectively to flagship programs. Differentiated application by program type and phased 24-month implementation reflect the practical constraints of an organization with limited evaluation expertise, no dedicated learning management infrastructure, and competing resource demands.

The study's limitations include its single-organization focus, which restricts direct generalizability; reliance on spontaneously generated participant comments rather than systematically elicited qualitative data; the inference-based derivation of learning objectives

from syllabi analysis rather than formal design documentation; and the design-oriented nature of the contribution: the framework has not been empirically validated through actual implementation, and claims regarding effectiveness remain theoretical propositions requiring longitudinal validation. Future research should prioritize: (1) a longitudinal validation study tracking framework implementation outcome across multiple program cohorts; (2) cross-organizational comparative studies examining evaluation practices in other MRO providers and Indonesian BUMN subsidiaries; (3) empirical validation of the evaluation ceiling paradox concept across organizational contexts; and (4) psychometric validation of the proposed instruments prior to operational deployment.

For practitioners, this study demonstrates that comprehensive training evaluation need not be prohibitively resource-intensive: Phillips (2003) estimates that robust ROI evaluation requires only 3–5% of total training budget, and the proposed framework leverages existing technology infrastructure throughout. The evaluation ceiling paradox identified in this study serves as a diagnostic signal for organizations whose uniformly high satisfaction scores mask unexamined assumptions about training effectiveness. Organizations facing similar evaluation gaps are advised to prioritize transfer-focused assessment, adapt frameworks to organizational context and program type rather than adopting generic prescriptions, build incrementally through phased implementation, and communicate evaluation's strategic value as a capability enabling evidence-based decisions rather than as an administrative burden. As leadership development continues to be recognized as critical to organizational competitiveness in the aviation MRO sector and beyond, systematic evaluation transforms training from organizational expenditure into measurable strategic investment.

REFERENCES

- Ahmad, N., Hamid, S. N. A., Wahab, S. A., & Zulkarnain, Z. A. (2024). Training evaluation and organizational performance: A structural equation modeling approach in the banking sector. *International Journal of Human Resource Management*, 35(4), 678–704. <https://doi.org/10.1080/09585192.2023.2298461>
- Aircraft-parts. (2025). *MRO workforce challenges and talent development in the global aviation industry*. Aircraft-Parts International. <https://www.aircraft-parts.com/mro-workforce-challenges>
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Ardondi, P., Mazzucchelli, S., & De Bernardi, P. (2025). Transfer of training in organizational settings: A systematic scoping review of 49 studies. *Human Resource Development Review*, 24(1), 45–89. <https://doi.org/10.1177/15344843241289741>
- Aziz, F. A. (2019). The urgency of employee training and development in Indonesian organizations: A comparative perspective. *Journal of Indonesian Human Resource Management*, 3(2), 12–25.
- Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, 41(1), 63–105. <https://doi.org/10.1111/j.1744-6570.1988.tb00632.x>

- 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>
- Blume, B. D., Ford, J. K., Surface, E. A., & Olenick, J. (2019). A dynamic model of training transfer. *Human Resource Management Review*, 29(2), 270–283. <https://doi.org/10.1016/j.hrmr.2017.11.004>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- De Jong, R., Gorissen, P., & Kessels, J. (2025). Boosting transfer of training through post-training goal setting: A randomized controlled trial. *Human Resource Development Quarterly*, 36(1), 29–54. <https://doi.org/10.1002/hrdq.21526>
- Faisal-E-Alam, M., Rahman, M. M., & Hossain, S. (2025). Beyond reaction: Multi-level evaluation and its incremental explanatory power in predicting training outcomes. *Journal of Workplace Learning*, 37(2), 101–118. <https://doi.org/10.1108/JWL-06-2024-0087>
- Fortenberry, J. L. (2025). Post-training evaluation in healthcare leadership development programs: A national survey. *Journal of Healthcare Management*, 70(1), 14–29. <https://doi.org/10.1097/JHM-D-24-00052>
- GMF Aero Asia. (2024). *GMF integrated annual report 2024*. PT Garuda Maintenance Facility Aero Asia Tbk.
- Güneş, E., Ülker, M., & Çelik, K. (2020). Competency gaps in aviation maintenance training: Evidence from Turkish MRO organizations. *International Journal of Aviation, Aeronautics, and Aerospace*, 7(3), 1–21. <https://doi.org/10.15394/ijaaa.2020.1485>
- Hirv-Biene, J., Kaist, T., & Saks, A. M. (2025). Environmental predictors of training transfer: A meta-analytic update. *Human Resource Development International*, 28(1), 12–37. <https://doi.org/10.1080/13678868.2024.2381245>
- Isbahi, M. B., Zuana, M. M. M., & Toha, M. (2024). The Multi-Social Relation of the Cattle Industry in the Plaosan Subdistrict Animal Market of Magetan Regency. *Malacca: Journal of Management and Business Development*, 1(1), 31–46. <https://doi.org/10.69965/malacca.v1i1.51>
- Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). *Kirkpatrick's four levels of training evaluation*. ATD Press.
- Knowles, M. S., Holton, E. F., & Swanson, R. A. (2015). *The adult learner: The definitive classic in adult education and human resource development* (8th ed.). Routledge.
- Kotter, J. P. (1996). *Leading change*. Harvard Business School Press.
- Mello, J. A. (2014). *Strategic human resource management* (4th ed.). Cengage Learning.
- Millmore, M., Lewis, P., Saunders, M., Thornhill, A., & Morrow, T. (2007). *Strategic human resource management: Contemporary issues*. Pearson Education.
- Phillips, J. J. (2003). *Return on investment in training and performance improvement programs* (2nd ed.). Butterworth-Heinemann.
- Phillips, J. J., & Phillips, P. P. (2016). *Handbook of training evaluation and measurement methods* (4th ed.). Routledge.
- Prahara, R. S., Fauziyah, N. D., & A'yuni, D. S. (2025). Quantitative study on the effect of age, education, experience, and wages on labor productivity in housing projects.



- Danadyaksa: Post Modern Economy Journal*, 2(2), 167–185.
<https://doi.org/10.69965/danadyaksa.v2i2.109>
- Richey, R. C., & Klein, J. D. (2007). *Design and development research: Methods, strategies, and issues*. Lawrence Erlbaum Associates.
- Stufflebeam, D. L. (1971). The relevance of the CIPP evaluation model for educational accountability. *Journal of Research and Development in Education*, 5(1), 19–25.
- Twitchell, S., Holton, E. F., & Trott, J. W. (2000). Technical training evaluation practices in the United States. *Performance Improvement Quarterly*, 13(3), 84–109.
<https://doi.org/10.1111/j.1937-8327.2000.tb00178.x>
- Warr, P., Bird, M., & Rackham, N. (1970). *Evaluation of management training*. Gower Press.