A DESIGN AND DEVELOPMENT STUDY ON MULTI-LEVEL POST-TRAINING EVALUATION FRAMEWORK FOR MANAGEMENT DEVELOPMENT PROGRAMS AT PT GARUDA MAINTENANCE FACILITY AERO ASIA TBK

Internship Final Project

Written by:

Muhammad Rafly Zinedine Zhilaal - 22311488

INTERNATIONAL UNDERGRADUATE PROGRAM IN MANAGEMENT

FACULTY OF BUSINESS AND ECONOMICS

UNIVERSITAS ISLAM INDONESIA

2026

A DESIGN AND DEVELOPMENT STUDY ON MULTI-LEVEL POST-TRAINING EVALUATION FRAMEWORK FOR MANAGEMENT DEVELOPMENT PROGRAMS AT PT GARUDA MAINTENANCE FACILITY AERO ASIA TBK

Internship Final Project

Submitted in partial fulfilment of the requirements for the Final Project Examination leading to the award of a Bachelor's Degree in International Undergraduate Program in Management, Faculty of Business and Economics, Universitas Islam Indonesia.

Written by:

Name : Muhammad Rafly Zinedine Zhilaal

Student Number : 22311488

INTERNATIONAL UNDERGRADUATE PROGRAM IN MANAGEMENT

FACULTY OF BUSINESS AND ECONOMICS

UNIVERSITAS ISLAM INDONESIA

2026

ADVISOR APPROVAL PAGE

A Design and Development Study on Multi-Level Post-Training Evaluation Framework for Management Development Programs At PT Garuda Maintenance Facility Aero Asia Tbk

Submitted by

Name	: Muhammad Rafly Zinedine Zhilaal
Student Number	: 22311488
Study Program	: Management
Concentration	: Human Resources

Approved by

Content Advisor,

Dr. Majang Palupi BBA., MBA., CPHCM.

Language Advisor,

Alfi Zakiya, S.Kom., S.Pd.

DECLARATION OF AUTHENTICITY

I declare the originality of this thesis, which was made by the author themselves. I do not use anyone else's work, words, expressions, and ideas without acknowledgment. I declare that all of the data, ideas, words, and the writing was obtained by the author themselves.

If this statement is false in the future, I am willing to accept any sanction complying with the determined regulation or its consequences.

Yogyakarta, 6 March 2026

M Rafly Zinedine Z

ACKNOWLEDGEMENTS

All praise be to Allah SWT for His endless blessings, mercy, and guidance that have enabled me to complete this final research assignment. Shalawat and greetings to Prophet Muhammad SAW, who has guided humanity toward a better path and a brighter era.

This final research was prepared as a requirement to obtain a Bachelor's Degree in the International Undergraduate Program in Management, Faculty of Business and Economics, Universitas Islam Indonesia. The completion of this research and my internship program at PT Garuda Maintenance Facility (GMF) Aero Asia Tbk would not have been possible without the support, assistance, and contributions from many parties. Therefore, I would like to express my sincere gratitude to:

1. Johan Arifin, S.E., M.Si., Ph.D., as the Dean of the Faculty of Business and Economics, Islamic University of Indonesia.
2. Abdur Rafik, S.E., M.Sc., as the Head of the Undergraduate Program in Management.
3. Dr. Majang Palupi BBA., MBA., CPHCM., as my thesis advisor. I am deeply grateful for your guidance, patience, and insightful feedback throughout this research process. Your academic rigor challenged me to think more systematically and complete this thesis with greater clarity and confidence.
4. Katiya Nahda, S.E., M.Sc., as the Secretary of the International Undergraduate Program in Management. I sincerely appreciate your support and assistance during my academic journey. Your responsiveness and willingness to help have provided reassurance during important stages of my study.

5. Alfi Zakiya, S.Kom., S.Pd., as my language advisor. Thank you for having your time checking my English grammar on my final project report.
6. All lecturers and staff of the International Undergraduate Program in Management, Faculty of Business and Economics, Universitas Islam Indonesia, for their valuable knowledge and support. I especially thank the IUP staff, Mrs. Alfi, Mr. Pasha, and Mrs. Annisa, for their assistance and dedication throughout my academic journey.
7. My beloved parents and family. I am profoundly thankful for your endless prayers, sacrifices, patience, and unconditional support. Everything I have achieved today stands on the foundation of your hard work, sincerity, and belief in me. I carry your hopes in every step I take, and my gratitude for you goes beyond what words can express.
8. Yudithia Dzaryati, as the Senior Manager of Leader & Talent Development at PT GMF Aero Asia Tbk, for the trust and opportunity to gain professional learning experience during my internship.
9. Enggar Diswandoro, as my supervisor during the internship program, for the guidance and direction provided throughout my internship period. I also extend my gratitude to Adena Darin Anbarunisa and Rahma Dilla Arnanda from the Leader & Talent Development unit for the warmth and meaningful learning experiences.
10. Iqlimasilka Zahra Putri. Thank you for staying beside me through every high and low of this journey. In moments of doubt, exhaustion, and pressure, you never once made me feel alone. Your constant encouragement, thoughtful words, and genuine care gave me the strength to keep going when I felt uncertain. This thesis carries not only my effort, but also your prayers, encouragement, and sincere support.
11. Fardan, Akmal, Arroyan, Raffael, Elian, Dewa, Damar, Galih, and Egan for the friendship, shared discussions, and support that made this journey more meaningful.

12. Many other parties whom I cannot mention individually, yet whose support and prayers have been invaluable throughout this process.

May this research provide meaningful contributions to academic development and practical implementation in the field of management.

Yogyakarta, 6 March 2026

M Rafly Zinedine Z

ABSTRACT

This research addressed the limitations of post-training evaluation practices in management development programs at PT Garuda Maintenance Facility (GMF) Aero Asia Tbk, an aviation Maintenance, Repair, and Overhaul organization operating in a safety-critical environment. Despite substantial investment in leadership development, evaluation practices were limited to Level 1 reaction measures, focusing primarily on participant satisfaction. The research aimed to analyze existing evaluation practices, identify gaps based on the Kirkpatrick-Phillips model, and develop a comprehensive post-training evaluation framework tailored to the MRO context. The research employed a qualitative descriptive-analytical approach within a design and development research paradigm. Primary data were obtained from training evaluation documents, training syllabi and materials, and participatory observation during the internship period. Thematic analysis served as the primary analytical technique, complemented by content analysis and descriptive quantitative analysis as supporting data. Findings indicated a systematic absence of learning measurement, behavior tracking, results linkage, and return on investment calculation. The research proposed an integrated evaluation framework encompassing all five Kirkpatrick-Phillips levels, including enhanced reaction surveys, pre-test and post-tests, 360-degree behavior assessment, results tracking templates, and ROI calculation guidelines, contextualized within the aviation MRO industry and Indonesian BUMN governance setting. It is concluded that systematic multi-level evaluation was essential to ensure accountability, learning transfer, and measurable organizational impact in leadership development programs within safety-critical industries.

Keywords: post-training evaluation framework, Kirkpatrick-Phillips model, leadership development, aviation MRO, return on investment, training effectiveness

ABSTRAK

Penelitian ini membahas keterbatasan-keterbatasan praktik evaluasi pasca pelatihan pada program pengembangan manajer di PT Garuda Maintenance Facility (GMF) Aero Asia Tbk, perusahaan Maintenance, Repair, and Overhaul penerbangan yang beroperasi dalam lingkungan keselamatan tinggi. Meskipun organisasi telah berinvestasi signifikan dalam pengembangan kepemimpinan, praktik evaluasi masih terbatas pada Level 1 reaksi yang berfokus pada kepuasan peserta. Penelitian ini bertujuan menganalisis praktik evaluasi yang ada, mengidentifikasi kesenjangan berdasarkan model Kirkpatrick-Phillips, serta mengembangkan kerangka evaluasi pasca pelatihan yang komprehensif sesuai konteks MRO. Penelitian menggunakan pendekatan kualitatif deskriptif-analitis dalam paradigma design and development research. Data primer diperoleh dari dokumen evaluasi pelatihan, silabus dan materi pelatihan, serta observasi partisipatif selama masa magang. Analisis tematik digunakan sebagai teknik analisis utama, didukung analisis konten dan analisis kuantitatif deskriptif sebagai data pendukung. Hasil penelitian menunjukkan tidak adanya pengukuran pembelajaran, pelacakan perubahan perilaku, keterkaitan dengan hasil bisnis, dan perhitungan return on investment. Penelitian ini menghasilkan kerangka evaluasi terintegrasi yang mencakup seluruh lima level Kirkpatrick-Phillips, meliputi survei reaksi yang disempurnakan, pre-test dan post-test, penilaian perilaku 360 derajat, template pelacakan hasil, serta panduan perhitungan ROI, yang dikontekstualisasikan dalam industri MRO penerbangan dan setting tata kelola BUMN Indonesia. Disimpulkan bahwa evaluasi multi-level yang sistematis diperlukan untuk menjamin akuntabilitas, transfer pembelajaran, dan dampak organisasi yang terukur dalam program pengembangan kepemimpinan di industri keselamatan tinggi.

Kata Kunci: *kerangka evaluasi pasca pelatihan, model Kirkpatrick-Phillips, pengembangan kepemimpinan, MRO aviasi, imbal hasil investasi, efektivitas pelatihan*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

1.1 Background

Human resources constitute one of the most critical assets in any organization. Despite rapid technological advancement, humans remain the cornerstone of organizational success, making continuous development of employee knowledge and skills essential (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 impact operational safety, service quality, and business competitiveness. Mello (2014) emphasized that strategic human resource management, including systematic training and development programs, serves as a key driver of organizational performance and competitive advantage.

Training represents a structured approach to enhance learning efficiency and drive performance improvement, transforming individual capabilities from unsatisfactory to proficient, and from adequate to exceptional (Aziz, 2019). In the aviation MRO sector, effective training becomes even more critical due to the industry's unique characteristics: stringent regulatory requirements, complex technical systems, high safety stakes, and an increasingly competitive global market. Research indicated that MRO leadership requires not only deep technical expertise but also strategic thinking, effective communication, decision-making under pressure, and the ability to foster continuous improvement cultures (Aviation Pros, 2016). Developing such multifaceted capabilities demands well-designed training programs supported by robust evaluation mechanisms to ensure effectiveness.

However, 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 employees, a stark contrast to China (80%), the Philippines (60%), and

Vietnam (20%) (Aziz, 2019). This low training investment correlates with Indonesia's persistently lower labor productivity compared to regional peers, as evidenced by GDP measurements since 2011. The resulting skills gap between employer needs and workforce capabilities poses ongoing challenges to organizational competitiveness. In the aviation MRO industry specifically, workforce development challenges are intensifying globally due to aging workforces, a surge in 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 under the Human Resources Department has established a comprehensive portfolio of management development programs designed to build leadership and technical capabilities across the organization. This portfolio includes three principal programs: the Operational Leadership Program (OLP) targeting senior managers and high-potential talent in strategic leadership development; the MRO Management program providing managers and talent with cross-functional knowledge of MRO business operations; and the MRO Finance program equipping department heads and talent with financial literacy essential for business decision-making.

A review of L&TD evaluation documentation and participatory observation conducted during the research period confirms that GMF's post-training evaluation process is formally limited to Level 1 (Reaction) assessment within the Kirkpatrick-Phillips framework. Evaluation instruments administered following management development programs collect participant satisfaction ratings across several dimensions, including training content, facilitation quality, and logistical arrangements. No formal post-training instruments addressing knowledge acquisition (Level 2), behavioral transfer (Level 3), organizational results (Level 4), or return on investment (Level 5) are administered as standardized components of the training evaluation process.

Within the Kirkpatrick model, different organizational measurement mechanisms are commonly associated with higher levels of evaluation. For instance, 360-degree feedback assessments are widely recognized as instruments for measuring behavioral change at Level 3 (Behavior) because they capture observable changes in employee performance through evaluations provided by supervisors, peers, and subordinates following a learning intervention (Kirkpatrick & Kirkpatrick, 2016). Similarly, organizational performance indicators such as departmental Key Performance Indicators (KPIs) are typically associated with Level 4 (Results) because they reflect organizational outcomes that may result from improvements in employee capability and performance (Kirkpatrick & Kirkpatrick, 2016).

However, in the organizational context of GMF, these mechanisms are not utilized as part of the training evaluation process. The 360-degree competency assessment system is implemented within the company solely for employee performance appraisal purposes and is administered by a different organizational unit, independent of the Leader & Talent Development (L&TD) function. As a result, the assessment is not linked to specific training programs and is not conducted as a post-training measurement to capture behavioral changes resulting from management development initiatives.

A similar separation exists with respect to departmental KPI monitoring. Although KPI tracking reflects organizational performance outcomes conceptually aligned with Level 4 evaluation, these indicators are monitored through operational performance management systems and are not systematically connected to training participation data or program-level evaluation timelines. Consequently, KPI measurements are used for operational management purposes rather than for assessing the organizational impact of training interventions. As a result, post-training evaluation within the L&TD unit remains concentrated at the reaction level, without a structured mechanism to measure learning outcomes, behavioral transfer, or organizational results attributable to management development programs (Kirkpatrick & Kirkpatrick, 2016).

The limitations of Level 1-only evaluation extend beyond operational data gaps and create strategic risks for the organization, as recognized in Strategic HRM literature (Mello, 2014; Millmore et al., 2007). From this perspective, relying solely on participant satisfaction to evaluate management development programs generates several organizational vulnerabilities. First, it creates a risk of resource misallocation. When budget decisions are based only on satisfaction scores, significant investments in leadership development may be continued or expanded despite producing no measurable behavioral change or business impact, while programs that challenge participants and potentially produce deeper learning outcomes may be discontinued due to lower satisfaction ratings (Mello, 2014; Kirkpatrick & Kirkpatrick, 2016). In an organization the size of GMF, where management development programs involve substantial investment, such misallocation has clear financial implications. Second, reliance on Level 1 evaluation weakens the strategic accountability of the Leader and Talent Development (L&TD) function. Without evidence linking training investments to behavioral outcomes or organizational performance, the unit cannot build credible business cases for program funding within GMF's BUMN governance context, where financial justification for operational expenditures is routinely expected (Millmore et al., 2007). This limitation risks positioning the L&TD function as merely administrative rather than strategic, thereby reducing its influence in talent development and resource allocation decisions at the leadership level. Third, and most critically in the context of an aviation MRO organization, Level 1-only evaluation creates blind spots regarding leadership behavior transfer. Aviation MRO operations operate under strict regulatory oversight, including authorities such as the Directorate General of Civil Aviation (DGCA), where leadership competency gaps can affect safety and regulatory compliance (Aviation Pros, 2016). If evaluation instruments only measure whether participants found training satisfactory, the organization lacks evidence on whether safety-critical leadership behaviors, such as decision-making under pressure, regulatory adherence, and crew resource management, are actually applied in operational settings. As a result, behavioral deficiencies may remain undetected until they appear in audit findings,

safety incidents, or regulatory non-compliance events, when the cost of corrective action is significantly higher than if systematic evaluation and earlier intervention had been implemented.

The limitation of satisfaction-based evaluation is well-documented in training literature. Kirkpatrick (2016), whose four-level evaluation model has served as the field's gold standard for over six decades, emphasizes that while Level 1 (Reaction) assesses how participants feel about training, it provides no evidence of actual learning, behavior change, or organizational impact. Research by Bates (2004) critiqued the widespread tendency to equate participant satisfaction with training effectiveness, noting that positive reactions neither guarantee knowledge acquisition nor predict workplace application. The Kirkpatrick framework delineates four progressive evaluation levels: Level 1 (Reaction) measures satisfaction and perceived relevance; Level 2 (Learning) assesses knowledge and skill acquisition through tests and demonstrations; Level 3 (Behavior) evaluates workplace application of learned concepts; and Level 4 (Results) examines impact on organizational performance metrics such as productivity, quality, safety, or profitability. Phillips and Phillips (2016) subsequently added Level 5 (Return on Investment), comparing monetary benefits against training costs to provide financial justification for development investments.

Despite the comprehensive nature of this framework, research reveals that most organizations concentrate evaluation efforts at lower levels while neglecting higher-level assessment. A meta-analysis by Twitchell et al. (2000) examining training evaluation practices over 40 years found that 86-100% of companies conduct Level 1 evaluation, 71-90% conduct Level 2, but only 43-83% reach Level 3, and merely 21-49% implement Level 4. This pattern reflects the reality that Levels 1 and 2 are relatively easy and inexpensive to implement, while Levels 3 and 4 demand greater complexity, time investment, and resources (Bassi et al., 1996). However, research by Ahmad et al. (2024) demonstrated that comprehensive multi-level evaluation yields substantial organizational benefits, finding statistically significant relationships between reaction, learning, behavior, and

results that enable targeted program improvements enhancing both employee capability and organizational performance.

GMF's situation reflects a common organizational challenge: how to move beyond basic satisfaction measurement to comprehensive evaluation that demonstrates training value and drives continuous improvement. Current evaluation practices, while providing some useful feedback, may not generate the strategic insights needed to justify training investments to senior leadership, optimize program design, ensure knowledge transfer to workplace application, or link development activities to business outcomes. This research therefore seeks to investigate GMF's current training evaluation practices across its management development portfolio and develop a systematic framework for comprehensive assessment aligned with evidence-based evaluation standards.

The theoretical foundation for addressing GMF's evaluation gap lies in systematically implementing the Kirkpatrick-Phillips model within the organization's existing training infrastructure. This approach would enable comprehensive assessment spanning participant reactions, learning gains, behavioral changes, business results, and return on investment. Importantly, comprehensive evaluation need not be prohibitively resource-intensive. Phillips (2003) estimates that robust ROI evaluation typically requires only 3-5% of overall training budget, a modest investment offset by improved outcomes achieved through data-driven program refinement. Moreover, evaluation instruments can integrate into existing workflows: reaction surveys are already administered, pre/post knowledge tests add minimal session time, behavior follow-up leverages existing performance systems, and results tracking utilizes operationally-monitored metrics.

This research therefore proposes to develop a comprehensive post-training evaluation framework specifically designed for GMF's management development programs. Grounded in the Kirkpatrick-Phillips model and contextualized for the MRO operational environment, the framework will provide practical, implementable evaluation instruments spanning all five levels. By creating enhanced reaction assessments, learning measurement tools, behavior change

surveys, results tracking templates, and ROI calculation frameworks, this research aimed to bridge the gap between GMF's current practices and evidence-based evaluation standards. The framework will enable GMF to measure training effectiveness holistically, demonstrate organizational value, identify improvement opportunities, and make data-driven decisions about leadership development investments, ultimately ensuring that the organization's substantial training investment translates into measurable capability improvements and business results.

Beyond its practical contribution, this research offers a theoretically novel adaptation of the Kirkpatrick-Phillips framework to the aviation MRO context. Existing evaluation scholarship originates predominantly from Western corporate, healthcare, or military settings, leaving safety-critical MRO environments underexplored. MRO leadership evaluation carries distinct requirements that generic frameworks do not inherently accommodate: instruments must reflect the integration of technical and strategic competencies, regulatory accountability norms, and safety-conscious operational culture. A second contextual layer arises from the Indonesian BUMN governance setting, where hierarchical organizational norms, accountability imperatives, and resource constraints shape both instrument design and score interpretation. Together, these dual contextualization positions the resulting framework as a contribution to cross-cultural human resource development scholarship in non-Western, safety-critical industrial settings, not merely an implementation template.

1.2 Problem Foundation

Based on the background described above, several critical problems can be identified across GMF's training and development programs:

A. Limited Scope of Current Evaluation

GMF’s post-training evaluation is limited to Kirkpatrick Level 1 (Reaction). Although the organization uses a 360-degree competency assessment, corresponding to the behavioral dimension of Level 3, and tracks departmental KPIs that could inform Level 4, neither mechanism is integrated into the training evaluation cycle, so its baseline data cannot be used to measure behavioral change

(Kirkpatrick & Kirkpatrick, 2016). Similarly, KPI data is not linked to training participation or program timelines, preventing its use as an organizational results indicator within the Kirkpatrick-Phillips framework (Phillips, 2012). Therefore, the issue is not the absence of data but the lack of systematic integration connecting these data sources to training outcomes.

B. Absence of Learning Measurement (Level 2)

There is no systematic assessment of knowledge or skill acquisition during the training. GMF does not conduct pre-training baseline assessments or post-training knowledge tests. Consequently, the organization cannot:

1) Verify that participants gained the intended knowledge, skills, or attitudes from the training
2) Identify remaining knowledge or skill gaps that may require additional support or remedial learning
3) Demonstrate to stakeholders that the training investment produced measurable learning outcomes

This gap is explicitly recognized by participants themselves, as evidenced by feedback requesting pre- and post-assessment mechanisms.

C. No Behavior Change Tracking (Level 3)

Training evaluation concludes immediately upon program completion, with no follow-up to determine whether participants apply learned concepts in their daily work. GMF has no mechanism to:

1) Assess whether managers apply new leadership behaviors in their day-to-day work after returning to their departments
2) Identify workplace or environmental barriers that may hinder the transfer of learning into actual job performance
3) Provide post-training support and reinforcement to sustain behavior change over time

Participant feedback requesting "continues training" and "*update perkembangan perusahaan setelah training*" (updates on company developments after training) suggests desire for ongoing engagement that current practices do not address.

D. Lack of Business Results Linkage (Level 4)

GMF does not connect leadership training participation to organizational performance metrics. There is no system to:

1) Track business outcomes in departments led by training program graduates
2) Compare performance between trained and untrained managers to isolate training impact
3) Demonstrate that leadership development contributes to critical MRO performance metrics such as quality, productivity, retention, and operational efficiency

E. Inability to Calculate Return on Investment (ROI)

Without comprehensive evaluation data, GMF cannot determine whether the substantial investment in OLP, which includes instructor fees, participant time, materials, and facilities, generates adequate returns. The organization lacks:

1) Ability to conduct cost–benefit analysis of leadership development programs to determine net financial value
2) Convert training outcomes into monetary terms to provide quantified financial evidence for executive decision-making
3) Compare ROI across programs to prioritize, scale, or modify training investments based on financial return

Research Questions

To address the evaluation gap and bridge the disconnect between training expectations and measurement capabilities, this research seeks to answer the following questions:

1. What are the current training evaluation practices for management development programs at PT GMF Aero Asia and what gaps exist compared to comprehensive evaluation frameworks based on the Kirkpatrick-Phillips model?
2. How can comprehensive post-training evaluation framework based on the Kirkpatrick-Phillips model be designed and adapted for GMF's MRO

management development programs to measure learning, behavior change, business results, and return on investment?

3. What implementation guidelines and practical instruments are needed to operationalize the comprehensive evaluation framework within GMF's organizational context and resource constraints?

1.3 Research Objectives

Based on the problem foundation identified above, this research aims to achieve the following objectives:

1.3.1 Primary Objective

The objective of this research is to develop a comprehensive post-training evaluation framework for leadership training and development programs at PT GMF Aero Asia, encompassing the Operational Leadership Program (OLP), MRO Management, and MRO Finance, based on Kirkpatrick's Four Levels of Training Evaluation and Phillips' ROI Methodology.

1.3.2 Specific Objectives

This research has several specific objectives as follow:

1. To analyze GMF's current training evaluation practices across its management development programs and identify gaps in existing assessment mechanisms relative to the Kirkpatrick-Phillips framework.
2. To infer measurable learning objectives for each module across the OLP, MRO Management, and MRO Finance programs based on module content, participant feedback, and relevant competency requirements.
3. To design an enhanced Level 1 (Reaction) evaluation instrument that captures not only participant satisfaction but also learning confidence, behavioral intent, and environmental support factors, applicable across all three programs.
4. To develop Level 2 (Learning) assessment instruments, including pre-tests and post-tests calibrated to each program's learning objectives, that measure

knowledge and skill acquisition across the OLP, MRO Management, and MRO Finance programs.

5. To create Level 3 (Behavior) evaluation instrument utilizing 360-degree feedback methodology to assess leadership and competency behavior change 90 days post-training, prioritized for OLP and MRO Management.
6. To construct Level 4 (Results) tracking template that links management development program participation to relevant business metrics and key performance indicators in the MRO operational context.
7. To formulate Level 5 (ROI) calculation framework that enables GMF to quantify the return on investment of its leadership development programs, with selective application to flagship programs.
8. To provide differentiated implementation guidelines and recommendations for GMF's Leader & Talent Development unit to operationalize the proposed evaluation framework across future program cohorts.

1.4 Implications and Contributions

This research is expected to generate significant value across practical, theoretical, and methodological dimensions as follow:

1.4.1 Practical Contributions

1. Comprehensive Evaluation Tools

GMF will receive a complete set of ready-to-implement evaluation instruments for the Operational Leadership Program, including enhanced reaction surveys, pre/post knowledge tests, 360-degree behavior assessments, results tracking templates, and ROI calculation frameworks. These tools can be immediately deployed in future training cohorts.

2. Data-Driven Decision Making

The framework enables GMF's Leader & Talent Development unit to make evidence-based decisions about program improvements, content modifications, instructor selection, and resource allocation. Rather than relying solely on satisfaction feedback, L&TD will have objective data on learning effectiveness, behavioral outcomes, and business impact.

3. Accountability and Transfer of Learning

The 90-day behavior follow-up mechanism creates accountability for participants to apply learned concepts in their work. Knowing they will be assessed on behavioral change encourages active implementation rather than passive attendance.

4. Return on Investment Justification

The ability to calculate return on investment provides L&TD with compelling evidence to justify training budgets to senior management, demonstrate the value of leadership development to organizational performance, and secure continued investment in talent development initiatives.

5. Continuous Improvement Mechanism

Comprehensive evaluation data reveals specific strengths and weaknesses in training design and delivery, enabling systematic refinement of the training curriculum, teaching methods, and learning activities based on empirical evidence rather than intuition.

6. Participant Development Insights

Pre-test data helps identify individual knowledge gaps, allowing for targeted support or customized learning paths. Post-training assessments reveal which participants may need additional coaching or development interventions.

7. Adaptable Framework

While designed for GMF, the evaluation framework can be adapted by other MRO organizations seeking to strengthen leadership development evaluation practices, contributing to industry-wide professionalization of training assessment.

1.4.2 Theoretical Contributions

1. Contextual Application of Kirkpatrick-Phillips Model

This research demonstrates the practical application of the Kirkpatrick Four Levels + Phillips ROI framework in the specialized context of aviation MRO leadership development, contributing to literature on training evaluation in technical and safety-critical industries.

2. MRO Leadership Competency Framework

By inferring and validating learning objectives specific to MRO leadership roles, this research contributed to the limited academic literature on leadership competencies in the aviation maintenance sector, particularly for emerging markets like Indonesia.

3. Design Research Methodology

The research exemplified design and development research in human resource development, demonstrating how theoretical models can be translated into practical instruments through systematic design processes that balance academic rigor with organizational feasibility.

4. Evaluation in Resource-Constrained Settings

This research showed how comprehensive training evaluation can be implemented in organizations that lack dedicated evaluation specialists or sophisticated learning management systems, using accessible tools (Google Forms, Excel) and practical timelines.

CHAPTER II
METHODOLOGY

2.1 Company Profile

Source: GMF Aero Asia (n.d)

Figure 2.1. PT GMF Aero Asia Tbk Logo

The history of PT Garuda Maintenance Facility Aero Asia Tbk (the "Company" or "GMF") began with the establishment of Garuda Indonesia's Technical Directorate in 1949. After more than four decades of operating as one of the divisions under Garuda Indonesia, GMF was transformed into the Strategic Business Unit Garuda Maintenance Facility (SBU-GMF), which handles all maintenance activities of Garuda Indonesia's fleet. In 2002, Garuda Indonesia spun off SBU-GMF into a subsidiary. The Company thus officially became a subsidiary of Garuda Indonesia, under the name PT Garuda Maintenance Facility (GMF) Aero Asia Tbk.

As the leading Maintenance, Repair, and Overhaul (MRO) Company in Indonesia, GMF owns Line Maintenance including Overhaul, maintenance and repair of engines and component, modification processes, and cabin refurbishment. The Company continuously strengthens its core business by adding a number of additional maintenance services and various international standard facilities. GMF officially became the first issuer of the MRO industry in Indonesia to conduct an Initial Public Offering (IPO) to the public on October 2, 2017, and its shares were officially listed on the Indonesia Stock Exchange on October 10, 2017. This momentum became an important milestone for the Company to further strengthen GMF's position to become the Most Valuable MRO Company. Since its

establishment on April 26, 2002, the Company has continued to use the name PT Garuda Maintenance Facility Aero Asia. On October 10, 2017, the Company officially listed its shares on the Indonesia Stock Exchange (IDX), becoming a public company (Tbk), as reflected in its name change to PT Garuda Maintenance Facility Aero Asia Tbk.

2.1.1 Company Vision, Mission & Value

As a leading provider of aircraft maintenance and repair services, GMF Aero Asia operates with a clear and well-defined foundation guided by its vision, mission, core values, and structured organizational culture. These elements are outlined in the Company's Long-Term Plan (RJPP), most recently updated on September 27, 2023. In this update, GMF reaffirmed its strategic direction and commitment through the "GMF Roadmap 2020–2024." This roadmap serves as a blueprint for the company's business growth and transformation initiatives.

A. Vision

Reflecting the company's aspiration to become the most valued aircraft maintenance service provider in the eyes of customers, partners, and global stakeholders, GMF's vision is:

"Most Valuable MRO Company"

GMF has achieved one of its targets in business development by gaining recognition as a global MRO player, reaching the Top 10 MRO in 2019. The Company's strategic trajectory extends beyond this milestone. The Company is committed to further enhancing business value in the eyes of the public, both internally and externally, by diversifying its business activities beyond solely relying on the aviation sector.

B. Mission

GMF's mission captures the company's determination to provide integrated and reliable maintenance solutions as a tangible contribution to the nation:

"Integrated and Reliable Maintenance Solution as a Contribution to the Nation"

To achieve its mission, GMF upholds three key principles: maintaining high-quality, standardized maintenance to ensure airworthiness at competitive cost; continuously developing employee competence and facilities to deliver reliable and professional services; and fostering teamwork, safety, and mutual respect in an open work environment. These principles demonstrate GMF's commitment to providing integrated and dependable maintenance solutions.

C. Value

Based on the contents of the Circular Letter of the Minister of State-Owned Enterprise (SOE)/BUMN Number: SE-7/MBU/07/2020 dated July 1, 2020, every SOE, its subsidiaries, and affiliated companies are obliged to implement the Core Values of SOE Human Resources, which is called AKHLAK, become the Corporate Culture. Therefore, as a subsidiary, GMF has adopted the values of AKHLAK as the Corporate Values from 2020 until this moment.

The core values consist of *Amanah* (Trustworthy), *Kompeten* (Competent), *Harmonis* (Harmonious), *Loyal* (Loyal), *Adaptif* (Adaptive), and *Kolaboratif* (Collaborative). *Amanah* emphasizes integrity, accountability, honesty, and adherence to moral and ethical principles. *Kompeten* reflects a commitment to continuous learning, capability development aligned with current trends, and delivering high performance and outstanding achievements. *Harmonis* highlights caring for others, respecting differences, valuing diverse opinions, and appreciating contributions regardless of background. *Loyal* represents dedication and commitment to prioritizing organizational and national interests, demonstrating strong responsibility and willingness to contribute beyond expectations. *Adaptif* encourages innovation, agility, openness to change, and proactive transformation. *Kolaboratif* promotes open collaboration, synergy, and teamwork to create added value and achieve shared objectives.

2.1.2 Company Products/Services

GMF categorizes its business into two segments, namely (1) Reparation and Overhaul and (2) Maintenance.

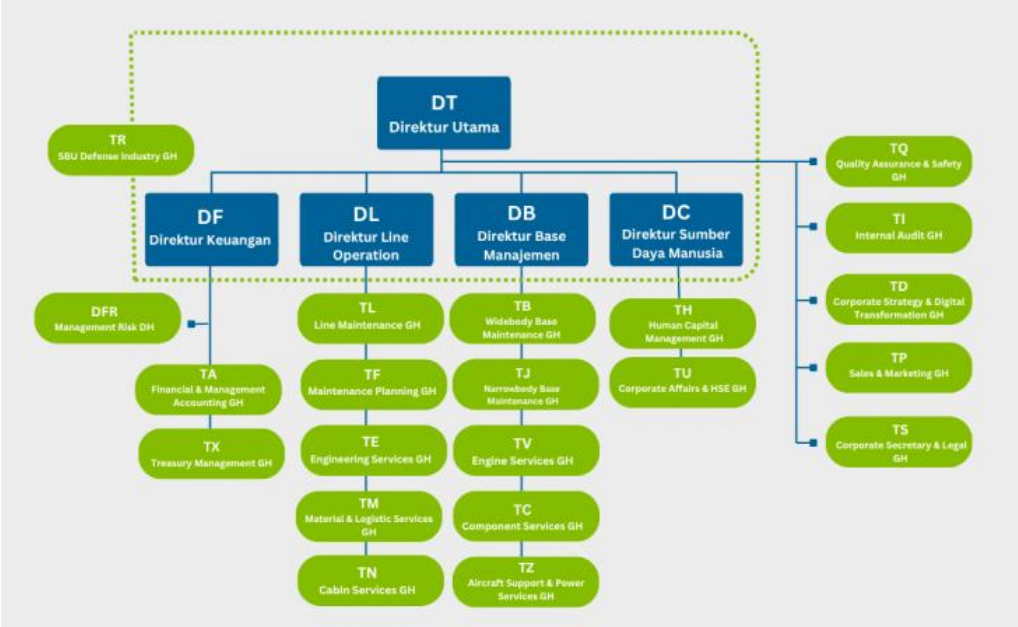
A. Reparation and Overhaul

GMF provides integrated Maintenance, Repair, and Overhaul (MRO) services across four main business lines: Airframe, Engine, Component, and Military & Defense. In Airframe, GMF operates four hangars capable of performing routine heavy checks, major modifications, structural repairs, cabin refurbishment and reconfiguration, exterior painting, and full aircraft overhaul, with certifications from DKPPU, FAA, EASA, CAA UK, and CASA for various aircraft types including A320/A320NEO, A330/A330NEO, B737 series (CL/NG/MAX), B747, B777, CRJ1000, and ATR72. In Engine services, GMF manages an Engine Workshop and Engine & APU Test Cell to maintain engines and APUs such as CFM56 series, PW100, and GTCP series supporting B737, A320, ATR, and A330 aircraft. Through its Component Services Unit, GMF delivers repair and overhaul for instruments, avionics, navigation systems, flight data recorders, gyros, and Wheel Brake & Landing Gear, supported by specialized workshops and advanced testing systems. In Military & Defense, GMF offers comprehensive airframe, engine, component, and support solutions, including C130 maintenance and modification, MRTT conversion, T56 engine overhaul, fighter aircraft maintenance, Boeing Business Jet services, military CFM56-3 and APU overhauls, as well as Ground Support Equipment manufacturing and maintenance.

B. Maintenance

GMF's Line Maintenance & Technical Ground Handling services support both domestic and international flights at Soekarno-Hatta Airport, delivering comprehensive light maintenance solutions including Pre-Flight Checks, Transit Checks, Daily Checks, A Checks (up to 600 flight hours), and other routine maintenance activities. The service covers various aircraft types such as B737 (including MAX), B747, B777, B787, A320/A320NEO, A330/A330NEO, A350, CRJ1000, and ATR72, as well as providing overnight transit support and emergency Aircraft on Ground (AOG) services. Operational reliability is strengthened through the Maintenance Control Center (MCC), which monitors maintenance activities to minimize unscheduled maintenance and technical delays, supported by an extensive network of 69 line stations across Indonesia.

2.1.3 Company Structure



Source: GMF Aero Asia (n.d)

Figure 1.2. PT GMF Aero Asia Tbk Company Structure

In 2024, GMF implemented a Parent Organization Structure based on President Director's Decree No. DT/SKEP-5015/24 dated August 21, 2024. PT GMF Aero Asia Tbk adopted a structural organization consisting of five directorates and 21 departments, managing around 5,000 employees under the leadership of the Chief Executive Officer (CEO).

The Director of Finance oversees financial planning, treasury and cash flow management, financial and management reporting, funding and investment planning, investor relations, and corporate tax management. The Director of Line Operation is responsible for daily aircraft maintenance operations, including line maintenance execution, maintenance scheduling, engineering services, material and logistics management, cabin maintenance, and ensuring aircraft airworthiness. The Director of Base Operation manages heavy maintenance activities performed in hangars, covering wide-body and narrow-body aircraft maintenance, engine services, component maintenance, specialized services, aircraft modifications, and industrial support solutions. The Director of Human Capital & Corporate Affairs

oversees human resource management and corporate affairs, including workforce planning, recruitment, remuneration, career development, industrial relations, general affairs, HSE implementation, corporate communication, and legal compliance

2.2 Research Approach

This research employed qualitative approach with a descriptive-analytical study design. The qualitative approach was chosen because it allowed the researcher to gain in-depth understanding of training evaluation practices at PT GMF Aero Asia and to develop a comprehensive evaluation framework grounded in contextual understanding of organizational needs, MRO industry characteristics, and the company's operational realities.

Qualitative research is exploratory and interpretive in nature, which aligns with the research objectives to investigate current training evaluation practices, identify gaps between existing practices and evidence-based evaluation standards, and design practical solutions tailored to GMF's specific context (Creswell & Creswell, 2018). This approach enabled the researcher to capture nuances, complexities, and contexts that cannot be adequately measured through quantitative methods alone (Lincoln & Guba, 1985).

Specifically, this research adopted the design and development research paradigm, which was a research approach focused on creating new products, tools, or frameworks to address practical problems in organizational settings (Richey & Klein, 2007). The DDR paradigm adopted in this study is substantively distinct from a conventional qualitative case study approach, and this distinction has methodological implications for how the research design, outputs, and quality criteria are understood. A qualitative case study, as defined by Creswell and Creswell (2018), aims to produce rich, bounded descriptive understanding of a phenomenon within its natural context. Its primary output is a set of interpretive findings or thematic analyses, and its quality is evaluated according to criteria of trustworthiness and transferability. Had this study been designed as a qualitative case study of GMF's training evaluation practices, it would have concluded with a

descriptive account of how evaluation is currently conducted, the perceptions of stakeholders, and the contextual factors shaping evaluation behavior -- but it would not have produced a designed artifact or implementation system.

DDR, by contrast, is explicitly artifact-oriented. As articulated by Richey and Klein (2007, p. 1), the paradigm is defined by a dual commitment to generating knowledge and producing a validated practical product. In DDR, empirical findings from the diagnostic phase are not terminal outputs; they function as inputs that directly drive the design of a practical solution. The quality of a DDR study is therefore evaluated not only on the analytical rigor of its diagnostic inquiry but also on the utility, feasibility, and contextual appropriateness of the designed artifact for implementation by the intended organizational user (Richey & Klein, 2007). This research follows a three-phase DDR structure that maps directly onto this logic: the diagnostic inquiry phase (Chapter III) establishes the evaluation gap and its organizational context; the theoretical synthesis phase (Chapter IV) establishes the design rationale and framework architecture; and the artifact development phase (Chapter V) produces the implementable evaluation instruments and deployment roadmap. This structure is fundamentally different from the bounded descriptive inquiry characteristic of qualitative case study research and justifies the DDR classification given the study's objective of producing an actionable, organizationally contextualized evaluation system rather than a descriptive account of existing practices. This approach is consistent with DDR applications in the HRD and management education literature, where the paradigm has been employed to develop training systems, competency frameworks, and performance improvement tools that integrate empirical context analysis with practical design outputs (Richey & Klein, 2007). In this context, the product being developed was comprehensive post-training evaluation framework based on the Kirkpatrick-Phillips Model (Kirkpatrick & Kirkpatrick, 2016; Phillips & Phillips, 2016), completed with practical instruments that can be implemented by GMF's Leader & Talent Development unit. The qualitative approach in this research enabled the researcher to:

1. Gain in-depth understanding of the training evaluation context at GMF through analysis of existing evaluation documents, observations of practices during the internship period, and understanding of organizational needs and constraints. This contextual understanding is critical because an effective evaluation framework must be adapted to GMF's operational realities, organizational culture, and resource capacity, rather than merely applying evaluation theory generically (Creswell & Creswell, 2018).
2. Identify patterns, themes, and insights from qualitative data such as participant feedback, comments and suggestions in existing evaluations, and characteristics of training content. Through thematic analysis of this data (Braun & Clarke, 2006), the researcher can identify evaluation needs based not only on academic literature but also on actual perceptions and experiences of stakeholders at GMF.
3. Design practical and implementable evaluation instruments by considering contextual factors such as available time, L&TD staff expertise, accessible technology (e.g., Google Forms, Microsoft Excel), and existing evaluation culture. The qualitative approach allows flexibility in design that is responsive to organizational needs and constraints (Richey & Klein, 2007).
4. Conduct expert validation of the designed instruments through review by academic supervisors and when possible input from practitioners at GMF. This qualitative validation process ensures that the developed instruments possess content validity and contextual appropriateness (Lincoln & Guba, 1985).

Although this research was qualitative in nature, quantitative data was used as supporting data to provide a descriptive picture of current evaluation practices. Quantitative data in the form of descriptive statistics from participant satisfaction scores were analyzed to describe the current state of Level 1 evaluation at GMF (Sugiyono, 2019). However, this quantitative data was not used for hypothesis testing or inferential analysis, but solely to provide context and support qualitative findings regarding the limitations of satisfaction-focused evaluation (Creswell & Creswell, 2018).

Thus, the descriptive-analytical qualitative approach with a design and development research paradigm was chosen in this research to gain in-depth understanding of training evaluation practices at GMF, identify evaluation gaps, design a comprehensive evaluation framework based on the Kirkpatrick-Phillips Model (Kirkpatrick & Kirkpatrick, 2016; Phillips & Phillips, 2016), and produce practical instruments that can be implemented in the MRO organizational context while considering the uniqueness of the aviation industry, technical complexity of aircraft maintenance operations, and the need for evaluation that measures not only satisfaction but also learning, behavior change, and business impact.

2.3 The Unit of Analysis

The unit of analysis for this research was situated at the division level, namely the Leader & Talent Development (L&TD) unit within the Human Resources Department of PT GMF Aero Asia. This unit oversees the end-to-end management of training and leadership development initiatives, encompassing the design, development, delivery, and evaluation of programs. The research concentrated on three flagship programs administered by L&TD: the Operational Leadership Program (OLP), MRO Management, and MRO Finance. The selection of L&TD as the analytical focus was grounded in its role as the process owner of training evaluation at GMF, making its existing practices, identified shortcomings, and the proposed evaluation framework directly aligned with its functional mandate. Moreover, the unit maintained the primary evaluation data utilized in this research and was responsible for implementing the resulting framework, thereby necessitating thorough understanding of its operational context, resources, and limitations to ensure feasibility and long-term applicability.

Although the primary level of analysis was the L&TD unit, the empirical data examined in this research was derived from three training programs involving participants from multiple directorates, including Human Capital & Corporate Affairs, Finance, Line Operations, and Base Operations. Consequently, while the research emphasized the evaluation practices and institutional needs of L&TD, it also incorporates insights from participants across various operational domains.

This broader perspective ensured that the proposed evaluation framework reflected cross-directorate experiences and supported the development of a more integrated and comprehensive training evaluation system.

The selection of the L&TD unit as the unit of analysis, rather than individual training participants, is grounded in the study's organizational research purpose and aligns with the Design and Development Research (DDR) paradigm (Richey & Klein, 2007). While individual participants may serve as a valid unit of analysis in studies examining learning outcomes or behavioral change, the research questions in this study operate at the organizational system level. They examine GMF's current evaluation practices, the design of a multi-level evaluation framework for management development programs, and the feasibility of implementing that framework within the organization. These questions concern institutional processes and system architecture rather than individual performance, making the L&TD unit the appropriate analytical focus. Consistent with DDR principles, the developed artifact is intended for institutional use, meaning the proposed evaluation framework is designed to be implemented and managed by the L&TD unit (Richey & Klein, 2007). Although participant-level data, including 638 quantitative responses and 584 qualitative comments, inform the diagnostic analysis, these data are analyzed in aggregate as organizational evidence rather than as individual learning records.

2.4 Data Sources & Data Collection Techniques

This research used two types of data sources: primary data and secondary data. The combination of both types enabled triangulation to enhance the credibility and depth of research findings. The research design prioritizes document-based and observational data over formal stakeholder interviews for two reasons. First, the available data sources were sufficient for the diagnostic and design objectives of this DDR study. The dataset includes 638 quantitative evaluation responses that reveal satisfaction patterns across the three programs and highlight the limitations of Level 1 evaluation. In addition, 584 qualitative participant comments provide insights into training quality, content relevance, and facilitation issues. Training

syllabi supply documentary evidence of program objectives and intended learning outcomes, enabling alignment analysis between program design and current evaluation practices. Participatory observation during the internship also allowed verification of evaluation workflows and organizational constraints not visible through documents alone (Creswell & Creswell, 2018). Together, these sources were adequate for diagnosing the evaluation gap and informing the framework design (Richey & Klein, 2007). Second, conducting formal interviews within an active internship relationship could introduce social desirability bias due to organizational power dynamics, potentially limiting candid responses (Creswell & Creswell, 2018). Relying on documents and observation mitigated this risk, though the absence of structured interviews remains a limitation for future research.

2.4.1 Primary Data

Primary data is data collected directly by the researcher for the specific purpose of this research. Primary data in this research included:

A. Training Evaluation Documents

Evaluation documents from three training programs that were analyzed including Operational Leadership Program (OLP) (2025), MRO Management (2025), and MRO Finance (2025). These documents included:

1) Training evaluation reports containing summary of participant satisfaction scores, inter-module comparison charts, and compilation of qualitative feedback
2) Raw evaluation data in Microsoft Excel format containing individual participant responses for each evaluation question

Collection Technique:

These documents were obtained during the internship at GMF's Leader & Talent Development (L&TD) unit. With the supervisor's approval, the researcher accessed the unit's internal document system and collected relevant evaluation files in digital format for analysis. Additionally, the researcher processed raw evaluation data in Excel and converted it into finalized PDF evaluation reports as part of the unit's standard evaluation workflow.

B. Training Syllabus and Materials

Training syllabi, lesson plans, and presentation materials from modules was done for all three training programs. These documents were used to understand training content and infer learning objectives that form the basis for designing Level 2 (Learning) evaluation instruments.

Collection Technique:

Syllabus documents and training materials were obtained from the L&TD unit's digital archives and from training instructors through coordination with the L&TD team during the internship period.

C. Participatory Observation

This observation was done by direct observation of training and evaluation processes that occurred during the researcher's internship period at GMF. This observation provided contextual understanding of how training is conducted, how evaluation was performed, and how evaluation data was used by the L&TD unit.

Collection Technique:

The researcher conducted participatory observation by engaging in L&TD unit activities during the internship period, including assisting in training delivery and the evaluation process. Field notes were made to document observations related to evaluation practices, challenges faced by the L&TD unit, and identified needs.

2.4.2 Secondary Data

Secondary data is data that existed previously and was collected for other purposes but is relevant to this research. Secondary data in this research included:

A. Academic Literature

Academic literature includes textbooks, scholarly journal articles, and academic publications on training evaluation (specifically the Kirkpatrick Model and Phillips ROI Methodology), ADDIE instructional design model, leadership competencies in the MRO industry, adult learning theory (andragogy), and principles of assessment instrument design.

Collection Technique:

Literature search was conducted through academic databases (Google Scholar, ResearchGate), university digital libraries, and credible online sources. Relevant literature was identified through keyword searches and snowballing technique (tracking references from relevant articles).

B. MRO Industry Sources

MRO industry sources include industry publications, trend reports, and articles from credible sources regarding workforce challenges in the MRO industry, competencies required for MRO leadership, and best practices in leadership development in the aviation sector.

Collection Technique:

Industry source search was conducted through aviation industry organization websites (e.g., Aviation Pros, Aircraft-parts), MRO consulting firm publications, and credible industry media.

C. GMF Organizational Documents

GMF documents included annual report or integrated report, organizational structure, job descriptions for managerial positions, and company strategic documents relevant to human resource development.

Collection Technique:

Organizational documents were obtained through access provided during the internship period, while adhering to confidentiality policies and obtaining GMF's approval for document use in academic research.

2.5 Validity and Reliability Testing

In qualitative research, the concepts of validity and reliability have different meanings compared to quantitative research. Qualitative research emphasizes credibility, transferability, dependability, and confirmability as research quality criteria (Lincoln & Guba, 1985).

2.5.1 Validity

Validity in qualitative research refers to the extent to which research findings accurately and credibly reflect the reality being studied. In this research, validity was ensured through several strategies:

A. Data Source Triangulation

This research used multiple data sources to validate findings: evaluation documents from three different training programs, training syllabi and materials, direct observation during internship, and academic and industry literature. Consistency of findings across different data sources enhanced the credibility of research conclusions.

B. Content Validity of Instruments

The evaluation instruments designed in this research (Level 1-5 evaluation instruments) are validated through:

1) Practitioner review: Given the researcher's embedded position as an intern within the L&TD unit, instrument components were iteratively refined through ongoing professional consultation with L&TD unit members throughout the internship period. Feedback from daily operational involvement, coordination with supervisors, and direct engagement with training activities informed the instrument design process, ensuring alignment with GMF's operational context and evaluation capabilities.
2) Expert validation: Instruments were reviewed by academic supervisors with expertise in Human Resources field.
3) Literature grounding: Instrument design was based on validated evaluation theory (Kirkpatrick-Phillips Model) and best practices documented in academic literature.

C. Thick Description

This research provided thick description of GMF's context, training program characteristics, existing evaluation practices, and the framework design process. This detailed description enabled readers to assess the transferability of findings to other contexts and understand the basis of research conclusions.

D. Audit Trail

The researcher systematically documented the entire research process, including data collection steps, analysis process, design decisions, and the rationale behind methodological choices. This documentation allowed other researchers or supervisors to trace the research logic and assess the validity of conclusions.

2.5.2 Reliability

Reliability in qualitative research refers to the consistency and dependability of the research process. In this research, reliability was ensured through the following:

A. Systematic Data Analysis Procedures

This research used systematic and well-documented data analysis procedures. Thematic analysis was conducted following clear stages: familiarization with data, initial coding, theme identification, theme review, and theme definition. This process was documented so it can be replicated by other researchers.

B. Consistency in Instrument Design

The designed evaluation instruments followed consistent theoretical framework (Kirkpatrick-Phillips Model) and used standard instrument design principles (e.g., clear question construction, consistent use of rating scales, alignment with learning objectives).

C. Peer Debriefing

The researcher conducted discussions with academic supervisors and when possible, with peers to discuss findings, interpretations, and design decisions. This peer debriefing process helped identify potential researcher bias and ensured consistency in analysis and conclusions.

2.6 Data Analysis Techniques

This research used qualitative data analysis approach with the main technique being Thematic Analysis to analyze qualitative data, supplemented by Descriptive Quantitative Analysis as supporting data to describe the current training

evaluation conditions. Thematic analysis is a qualitative analysis method used to identify, analyze, and report patterns (themes) in qualitative data (Braun & Clarke, 2006). This technique was chosen for its flexibility and ability to provide in-depth analysis of qualitative data such as training participant feedback, evaluation comments, and training content. The thematic analysis process in this research followed the six stages recommended by Braun and Clarke (2006):

2.6.1 Thematic Analysis

A. Stage 1: Familiarizing with Data

The researcher repeatedly read and examined all training evaluation documents, participant feedback, and training materials to gain in-depth understanding of the content and context of the data. Initial notes were made to record ideas and patterns that emerge during the reading process.

B. Stage 2: Initial Coding

Qualitative data was systematically coded to identify features that were interesting and relevant to the research questions. Codes were created to capture important aspects of the data, such as:

1) Codes for participant feedback themes (e.g., "request for more time", "need for practice", "request for videos", "content too fast")
2) Codes for current evaluation characteristics (e.g., "satisfaction focus", "no pretest", "no follow-up")
3) Codes for evaluation needs (e.g., "learning measurement", "behavior tracking", "business impact demonstration")

C. Stage 3: Searching for Themes

Identified codes were grouped into broader potential themes. These themes represented important patterns in the data that were relevant to the research questions. For example, codes like "request for more time", "content too dense", "limited discussion time" can be grouped into the theme "Training Time Constraints".

D. Stage 4: Reviewing Themes

Identified themes were reviewed to ensure they were truly supported by the data and internally coherent. Some themes may be merged, split, or discarded based on this review. The review process involved:

1) Ensuring data within each theme is coherent and meaningful.
2) Ensuring themes are clearly distinct from each other.
3) Validating themes by returning to raw data.

E. Stage 5: Defining and Naming Themes

Each theme was clearly defined and given a name that captures its essence. Theme definitions explained what the theme means and why it was important for answering the research questions.

Examples of themes identified in this research are the following:

1) Theme 1: Limitations of Satisfaction-Only Evaluation
    a. Includes data on how Level 1 evaluation cannot measure actual learning or behavior change
2) Theme 2: Participant Need for More Comprehensive Assessment
    a. Includes explicit requests from participants for pre/post assessment
3) Theme 3: Variation in Participant Baseline Knowledge
    a. Includes feedback about content being too fast or too slow, indicating need for pre-assessment
4) Theme 4: Gap Between High Satisfaction and Unmeasured Impact
    a. Includes patterns where training receives high ratings but its performance impact is unknown

F. Stage 6: Producing the Report

Defined themes were reported with representative data quotations to support each theme. This thematic analysis report becomes part of Chapter 3 (Research Findings) in the final report.

2.6.2 Descriptive Quantitative Analysis

As supporting data, this research also used descriptive quantitative analysis to describe the current state of Level 1 evaluation practices at GMF. Creswell and

Creswell (2018) defined descriptive quantitative analysis as a method used to summarize and describe the characteristics of a dataset through numerical measures, enabling researchers to identify patterns and tendencies within the data without extending inferences beyond the observed sample. Complementarily, Sugiyono (2019) further specified that descriptive quantitative analysis involves the use of statistical measures, including frequency distribution, percentage, mean, and standard deviation, to systematically summarize and interpret numerical data collected from a sample or population. Consistent with both frameworks, this research applied descriptive quantitative analysis to numerical data from GMF's training evaluation surveys, including:

A. Descriptive Statistics

1) Calculation of mean scores for each evaluation question (e.g., average instructor rating, average material rating)
2) Calculation of mode for ordinal scale questions (e.g., questions about time adequacy: 1=too short, 2=adequate, 3=too long)
3) Identification of highest and lowest scores to identify modules or aspects with best performance and those needing improvement

B. Inter-Program Comparison

1) Comparison of average satisfaction scores across the three training programs (OLP, MRO Management, MRO Finance)
2) Identification of patterns: which program received highest/lowest ratings, which aspects are consistent across programs, which aspects vary

C. Inter-Module Comparison

1) Analysis of satisfaction score variation across different modules within each program
2) Identification of modules with highest and lowest ratings to inform needs for deeper evaluation

It is important to note that this quantitative data was not the main focus of the research and was not used for statistical hypothesis testing or inferential

analysis. Quantitative data merely served as supporting data to provide a descriptive picture of the current evaluation state, which was then integrated with qualitative findings to build more comprehensive understanding of evaluation practices and needs at GMF.

2.6.3 Content Analysis

In addition to thematic analysis, this research also used content analysis to analyze training syllabi and materials. Content analysis was conducted for the following:

A. Learning Objectives Inference

Analyzing the content of each training module to infer learning objectives that form the basis for designing Level 2 evaluation instruments. This process involved:

1) Identifying key concepts taught in each module
2) Determining targeted cognitive levels (based on Bloom's Taxonomy: remember, understand, apply, analyze, evaluate, create)
3) Formulating learning objectives that are specific, measurable, achievable, relevant, and time-bound (SMART)

B. Alignment Analysis

Analyzing the alignment among training content, delivery methods, and existing evaluation. This analysis identified potential misalignments, such as:

1) Complex content that is not adequately evaluated
2) Content emphasized in training but not reflected in evaluation
3) Gaps between what is taught and what is measured

2.6.4 Triangulation and Synthesis

All findings from various analysis techniques (thematic analysis, descriptive quantitative analysis, content analysis) were triangulated and synthesized to produce holistic understanding of current training evaluation practices at GMF:

1) Gaps between existing practices and comprehensive evaluation standards
2) Specific needs for an evaluation framework adapted to GMF's context

This synthesis formed the basis for designing a comprehensive, practical, and evidence-based evaluation framework presented in Chapter 5 (Discussion).

2.7 Researcher Reflexivity and Positionality

This research was conducted by an active intern within GMF's Leader & Talent Development (L&TD) unit, the same unit whose evaluation practices were being examined. While this insider position enabled direct access to internal documents and operational workflows, it also presented potential bias risks, including insider bias (the tendency to interpret findings in ways that support organizational change), confirmation bias (as the framework designer may perceive results as validating its necessity), and social desirability bias (where participants provide favorable feedback due to institutional context, reflected in the prevalence of brief affirmative comments). To address these risks, the research applied data triangulation across three programs and multiple data sources, engaged in peer debriefing with the academic supervisor, actively sought disconfirming evidence, and anchored all recommendations in established theoretical frameworks rather than relying solely on internal organizational perspectives.

2.8 Ethical Consideration

This research was conducted in accordance with the ethical standards applicable to qualitative and design-based inquiry in organizational settings. The use of internal GMF documents, evaluation data, and training materials was undertaken with the explicit approval of the researcher's internship supervisor within the Leader & Talent Development unit, consistent with the access authorization described in Section 2.4.1. All participant evaluation data analyzed in this research were collected as part of GMF's standard post-training evaluation procedures prior to this research; no additional data were collected from participants specifically for research purposes, and no identifiable individual participant data were disclosed in this report. Quantitative data were presented exclusively in aggregated form and qualitative comments were reported thematically without

attribution to specific individuals. The researcher's dual role as intern and researcher was acknowledged and addressed through the reflexivity measures described in Section 2.7.

CHAPTER III
RESEARCH FINDINGS

3.1 Introduction

This chapter presents findings from an analysis of training evaluation practices at PT Garuda Maintenance Facility (GMF) Aero Asia, focusing on three management development programs: the Operational Leadership Program (OLP), MRO Management, and MRO Finance. Data were collected during the researcher's internship with GMF's Leader & Talent Development (L&TD) unit, including evaluation documents, participant response data, and training program materials.

The findings were organized into four sections. First, GMF's current training evaluation system was described, covering its procedures, instruments, and workflows. Second, descriptive statistics from 638 evaluation responses were presented, revealing patterns in participant reactions to instructor quality, material relevance, and training duration. Third, thematic analysis of 584 qualitative comments identified recurring feedback patterns and implicit evaluation needs. Fourth, the chapter highlighted the gap between GMF's current Level 1 evaluation and the comprehensive Kirkpatrick-Phillips framework, providing the empirical basis for the framework development discussed in subsequent chapters.

The analysis focused not only on what GMF currently measures, but also on what remained unmeasured and where gaps existed between current practices and evidence-based evaluation standards.

3.2 Overview of Current Training Evaluation Practices at GMF

3.2.1 Institutional Context and L&TD Unit Role

PT Garuda Maintenance Facility Aero Asia operates as Indonesia's premier aviation MRO provider, managing approximately 5,000 employees across four directorates (Finance, Line Operation, Base Operation, and Human Capital & Corporate Affairs). Within the Human Capital Management Division, the Leader & Talent Development (L&TD) unit bears primary responsibility for designing,

delivering, and evaluating all management development programs. The unit's portfolio encompasses three principal leadership development initiatives examined in this research: the Operational Leadership Program targeting senior managers and high-potential talent for strategic leadership roles, the MRO Management program providing cross-functional business knowledge, and the MRO Finance program equipping department heads with financial literacy for business decision-making.

A. GMF's Current Training Practices: Absence of Systematic Framework

Despite implementing various sophisticated training methods, GMF currently operates without a systematic instructional design framework organizing these activities into a coherent training development cycle. L&TD conducts multiple training-related activities, yet these occur somewhat independently rather than as integrated components of a structured process. Specifically, GMF's current practices include:

1. **Pre-Training Activities**
   a. **Training Needs Analysis (TNA)**: L&TD conducts need assessment at organizational, job, and individual levels to identify development requirements aligned with GMF's strategic objectives and operational challenges. However, this TNA process lacks formal connection to subsequent evaluation, with no systematic verification that identified needs were actually addressed through training outcomes.
   b. **360-Degree Pre-Assessment**: Selected programs employ 360-degree competency assessment to establish baseline leadership capabilities. However, without corresponding post-training 360-degree follow-up, these baseline assessments cannot demonstrate behavior change or training impact.
2. **Training Delivery Methods**
   a. **Classroom Instruction with Collaborative Elements**: Multi-day workshop formats emphasizing interactive discussion, case analysis, group problem-solving, and peer learning rather than passive lecture consumption.

b. **Gamification Integration**: Incorporation of interactive challenges, scenario-based exercises, and workshop activities designed to enhance engagement and practical application.
c. **E-Learning Components**: Digital content delivery and online resources enabling flexible access to supplementary materials.
d. **On-the-Job Training (OJT) Planning**: Training sessions include planning for workplace application, enabling participants to identify opportunities for implementing learned concepts in their operational contexts.

3. Post-Training Evaluation: The Critical Gap

GMF's current post-training evaluation remains limited exclusively to Kirkpatrick Level 1 (Reaction assessment) through satisfaction surveys administered immediately upon training completion. As documented throughout this chapter, systematic absence exists regarding learning verification (Level 2), behavioral application tracking (Level 3), business results measurement (Level 4), and return on investment calculation (Level 5).

B. The Missing Framework

GMF currently lacks a systematic training development framework that integrates needs assessment, design, delivery, and evaluation into a coherent end to end process. Although the organization applies evidence-based methods such as robust TNA practices, multi modal delivery, collaborative learning, gamification, e learning, and pre training 360-degree assessments, these elements operate in isolation. As a result, TNA findings are not validated through learning assessment, delivery methods are not evaluated for effectiveness, 360-degree assessments lack post training comparison, and training investments cannot be linked to measurable organizational impact. The core issue is not the quality of individual activities, but the absence of an overarching framework connecting inputs to outcomes. This research proposed adopting the ADDIE model to systematize the process, particularly by strengthening the evaluation component while leveraging existing strengths. The urgency was heightened by GMF’s context as a BUMN subsidiary facing governance and performance accountability pressures aligned with its 2024 strategic priorities of profitability, debt restructuring, and equity improvement,

while operating with limited evaluation expertise, no dedicated LMS infrastructure, and constrained resources, which reinforced the need for a structured and sustainable evaluation framework (GMF Integrated Report, 2024).

3.2.2 Current Evaluation Process and Workflow

GMF's established training evaluation process follows a standardized workflow applicable across all management development programs. The evaluation sequence unfolds as follows:

A. Pre-Training Phase

During this phase, the post-training evaluation findings from the previous batch are reviewed by the L&TD team prior to program delivery in order to identify important feedback, areas for improvement, and instructional improvements. With a few minor adjustments to reflect particular module titles and instructor names, the assessment tool is the same for all programs. Nevertheless, there is no pre-training evaluation of participants' baseline capabilities, knowledge, or abilities.

B. Training Delivery Phase

Management development programs typically span in several days, with multiple modules delivered by different instructors. During this period, no formal learning assessment takes place. Training methodology emphasizes classroom presentation, group discussion, and selective workshop activities, with instructors maintaining primary responsibility for content delivery while L&TD staff manage logistics and participant support.

C. Post-Training Evaluation

Upon completion of each module (or in some cases, at program conclusion), participants receive evaluation forms via Google Forms links distributed. The digital nature of the instrument enables convenient completion via smartphones or computers, with responses automatically captured in spreadsheet format. Participant completion occurs immediately post-session, typically within the training venue or shortly thereafter.

D. Data Processing and Reporting Phase

Raw evaluation data, captured in Microsoft Excel format, undergoes processing by L&TD staff (including the researcher during the internship period). This processing involves:

1) Calculating average scores for each evaluation question by instructor/module
2) Generating summary statistics (mean, mode) for each aspect (instructor, material, time)
3) Creating comparison charts showing relative performance across modules
4) Compiling qualitative comments into categorized lists
5) Producing finalized evaluation reports

E. Report Distribution and Follow-Up

Completed evaluation reports reach multiple stakeholders, including training instructors (providing feedback on delivery effectiveness), L&TD leadership (informing programmatic decisions), and occasionally broader management audiences. However, no systematic process exists for:

1) Communicating results back to participants
2) Conducting follow-up assessment of learning application
3) Linking evaluation findings to curriculum modification decisions
4) Tracking performance metrics of trained versus untrained managers

3.2.3 Current Evaluation Instrument Structure

The standardized evaluation instrument employed across all three programs comprises eleven closed-ended questions and one open-ended question, organized across three conceptual domains:

A. Instructor Aspect (4 questions)

This domain assesses participants' perceptions of instructor quality:

1) Instructor has a good knowledge about the course material.
2) Instructor has a good ability to present the material.
3) Instructor always comes on time.
4) Instructor used the right method when he gives the material.

Each question employs a 5-point Likert scale:1=Poor, 2=Fair, 3=Moderate, 4=Good, 5=Excellent.

B. Material Aspect (4 questions)

This domain evaluates participants' reactions to training content:

1) Theory in training material is very complete.
2) Information giving in this training gives a great help in my job.
3) The training content is very up to date.
4) There are enough practices in this training.

Questions similarly utilize the 5-point scale from Poor to Excellent.

C. Time Aspect (2 questions)

This domain captures perceptions of training duration:

1) The course time is enough.
2) Time devoted to each session is fully & effectively.

These questions employ a 3-point scale: 1=Too Short, 2=As Much As Needed, 3=Too Long.

D. Open-Ended Feedback (1 question)

The instrument concludes with a free-response prompt: "Give us suggestion for our improvement." This open-ended question generates qualitative data revealing participants' priorities, concerns, and unmet needs.

3.2.4 Positioning Within Kirkpatrick-Phillips Framework

A concentration at Level 1 (Reaction) is found when GMF's existing procedures are examined using the Kirkpatrick-Phillips evaluation framework. Only participant satisfaction and emotive responses to training, such as perceptions of teacher quality, perceived content relevance, and time adequacy, are measured by the current instrument. These statistics only cover the base level of Kirkpatrick's hierarchy, despite offering insightful commentary on participant involvement and training delivery. Critically absent from current practices are the following:

A. Level 2 (Learning) Assessment

GMF conducts no systematic measurement of knowledge or skill acquisition. No pre-tests establish baseline competency, no post-tests verify

learning gains, and no demonstrations confirm skill development. Consequently, the organization cannot determine whether high satisfaction ratings correspond to actual learning, cannot identify participants requiring additional support, and cannot validate that training objectives were achieved.

B. Level 3 (Behavior) Evaluation

No follow-up assessment occurs to determine whether participants apply learned concepts in their work contexts. Evaluation concludes at program completion, with no mechanism to track behavioral change, identify application barriers, or provide post-training reinforcement. L&TD possesses no data on whether managers implement new practices, whether leadership behaviors change, or whether trained competencies transfer to workplace performance.

C. Level 4 (Results) Tracking

Training participation remains unconnected to organizational performance metrics. No system exists to compare business outcomes (productivity, quality, safety, customer satisfaction) between departments led by trained versus untrained managers, nor to track performance trends following training participation. Without results-level data, GMF cannot demonstrate training's contribution to strategic objectives or justify development investments through performance evidence.

D. Level 5 (ROI) Calculation

Given the absence of learning, behavior, and results data, return on investment analysis proves impossible. L&TD cannot quantify training benefits in monetary terms, cannot compare program costs against measurable returns, and cannot provide financial justification for leadership development investments to senior management.

The gap between current procedures and thorough review is a strategic challenge as well as a methodological one. Although GMF makes significant investments in leadership development, including funding for facilities, materials, participant time, and instructor fees, the organization does not have proof that these expenditures result in improved organizational performance, competency

development, or behavioral application. While the evaluation system gives information on participant satisfaction, it does not provide information regarding the effectiveness of the training, learning transfer, or business impact.

3.3 Thematic Analysis of Qualitative Feedback

This section presents a thematic analysis of 584 open-ended comments across the three programs, examining participants' priorities, concerns, and implicit evaluation needs that satisfaction scores alone cannot capture (Bates, 2004; Kirkpatrick & Kirkpatrick, 2016). Open-ended responses yield interpretive depth beyond what rating scales can measure, as high scores frequently mask substantive developmental needs invisible to quantitative instruments (Twitchell et al., 2000; Creswell & Creswell, 2018). Following Braun and Clarke's (2006) six-phase framework, comprising familiarization with the data, generating initial codes, constructing themes, reviewing and refining themes, defining and naming themes, and producing the final report, recurring patterns are identified and their implications for evaluation framework development are examined.

3.3.1 Analytical Approach

The analysis employed systematic coding procedures to identify themes within participant suggestions. Each comment was independently reviewed, with relevant keywords and phrases coded according to predefined thematic categories derived from both a priori theoretical constructs (e.g., transfer-related comments) and emergent patterns in the data itself. Thematic categories included the following:

a. **Time/Duration concerns** - Comments about training length, pace, or scheduling
b. **Need Assessment/Pre-Post Test** - Explicit requests for learning measurement
c. **Practice/Hands-on activities** - Requests for application opportunities
d. **Video/Multimedia requests** - Suggestions for enhanced instructional media
e. **Material Distribution** - Comments about access to training resources
f. **Too Technical/Complex** - Perceptions of content difficulty or abstraction

g. **Need More Examples** - Requests for practical, real-world illustrations
h. **Continues/Follow-up** - Suggestions for post-training support or updates
i. **Facilities** - Comments about physical/logistical arrangements
j. **Positive feedback** - General commendations without specific suggestions

Frequency analysis quantified theme prevalence across programs, while exemplar selection identified representative comments illustrating each pattern. To protect participant confidentiality, each respondent whose comment is cited in this section is assigned an anonymous code (e.g., R1, R2, R3, and so forth), assigned sequentially across all three programs in the order in which quotes appear in the analysis, allowing readers to distinguish between different respondents without disclosing personal or positional information. All cited responses were drawn directly from the raw evaluation data collected during the internship period and cross-referenced against original module-level worksheets to ensure authenticity.

3.3.2 Operational Leadership Program (OLP): Qualitative Patterns

A. Response Characteristics

OLP generated 135 substantive comments from 143 total responses (94.4% response rate), indicating high participant engagement with the open-ended feedback opportunity. The elevated response rate suggests that participants perceived the training as meaningful and meriting thoughtful reflection.

B. Theme Frequency Distribution

1) **Positive feedback** (51 mentions, 37.8% of comments) - Approximately one-third of comments expressed general satisfaction without specific improvement suggestions: *"Materinya bagus, semoga semakin banyak orang yang berkesempatan mendengarkan"* (R1) ("The material was good, hopefully more people get the opportunity to listen"); *"Excellent sharing session by Pak Andi. Would love to see more sharing session like this in other internal forums."* (R2)
2) **Material distribution** (33 mentions, 24.4%) - Participants frequently requested advance access to training materials: *"Materi yang padat dan penuh dengan informasi vital"* (R3) ("Material is dense and full of vital

information"); requests for softcopy materials to facilitate review and deeper engagement.

3) **Practice/Hands-on** (24 mentions, 17.8%) - Substantial proportion requested more application opportunities: *"Kurangnya korelasi antara materi dan workshop"* (R4) ("Lacking correlation between material and workshop"); *"Penambahan aktivitas praktek, seperti utk negosiasi"* (R5) ("Addition of practice activities, such as for negotiation")
4) **Time/Duration** (21 mentions, 15.6%) - Participants expressed mixed perceptions: some requested more time for specific activities, others suggested streamlining: *"Materi sangat bagus, aktual informasi terkini. Forum diskusi perlu waktu lebih"* (R6) ("Material is very good, current information. Discussion forum needs more time").
5) **Need More Examples** (11 mentions, 8.1%) - Requests for concrete illustrations: *"Great sharing, banyak contoh experience leadership aplikatif"* (R7) ("Great sharing, many practical leadership experience examples").
6) **Need Assessment/Pre-Post Test** (9 mentions, 6.7%) - **Critical finding:** Explicit participant requested for learning measurement: *"Apabila ada pre asses dan post asses di setiap materi akan lebih bisa memapping knowledge participant before after"* (R8) ("If there are pre-assessment and post-assessment for each material, it would better map participants' knowledge before and after"). This unprompted feedback reveals participant awareness of evaluation gaps and desire for more rigorous assessment.
7) **Continues/Follow-up** (6 mentions, 4.4%) - Requests for ongoing engagement: *"Perlu continues training"* (R9) ("Need continuous training"); *"Update perkembangan perusahaan setelah training"* (R10) ("Updates on company developments after training").

C. Interpretive Synthesis

The OLP qualitative data revealed several significant patterns. First, despite high quantitative satisfaction ratings, participants articulated specific improvement areas, suggesting that Level 1 data alone inadequately captured their full

perspective. Second, the explicit requested for pre/post assessment indicated sophisticated participant understanding that satisfaction differed from learning, they desired measurement of knowledge acquisition. Third, practice/hands-on themes suggested that while participants appreciated content quality, they sought more opportunities to apply concepts, foreshadowing transfer challenges if behavioral application received insufficient emphasis. Fourth, follow-up requests intimate awareness that one-time training events required reinforcement for sustained impact.

3.3.3 MRO Management Program: Qualitative Patterns

A. Response Characteristics

MRO Management produced 199 comments from 220 responses (90.5% response rate), maintaining high engagement though slightly lower than OLP's rate.

B. Theme Frequency Distribution

1) **Positive feedback** (111 mentions, 55.8% of comments) - The majority of comments (over half) expressed satisfaction without specific suggestions: *"Excellent"* (R11); *"Good"* (R12); *"Sangat menguasai dan cukup baik dalam menyampaikan"* (R13) ("Very knowledgeable and quite good in delivering the material"). This high proportion of generic positive feedback may reflect either genuine contentment or limited critical engagement with evaluation questions.
2) **Material distribution** (11 mentions, 5.5%) - Fewer requests than OLP, possibly reflecting different participant expectations or program structure.
3) **Time/Duration** (9 mentions, 4.5%) - Some participants perceived time constraints: *"Banyaknya materi-materi yang diberikan kurang pas dengan alokasi waktu"* (R14) ("The amount of material given does not match the time allocation").
4) **Video/Multimedia** (7 mentions, 3.5%) - Recurring requests for enhanced visual content: *"Ditambahkan beberapa video related materi"* (R15) ("Add some videos related to the material"); *"Kasih video biar ngga ngantuk"* (R16) ("Give videos so we do not get sleepy").

5) **Need More Examples** (4 mentions, 2.0%) - Requests for practical illustrations, though less frequent than OLP.
6) **Need Assessment/Pre-Post Test** (2 mentions, 1.0%) - Minimal explicit evaluation requests, contrasting with OLP's more frequent mentions.

C. Interpretive Synthesis

MRO Management's qualitative pattern differed markedly from OLP, with substantially higher proportion of generic positive feedback and lower frequency of specific improvement suggestions. This pattern may reflect several factors: (1) the functional/technical nature of content versus OLP's strategic/leadership focus may generate less reflective feedback; (2) the diverse participant backgrounds (spanning all directorates and business units) may reduce shared concerns; (3) the nine-module structure, with participants evaluating multiple discrete instructors, may fragment attention and reduce systematic critique. The lower frequency of assessment-related requests did not necessarily indicate satisfaction with evaluation practices but may reflect different participant priorities when encountering operational/technical content.

3.3.4 MRO Finance Program: Qualitative Patterns

A. Response Characteristics

MRO Finance yielded 250 comments from 275 responses (90.9% response rate), maintaining high engagement.

B. Theme Frequency Distribution

1) **Positive feedback** (139 mentions, 55.6% of comments) - Similar to MRO Management, over half expressed general satisfaction: *"Good"* (R17); *"Excellent"* (R18); *"Sudah bagus"* (R19) ("Already good").
2) **Material distribution** (25 mentions, 10.0%) - Moderate frequency, with participants requesting advance access to materials for pre-reading.
3) **Too Technical/Complex** (10 mentions, 4.0%) - **Critical pattern unique to Finance**: Multiple comments identified comprehension challenges: *"Penyampaian masih terlalu cepat untuk peserta yang awam"* (R20) ("Delivery is still too fast for participants who are novices"); *"Penjelasan*

definisi seharusnya bisa lebih jelas lagi agar mudah dimengerti oleh background non akuntansi" (R21) ("Definition explanations should be clearer so they can be understood by non-accounting backgrounds"); *"Banyak istilah baru untuk orang yang bukan dari bidang akuntansi, sehingga membingungkan"* (R22) ("Many new terms for people not from the accounting field, so it is confusing").

4) **Time/Duration** (10 mentions, 4.0%) - Balance between requests for more time and perceptions of sufficient duration.
5) **Need More Examples** (6 mentions, 2.4%) - Requests for concrete financial scenarios relevant to MRO operations.

C. Interpretive Synthesis

The Finance program's distinctive "Too Technical/Complex" theme reveals a critical participant characteristic often invisible in quantitative satisfaction ratings: baseline knowledge variation. Participants with engineering, operations, or technical backgrounds encountered financial terminology and concepts as foreign language, creating comprehension barriers despite instructor quality. These comments implicitly advocate for pre-assessment to identify baseline knowledge gaps and enable differentiated instruction or prerequisite materials. The pattern suggests that undifferentiated training delivery may inadequately serve audiences with heterogeneous educational backgrounds, a challenge that learning assessment (Level 2) would help diagnose and address.

3.3.5 Cross-Program Thematic Synthesis

A. Universal Themes

Several patterns transcended individual programs:

1) **Practice/Application Gap** - Across all programs (most pronounced in OLP with 24 mentions), participants sought more hands-on opportunities. This pattern suggests that despite high satisfaction with content presentation, participants recognized the gap between knowing concepts and applying them—the transfer challenge documented in literature. The consistent requests for practice opportunities intimate participant awareness that

leadership and business competencies require behavioral practice, not merely cognitive understanding.

2) **Material Distribution Patterns** - The recurring request for advance materials (33 mentions in OLP, 11 in Management, 25 in Finance) reflects adult learning principles wherein professionals prefer preparatory reading to maximize face-to-face engagement. This pattern also suggests that participants desire resources enabling post-training reference and reinforcement.
3) **Time Paradox** - Comments across programs revealed the paradoxical perception that training simultaneously felt lengthy yet insufficient for deep engagement with certain topics. This pattern reflects the challenge of balancing breadth (covering comprehensive content) versus depth (allowing extended exploration and practice).
4) **Assessment Awareness** - While most prevalent in OLP (9 mentions), explicit requests for pre/post assessment appeared in all programs. These comments reveal sophisticated participant understanding that their satisfaction does not equate to verified learning, they desire objective measurement confirming knowledge acquisition.

B. Program-Specific Patterns

Each program also displayed distinctive thematic emphases:

1) **OLP's higher critical engagement**: Longer comments, more specific suggestions, greater requests for assessment, potentially reflecting the senior management audience's experience and expectations for rigorous development.
2) **Management's brevity**: Shorter, more perfunctory comments, possibly reflecting the functional/technical content's nature or the nine-module structure creating evaluation fatigue.
3) **Finance's comprehension challenges**: Unique emphasis on content complexity and terminology difficulties, revealing the challenge of cross-functional training for audiences with heterogeneous backgrounds.

C. Implicit Evaluation Needs

The thematic analysis reveals evaluation needs that participants articulated indirectly:

1) **Baseline Assessment Need** - The Finance program's comprehension complaints implicitly request pre-assessment to identify knowledge gaps and enable appropriate instructional adaptation.
2) **Learning Verification Need** - Explicit requests for pre/post testing demonstrate participant recognition that satisfaction differs from learning and desire for objective competency verification.
3) **Application Support Need** - Practice requests and follow-up suggestions indicate awareness that training transfer requires ongoing support beyond one-time events.
4) **Relevance Validation Need** - Requests for more examples and real-world applications suggest that participants seek concrete evidence of content's practical utility for their work contexts.

These implicit needs, when combined with the documented absence of Levels 2-5 evaluation (Section 3.2), establish the empirical foundation for comprehensive framework development undertaken in Chapter 5.

3.4 Descriptive Statistical Analysis of Level 1 Evaluation Data

This section presents descriptive statistical findings from 638 individual evaluation responses collected across three management development programs during 2024-2025: Operational Leadership Program (143 responses), MRO Management (220 responses), and MRO Finance (275 responses). The analysis examines patterns in participant ratings, variations across programs and modules, and the overall characteristics of GMF's Level 1 evaluation data. Descriptive statistics are used here to organize, summarize, and present data in a meaningful and interpretable form, allowing researchers to identify patterns and tendencies within a dataset without extending conclusions beyond the observed sample (Creswell & Creswell, 2018). This includes the use of statistical measures such as

mean, standard deviation, and mode to systematically describe and interpret numerical data collected from the evaluated sample (Sugiyono, 2019).

3.4.1 Operational Leadership Program (OLP)

A. Program Overview

The Operational Leadership Program represents GMF's flagship leadership development initiative, targeting 30 participants comprising 11 Senior Managers and 19 Talent-designated high-potential employees. Delivered over three days, the program encompassed five modules: Economic Outlook & MRO Global Outlook, Operational & Business Excellence, Innovation & Intrapreneurship (including workshop session), Leading at Higher Level, and Business Negotiation. The curriculum combined strategic perspective-building, operational excellence concepts, innovation frameworks, and leadership principles relevant to senior management responsibilities in the MRO environment.

B. Response Characteristics

The evaluation yielded 143 module-level responses (participants evaluated each instructor/module separately), representing a response completion rate exceeding 95%. The high response rate reflects L&TD's systematic follow-up procedures and participants' engagement with the evaluation process.

a. Instructor Aspect Performance

Across all OLP modules, instructor-related ratings demonstrated consistently high scores:

1) Knowledge of course material: M = 4.92, SD = 0.27, Mode = 5 (Range: 4-5)
2) Ability to present material: M = 4.88, SD = 0.35, Mode = 5 (Range: 3-5)
3) Punctuality: M = 4.95, SD = 0.22, Mode = 5 (Range: 4-5)
4) Appropriateness of teaching methods: M = 4.90, SD = 0.32, Mode = 5 (Range: 3-5)

Overall instructor aspect average: 4.91/5.00

The statistical patterns reveal several noteworthy characteristics. First, mean scores clustered tightly near the scale maximum, with overall instructor ratings

averaging 4.91 on a 5-point scale. Second, standard deviations remained small (0.22-0.35), indicating minimal variability in participant perceptions and suggesting broad consensus regarding instructor quality. Third, the mode value of 5 (Excellent) dominated across all instructor questions, with the majority of responses rating instructors at the highest level. Fourth, the restricted range with minimum values of 3-4 and maximum consistently at 5 indicates ceiling effects, wherein the instrument's limited scale constrains measurement of potential variation in instructor performance.

b. Material Aspect Performance

Content-related ratings similarly displayed high but slightly more varied scores:

1) Theory completeness: M = 4.83, SD = 0.43, Mode = 5 (Range: 3-5)

2) Job relevance of information: M = 4.85, SD = 0.43, Mode = 5 (Range: 3-5)

3) Content currency: M = 4.83, SD = 0.43, Mode = 5 (Range: 3-5)

4) Adequacy of practice opportunities: M = 4.71, SD = 0.55, Mode = 5 (Range: 3-5)

Overall material aspect average: 4.80/5.00

Material ratings, while remaining highly positive, exhibited marginally greater variability than instructor ratings (SD = 0.43-0.55 versus 0.22-0.35). Most notably, the question regarding practice opportunities ("There are enough practices in this training") received the lowest average score (4.71) and highest standard deviation (0.55), suggesting this dimension generated more diverse participant reactions. This pattern intimates that while participants perceived instructors highly favorably, some expressed reservation regarding the sufficiency of hands-on application opportunities, a finding corroborated by qualitative feedback examined in Section 3.3.

c. Time Aspect Performance

Duration-related ratings utilized 3-point scale (1 = Too Short, 2 = As Much As Needed, 3 = Too Long):

1) Course time sufficiency: M = 2.67, SD = 0.53, Mode = 3 (Range: 1-3)

2) Session time effectiveness: M = 2.78, SD = 0.45, Mode = 3 (Range: 1-3)

Overall time aspect average: 2.72/3.00

The time dimension revealed more nuanced patterns than instructor and material aspects. While the mode remained at 3 ("Too Long"), mean scores of 2.67-2.78 suggested mixed perceptions, with substantial minorities rating time as inadequate (score = 1) or appropriate (score = 2). The broader standard deviations (0.45-0.53) relative to the 3-point scale indicated considerable dispersion in time perceptions. These quantitative patterns align with qualitative feedback themes (discussed in Section 3.4) wherein participants simultaneously requested more time for certain activities while perceiving overall training as occasionally lengthy.

3.4.2 MRO Management Program

A. Program Overview

The MRO Management program provides cross-functional business knowledge to 24 participants (18 Managers, 6 Talent) across various directorates. Delivered over three days (April 21-23, 2025), the curriculum encompassed nine modules spanning business landscape, marketing, supply chain, line/base/engine/component maintenance operations, quality awareness, and ERP systems. The program aims to broaden participants' understanding of GMF's integrated MRO operations beyond their specialized functional areas.

B. Response Characteristics

The evaluation generated 220 module-level responses, representing approximately 90% completion rate across participants and modules.

a. Instructor Aspect Performance

1) Knowledge of course material: M = 4.88, SD = 0.32, Mode = 5 (Range: 4-5)
2) Ability to present material: M = 4.84, SD = 0.42, Mode = 5 (Range: 3-5)
3) Punctuality: M = 4.92, SD = 0.29, Mode = 5 (Range: 3-5)
4) Appropriateness of teaching methods: M = 4.82, SD = 0.42, Mode = 5 (Range: 3-5)

Overall instructor aspect average: 4.86/5.00

Instructor ratings for MRO Management closely paralleled OLP patterns, with mean scores ranging 4.82-4.92 and overall average of 4.86. The consistency across programs suggests systematic instructor quality rather than program-specific variation. Slightly larger standard deviations (0.29-0.42) compared to OLP may reflect the greater number of instructors (nine versus five) and resulting increased variability in delivery styles and participant perceptions.

b. Material Aspect Performance

1) Theory completeness: M = 4.77, SD = 0.46, Mode = 5 (Range: 3-5)
2) Job relevance of information: M = 4.83, SD = 0.39, Mode = 5 (Range: 3-5)
3) Content currency: M = 4.78, SD = 0.52, Mode = 5 (Range: 1-5)
4) Adequacy of practice opportunities: M = 4.73, SD = 0.56, Mode = 5 (Range: 3-5)

Overall material aspect average: 4.78/5.00

Material ratings exhibited patterns comparable to OLP, with practice opportunities receiving lowest average (4.73) and highest variability (SD = 0.56). Notably, content currency displayed the widest range (1-5), including one or more responses at the scale's minimum, potentially indicating isolated but strongly negative perceptions of certain modules' timeliness.

c. Time Aspect Performance

1) Course time sufficiency: M = 2.89, SD = 0.32, Mode = 3 (Range: 2-3)
2) Session time effectiveness: M = 2.90, SD = 0.31, Mode = 3 (Range: 2-3)

Overall time aspect average: 2.90/3.00

Time perceptions for MRO Management skewed more strongly toward "Too Long" (score = 3) than OLP, with means approaching 2.90 and narrower standard deviations (0.31-0.32). The restricted range (2-3, versus OLP's 1-3) indicated no participants rated time as insufficient, suggesting the nine-module curriculum may have felt comprehensive or potentially extensive to participants.

3.4.3 MRO Finance Program

A. Program Overview

The MRO Finance program targets 32 participants (19 Department Heads, 13 Talent) across all four directorates, providing financial literacy for non-finance managers. Conducted over three days (February 10-12, 2025), the curriculum included eight modules: Corporate Financial Overview, Understanding Financial Statements, Evaluating Financial Performance, Operating & Capital Budgeting, Working Capital Management, COPA & Cost Analysis, SO-PO Creating, and Risk Management.

B. Response Characteristics

The evaluation produced 275 module-level responses, representing approximately 91% completion rate.

a. Instructor Aspect Performance

1) Knowledge of course material: M = 4.65, SD = 0.51, Mode = 5 (Range: 2-5)
2) Ability to present material: M = 4.59, SD = 0.54, Mode = 5 (Range: 2-5)
3) Punctuality: M = 4.72, SD = 0.54, Mode = 5 (Range: 1-5)
4) Appropriateness of teaching methods: M = 4.55, SD = 0.57, Mode = 5 (Range: 2-5)

Overall instructor aspect average: 4.63/5.00

MRO Finance instructor ratings, while remaining positive, displayed notably lower averages (4.55-4.72) and greater variability (SD = 0.51-0.57) compared to both OLP and MRO Management. The expanded range (including minimum scores of 1-2) suggests some participants perceived certain instructors less favorably, potentially reflecting the challenge of communicating technical financial concepts to audiences with diverse educational and professional backgrounds.

b. Material Aspect Performance

1) Theory completeness: M = 4.54, SD = 0.59, Mode = 5 (Range: 2-5)
2) Job relevance of information: M = 4.58, SD = 0.56, Mode = 5 (Range: 2-5)

3) Content currency: M = 4.57, SD = 0.53, Mode = 5 (Range: 2-5)
4) Adequacy of practice opportunities: M = 4.45, SD = 0.66, Mode = 5 (Range: 2-5)

Overall material aspect average: 4.54/5.00

Material ratings for Finance similarly exhibited lower averages and wider dispersion than other programs. Practice opportunities received the lowest score (4.45) and highest variability (SD = 0.66), consistent with the pattern observed across all programs but more pronounced in Finance. These patterns may reflect the inherently abstract nature of financial concepts, which participants with non-accounting backgrounds found challenging to connect with concrete workplace applications.

c. Time Aspect Performance

1) Course time sufficiency: M = 2.73, SD = 0.47, Mode = 3 (Range: 1-3)
2) Session time effectiveness: M = 2.72, SD = 0.45, Mode = 3 (Range: 2-3)

Overall time aspect average: 2.72/3.00

Time perceptions for Finance resembled OLP more closely than MRO Management, with means around 2.72-2.73 and standard deviations indicating moderate variability. The presence of scores across all three scale points suggests more diverse perceptions regarding duration appropriateness compared to MRO Management's more uniform pattern.

3.4.4 Cross-Program Comparative Analysis

A. Instructor Quality Consistency

Across all three programs, instructor-related ratings remained high, with overall averages ranging 4.63-4.91 on the 5-point scale. This consistency suggests that GMF's L&TD unit successfully engages qualified, competent instructors regardless of program content area. However, the Finance program's lower and more variable ratings (4.63 average, SD = 0.51-0.57) compared to OLP (4.91 average, SD = 0.22-0.35) may indicate that technical financial content presents greater instructional challenges, requiring specialized pedagogical approaches for non-specialist audiences.

B. Material Perception Patterns

Material aspect ratings similarly demonstrated consistency across programs (overall averages: OLP = 4.80, Management = 4.78, Finance = 4.54), with Finance again displaying lower and more variable perceptions. Notably, across all programs, practice opportunities consistently received the lowest average ratings and highest standard deviations, suggesting systematic participant desire for more hands-on application activities regardless of content domain.

C. Time Duration Convergence

Time perceptions showed interesting variations, with Management participants perceiving longer relative duration (2.90) compared to OLP and Finance (both 2.72). This pattern likely reflects program structure: Management's nine modules across three days versus OLP's five modules and Finance's eight modules, potentially creating a more densely packed curriculum experience.

D. Statistical Limitations and Implications

The consistently high ratings across all dimensions—with means clustering between 4.5-5.0 on 5-point scales—suggest ceiling effects that limit the instrument's discriminative power. With modal responses consistently at the maximum value and restricted ranges excluding lower scale points, the current instrument may inadequately capture variation in training quality or participant experience. This compression toward the scale ceiling raises questions about whether differences among instructors, modules, or programs truly are minimal (as the data suggest) or whether the instrument lacks sensitivity to detect meaningful variation.

Moreover, the exclusive focus on participant satisfaction provides no validation of whether high ratings correspond to actual learning, behavioral change, or organizational impact—the central limitation that motivates comprehensive evaluation framework development.

3.5 Gap Analysis: Current Practice Versus Comprehensive Evaluation Framework

The gap analysis presented in this section applies the Kirkpatrick-Phillips five-level framework introduced in Section 1.2 as the evaluative benchmark. Detailed theoretical elaboration of this framework is provided in Chapter IV. The gap analysis reveals not merely methodological limitations but strategic vulnerabilities that constrain organizational learning, accountability, and evidence-based decision-making regarding leadership development investments.

3.5.1 Level 1 (Reaction) Evaluation: Strengths and Limitations

A. Current Strengths

GMF's Level 1 evaluation demonstrates several commendable practices:

1) **High response rates** (90-95%) indicate effective implementation procedures and participant engagement
2) **Systematic data collection** using digital instruments ensures efficient capture and processing
3) **Consistent application** across programs demonstrates standardized practices
4) **Timely execution** with immediate post-training administration captures fresh impressions
5) **Stakeholder reporting** provides instructors and L&TD leadership with feedback

These strengths establish a solid foundation for evaluation practice and demonstrate organizational commitment to systematic assessment.

B. Critical Limitations

However, the current Level 1 approach exhibits significant weaknesses when assessed against evaluation best practices:

1. **Narrow Construct Coverage.** The existing instrument focuses exclusively on satisfaction—how participants feel about training—while neglecting other reaction-level constructs documented in literature as predictive of transfer. Specifically, GMF's evaluation omits:

a) **Perceived utility/relevance**: While one question addresses material's job relevance, no items assess participants' confidence that they can apply learned content in their specific work contexts.
b) **Behavioral intention**: The instrument does not measure participants' commitment to implement new practices, a construct that research identifies as predictive of actual transfer (De Jong et al., 2025).
c) **Learning confidence**: No items assess participants' self-efficacy regarding their mastery of content, despite research documenting confidence as a transfer enabler (Blume et al., 2019).
d) **Environmental support perception**: The evaluation excludes questions about whether participants anticipate supervisor support, peer reinforcement, or resource availability for applying learned concepts—factors critical for transfer (Burke & Hutchins, 2007).

2. **Limited Discriminative Power.** The consistent high ratings (means 4.5-5.0) and ceiling effects suggest the instrument lacks sensitivity to detect meaningful variation in training quality or participant experience. This compression may stem from:
   a) **Scale limitations**: The 5-point scale with only positive anchors (Poor/Fair/Moderate/Good/Excellent) may encourage socially desirable responding.
   b) **Question phrasing**: Items framed positively ("Instructor has a good knowledge...") may bias responses upward.
   c) **Lack of behavioral specificity**: Abstract items about instructor "ability" or material "completeness" may resist differentiation.
3. **Timing Constraints.** Immediate post-training evaluation captures only initial reactions, not considered judgments developed after workplace application attempts. Research demonstrates that participant perceptions of training utility often shift substantially weeks after completion, as attempts to apply learning reveal practical challenges or unexpected relevance (Hirv-Biene et al., 2025).

4. **Qualitative Data Under-Utilization.** While GMF collects open-ended comments, the current analysis merely compiles these into lists rather than employing systematic thematic coding to identify patterns, prioritize concerns, or track recurring issues across cohorts. The rich qualitative data remain under-analyzed relative to their potential value.

3.5.2 Level 2 (Learning) Assessment: Complete Absence

A. The Gap

GMF conducts no systematic learning assessment. No pre-tests establish baseline knowledge, no post-tests verify acquisition, no skills demonstrations confirm capability development, and no attitudinal surveys measure perspective shifts. Consequently, the organization cannot answer fundamental questions:

1) Did participants actually learn intended content?
2) Which specific competencies did they master versus struggle with?
3) Do baseline knowledge levels vary substantially, suggesting need for differentiated instruction?
4) Can knowledge gains be quantified to demonstrate training value?
5) Which participants may require additional support or remediation?

B. Implications

This gap creates several organizational vulnerabilities:

1) **Assumption Without Verification.** High satisfaction ratings create implicit assumption that learning occurred, yet satisfaction poorly predicts actual knowledge acquisition (Bates, 2004). GMF assumes effectiveness without evidence.
2) **Inability to Diagnose Learning Failures.** When participants fail to apply training in workplace, L&TD cannot determine whether failure stems from inadequate learning, environmental barriers, or motivational factors. Without learning data, diagnostic capacity remains limited.
3) **Lost Opportunity for Participant Support.** Learning assessment data would enable L&TD to identify struggling participants and provide targeted coaching, additional resources, or remedial sessions. Current

practices allow participants to complete training with undetected knowledge gaps.

4) **Curriculum Improvement Limitations.** Without learning data revealing which content participants mastered versus struggled with, curriculum refinement decisions rely on satisfaction feedback and instructor intuition rather than empirical evidence of learning effectiveness.
5) **Stakeholder Accountability Weakness.** L&TD cannot demonstrate to senior management that training investments produced documented competency development. In an organizational context emphasizing accountability and performance improvement, inability to verify learning undermines training's credibility.

C. Participant Awareness

Notably, OLP participants explicitly requested pre/post assessment (*"apabila ada pre asses dan post asses di setiap materi akan lebih bisa memapping knowledge participant before after"*), demonstrating sophisticated awareness that satisfaction differs from learning and that objective measurement would provide value. This unsolicited feedback reveals that evaluation gaps exist not merely from academic perspective but align with participant expectations for rigorous professional development.

3.5.3 Level 3 (Behavior) Evaluation: No Transfer Assessment

A. The Gap

GMF's evaluation concludes at program completion, with no follow-up to determine whether participants apply learned concepts in their work. No mechanism exists to the following:

1) Assess whether managers implement new leadership practices
2) Identify organizational barriers hindering transfer
3) Provide post-training support and reinforcement
4) Create accountability for application
5) Track behavior change over time

B. Implications

The absence of behavior evaluation creates the most critical gap in GMF's system, as transfer represents training's ultimate purpose without transfer assessment as follow:

1) **Training's Core Purpose Remains Unmeasured.** Organizations invest in leadership development to change managerial behaviors and improve performance, not merely to inform or entertain. If trained leaders do not apply new practices, training investment yields no organizational return, regardless of how much participants enjoyed the experience or what they learned. GMF cannot determine whether this fundamental objective is achieved.
2) **Transfer Problem Remains Invisible.** Research consistently documents that majority of training fails to transfer to workplace application (Ardondi et al., 2025). Without assessment, GMF cannot know whether its programs experience similar transfer challenges or achieve higher application rates. The organization operates without visibility into training's most critical outcome.
3) **Opportunity for Improvement Lost.** Transfer assessment data would reveal common barriers—unsupportive supervisors, lack of resources, competing priorities, peer resistance—enabling organizational response. Current practices allow barriers to persist unidentified and unaddressed.
4) **No Accountability Mechanism.** Without follow-up assessment, participants face no accountability for applying learned concepts. Research demonstrates that anticipated assessment creates motivation for transfer (De Jong et al., 2025). GMF's evaluation system inadvertently signals that application is optional rather than expected.
5) **Individual Development Tracking Impossible.** L&TD cannot identify high-performing managers who exemplify learned behaviors, nor struggling managers requiring additional coaching. Individual development remains untracked beyond program attendance.

C. Participant Desires

Comments requesting "continues training" and "update *perkembangan perusahaan setelah training*" suggest participant awareness that one-time events require ongoing engagement for sustained impact. These unsolicited requests intimate desire for continued connection between L&TD and training alumni—a relationship that behavior evaluation would formalize.

3.5.4 Level 4 (Results) Tracking: Business Impact Unknown

A. The Gap

No system connects leadership training participation to organizational performance metrics. GMF does not do the following:

1) Track business outcomes in departments led by training graduates
2) Compare performance between trained and untrained managers
3) Assess whether training correlates with improvements in productivity, quality, safety, customer satisfaction, or other strategic metrics
4) Link development activities to organizational strategic objectives

B. Implications

The following are the implications without results-level data:

1) **Strategic Value Undemonstrated.** GMF's 2024 strategic priorities emphasize profitability enhancement, operational excellence, and competitive positioning (GMF Integrated Report, 2024). Leadership development presumably supports these objectives, yet no evidence demonstrates this contribution. L&TD cannot show senior management that training investments advance strategic goals.
2) **Resource Allocation Lacks Evidence Base.** Decisions about which programs to continue, expand, modify, or discontinue occur without performance data. Resource allocation relies on satisfaction scores and anecdotal feedback rather than business impact evidence.
3) **Competitive Disadvantage.** In MRO industry characterized by safety imperatives, operational precision, and customer satisfaction, leadership quality directly influences organizational performance (Aviation Pros,

2016). GMF's inability to measure leadership development's performance impact represents strategic vulnerability—competitors with more rigorous evaluation may optimize their programs more effectively.

4) **Lost Opportunity for Case-Building.** Results data would enable L&TD to construct compelling business cases for training investments, demonstrating tangible returns. Without such data, training remains perceived as expense rather than strategic investment.

3.5.5 Level 5 (ROI) Calculation: Financial Justification Impossible

A. The Gap

Given absence of learning, behavior, and results data, return on investment calculation proves unfeasible. GMF cannot do the following:

1) Quantify training benefits in monetary terms
2) Compare program costs against measurable returns
3) Calculate ROI ratios for executive decision-making
4) Provide financial justification demonstrating training's profitability

B. Implications

The following are the implications without ROI capability:

1) **Executive Language Missing.** Senior management makes strategic decisions in financial terms—budgets, costs, returns, profitability. L&TD cannot communicate training value in this language, potentially marginalizing development in resource allocation discussions.
2) **Comparative Assessment Impossible.** Without ROI data, GMF cannot compare training investments against alternative uses of funds or benchmark against industry standards for development spending efficiency.
3) **Vulnerability in Resource Competition.** During budget constraints, initiatives lacking financial justification face disproportionate pressure. L&TD's inability to demonstrate ROI creates vulnerability when competing for resources.

However, it warrants emphasis that ROI calculation represents an advanced evaluation practice appropriate for selective, high-investment programs rather than

universal application (Phillips & Phillips, 2016). The more fundamental gap involves missing learning, behavior, and results data—without which ROI calculation remains impossible regardless of analytical sophistication.

3.5.6 Systemic Consequences of Evaluation Gaps

The documented gaps across Kirkpatrick-Phillips levels create interconnected organizational consequences as follow:

A. Limited Organizational Learning

Comprehensive evaluation enables systematic program improvement through empirical feedback. Without learning, behavior, and results data, program refinement decisions rely on incomplete information (satisfaction scores and anecdotal impressions) rather than evidence revealing which elements work, which need modification, and where targeted improvements would yield greatest benefit.

B. Accountability Vacuum

Rigorous evaluation creates accountability at multiple levels: participants accountable for learning and application, instructors accountable for teaching effectiveness, L&TD accountable for program quality, and leadership development accountable for organizational contribution. Current practices limit accountability to satisfaction delivery rather than competency development, behavioral change, or performance improvement.

C. Evidence-Based Decision-Making Constraints

Decisions about program continuation, expansion, modification, resource allocation, and strategic priorities occur without empirical foundation. This limitation affects not only training operations but broader organizational development strategy.

D. Missed Opportunity for Leadership Excellence

The MRO industry faces workforce challenges requiring sophisticated leadership development (Aircraft-parts, 2025). GMF's substantial investment in leadership programs demonstrates recognition of this imperative. However, evaluation gaps prevent the organization from optimizing these investments, tracking development outcomes, or demonstrating return. A significant opportunity

remains unrealized because excellence in leadership development depends on excellence in evaluation.

3.6 Chapter Summary and Implications

This chapter has presented comprehensive findings regarding GMF's current training evaluation practices, revealing both established strengths and systematic gaps relative to evidence-based evaluation standards. The analysis encompassed 638 quantitative evaluation responses and 584 qualitative comments across three management development programs, yielding several critical conclusions.

A. Current State: Solid Foundation, Significant Gaps

GMF's L&TD unit has established systematic Level 1 evaluation practices characterized by high response rates, consistent procedures, and effective data management. Quantitative findings reveal uniformly high participant satisfaction across instructor quality (overall averages 4.63-4.91/5.00), material relevance (4.54-4.80/5.00), and reasonable time perceptions. These results attest to GMF's success in delivering well-received training experiences.

However, qualitative analysis reveals that beneath high satisfaction ratings lie substantive concerns and unmet needs. Participants explicitly request pre/post learning assessment, seek more hands-on practice opportunities, desire ongoing post-training engagement, and identify challenges with content complexity requiring baseline assessment. These patterns suggest that Level 1 data alone provides incomplete understanding of training effectiveness and participant needs.

B. Critical Evaluation Gaps

The gap analysis documents systematic absence of assessment at Kirkpatrick-Phillips Levels 2-5:

1) **No learning measurement** prevents verification of knowledge acquisition, diagnosis of learning difficulties, or demonstration of competency development

2) **No behavior evaluation** leaves transfer outcomes invisible, eliminates accountability for application, and prevents identification of transfer barriers
3) **No results tracking** means business impact remains unknown and strategic value undemonstrated
4) **No ROI calculation** precludes financial justification of training investments

These gaps create organizational vulnerabilities: inability to demonstrate training value, constraints on evidence-based decision-making, limited capacity for systematic improvement, and reduced accountability for development investments.

C. Implications for Framework Development

The findings establish clear empirical foundation for comprehensive evaluation framework development:

1) **Need for Enhanced Level 1 Assessment** - Current reaction evaluation requires expansion to include transfer-predicting constructs (behavioral intention, learning confidence, environmental support perception) rather than satisfaction alone.
2) **Imperative for Learning Measurement** - Participant requests, combined with learning assessment's fundamental importance, necessitate systematic Level 2 instruments including pre-tests, post-tests, and skill demonstrations aligned with specific learning objectives.
3) **Critical Importance of Behavior Evaluation** - Transfer assessment represents the most significant gap and highest priority, requiring 360-degree feedback or similar multi-source rating approaches assessing leadership behavior change 3-6 months post-training.
4) **Strategic Value of Results Tracking** - Linking training to business metrics would enable demonstration of organizational value and evidence-based decision-making, though methodological complexities require practical, implementable approaches.

5) **Selective ROI Application** - Financial return calculation, while valuable for high-investment initiatives, should focus on flagship programs (e.g., OLP) rather than universal implementation.

The subsequent chapters translate these findings into practical, evidence-based evaluation instruments that address identified gaps while aligning with GMF's organizational context and resource constraints. The framework is designed to provide L&TD with actionable tools that improve programs, demonstrate value, and support GMF's strategic leadership development objectives.

CHAPTER IV
LITERATURE REVIEW

4.1 Theoretical Foundation

This chapter establishes the theoretical foundation and conceptual framework guiding the development of a comprehensive post-training evaluation system for PT GMF Aero Asia's management development programs. The literature review grounds the research in established theoretical frameworks including instructional design models (ADDIE), training evaluation theory (Kirkpatrick-Phillips), transfer of learning, and adult learning principles (andragogy). It synthesizes empirical evidence validating these frameworks, defines key concepts essential to systematic training development and evaluation, and articulates relationships among these concepts within the aviation MRO leadership development context.

4.1.1 Primary Theory

This section presents the foundational theoretical frameworks undergirding the research: the ADDIE instructional design model positioning evaluation within the complete training lifecycle, the Kirkpatrick-Phillips evaluation framework providing systematic assessment across five hierarchical levels, transfer of learning theory explaining workplace application dynamics, and andragogical principles informing adult professional development. Together, these theories provide the conceptual architecture for developing comprehensive evaluation approaches appropriate to organizational leadership development contexts.

4.1.1.1 The ADDIE Model: Systematic Instructional Design Framework

The ADDIE model (Analysis, Design, Development, Implementation, Evaluation) represents the foundational framework for systematic instructional design, providing clear stages for creating effective training programs while accommodating iterative refinement throughout the design process (Kurt, 2017; Peck, 2025). Developed in 1975 by the Center for Educational Technology at Florida State University for the U.S. Army, ADDIE has become the dominant

instructional design paradigm, with over 100 contemporary models based on or derived from its core principles (Allen, 2006; Research.com, 2025). The ADDIE framework comprises five interconnected phases that form a complete training development cycle:

A. Analysis Phase

Establishes foundation for all subsequent activities through systematic examination of multiple dimensions (Peck, 2025). Organizational analysis examines how training aligns with strategic business objectives. Job-task analysis identifies specific competencies required for effective performance. Learner analysis profiles target audience characteristics, existing knowledge, and learning needs. Context analysis evaluates resources, constraints, and environmental factors. Critically, this phase includes Training Needs Analysis (TNA), a systematic process identifying gaps between current and desired employee capabilities (AIHR, 2024; Nunes et al., 2025). Research emphasizes that while analysis represents the most important phase, it paradoxically remains the most frequently neglected, with organizations often pressuring training departments to develop programs immediately while bypassing thorough needs assessment (Peck, 2025; Moore, 2023).

B. Design Phase

Translates identified needs into structured learning objectives, assessment strategies, instructional methods, and content outlines (InstructionalDesign.org, 2018). Learning objectives specify observable measurable outcomes. Assessment instruments align with objectives to verify achievement. Instructional strategies match content types and learner characteristics. Content sequencing facilitates progressive skill development. The design phase produces comprehensive blueprints guiding subsequent development activities (Papadopoulou et al., 2022).

C. Development Phase

Transforms blueprints into tangible learning resources (Ahmadi, 2020; Rafiola & Umar, 2024). This encompasses creating instructional content, designing learning activities and exercises, producing multimedia elements, developing

assessment instruments, and constructing participant materials. Contemporary development increasingly incorporates digital technologies alongside traditional classroom materials and includes pilot testing and iterative refinement based on feedback (Research.com, 2025).

D. Implementation Phase

Involves delivering training to target learners, requiring careful attention to instructor preparation, logistical arrangements (scheduling, facilities, technology), participant engagement strategies, and real-time monitoring of learning progress (Chee et al., 2023). Implementation varies substantially across delivery modalities. Effective implementation also includes formative evaluation, providing real-time feedback enabling immediate adjustments (Kurt, 2017).

E. Evaluation Phase

Assesses training effectiveness and identifies improvement opportunities (Kurt, 2017; InstructionalDesign.org, 2018). Evaluation operates at two levels. Formative evaluation occurs throughout the ADDIE process, providing continuous feedback that refines design and delivery. Summative evaluation occurs post-implementation, examining whether training achieved intended outcomes and generated organizational value. The evaluation phase answers critical questions: Does training resolve identified performance problems? Do participants acquire intended competencies? Do behavioral changes manifest in workplace application? What organizational results emerged? Importantly, the evaluation phase connects directly to the analysis phase, creating a feedback loop where evaluation findings inform subsequent training cycles (Educational Technology, 2018).

The ADDIE framework's enduring influence stems from its systematic yet flexible approach. However, research consistently documents that organizations implement ADDIE's first four phases more systematically than the evaluation phase. Time constraints, resource limitations, and lack of evaluation expertise lead many organizations to neglect comprehensive assessment, implementing only superficial satisfaction surveys that fail to provide actionable insights for

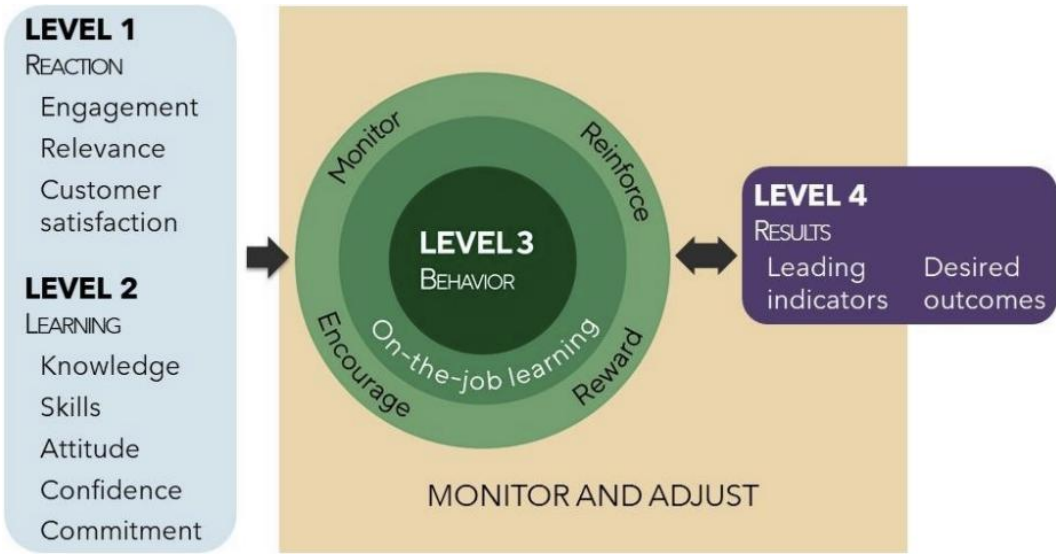
improvement or convincing evidence of value (Peck, 2025; Educational Technology, 2018).

4.1.1.2 The Kirkpatrick-Phillips Model: Training Evaluation Framework

While ADDIE positions evaluation within the broader instructional design process, the Kirkpatrick-Phillips framework provides specific theoretical architecture for understanding what comprehensive training evaluation should assess and how assessment can progress from simple to sophisticated levels.

A. Kirkpatrick's Four Levels

Donald Kirkpatrick introduced his four-level evaluation model in 1959, establishing a framework that has dominated training evaluation discourse for over six decades (Kirkpatrick & Kirkpatrick, 2016). The model conceptualizes evaluation as progressing through increasingly rigorous and valuable levels:



Source: Kirkpatrick Partners (2021).

Figure 4.1. Kirkpatrick New World Model

1) **Level 1 (Reaction)**

This level assesses participants' immediate responses to training, capturing satisfaction with instructors, materials, delivery methods, and facilities, along with perceptions of training's relevance and utility. Level 1 evaluation

typically employs surveys administered immediately post-training. While Level 1 assessment is nearly universal (86-100% of organizations implement reaction evaluation), critics note that satisfaction correlates weakly with learning or behavioral application (Bates, 2004; Twitchell et al., 2000). Participants may enjoy entertaining presentations without acquiring competencies, or conversely, challenging learning experiences may generate lower satisfaction despite producing superior outcomes.

2) **Level 2 (Learning)**

This level assesses knowledge and skill acquisition resulting from training. This level verifies that participants actually learned intended content through pre-tests and post-tests, skills demonstrations, case analyses, or knowledge assessments. Level 2 evaluation answers the fundamental question: Does learning occur? Without learning assessment, organizations cannot determine whether high satisfaction ratings reflect actual competency development or merely pleasant experiences. Despite learning assessment's importance, only 30-54% of organizations conduct systematic Level 2 evaluation (Twitchell et al., 2000).

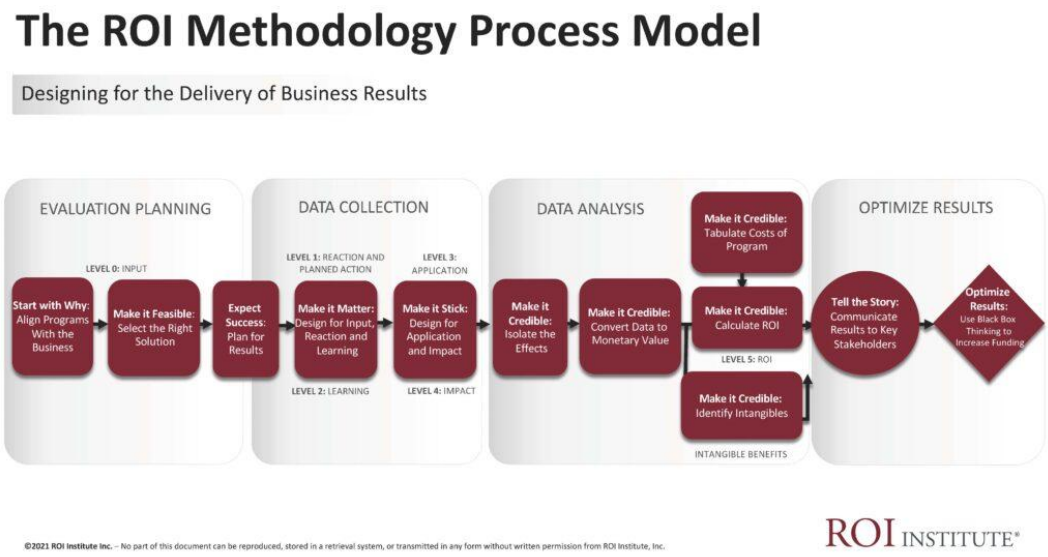
3) **Level 3 (Behavior)**

This level evaluates whether participants apply learned knowledge and skills in their workplace contexts. This level addresses training's ultimate purpose: changing on-the-job behavior and performance. Level 3 assessment typically occurs 3-6 months post-training through multi-source feedback (360-degree assessments), supervisor observations, self-reports, or performance monitoring. Behavior evaluation proves critical because research consistently documents that the majority of training fails to transfer to workplace application despite participants learning content (Ardondi et al., 2025). However, only 21-49% of organizations implement Level 3 evaluation, with transfer assessment remaining the most neglected yet arguably most important evaluation level (Fortenberry, 2025).

4) **Level 4 (Results)**

This level examines training's impact on organizational performance metrics such as productivity, quality, customer satisfaction, safety, turnover, or financial outcomes. This level demonstrates training's business value by linking participation to strategic objectives and operational improvements. Level 4 evaluation faces substantial methodological challenges (isolating training effects from other influences, time lags between training and results, comparison group difficulties), explaining why only 13-49% of organizations attempt results measurement (Twitchell et al., 2000; Fortenberry, 2025).

B. Phillips' ROI Methodology (Level 5)



Source: ROI Institute (2021).

Figure 2.2. The ROI Methodology Process Model

Jack and Patti Phillips extended Kirkpatrick's framework by adding fifth level: Return on Investment (ROI) calculation (Phillips & Phillips, 2016). Level 5 converts Level 4 results into monetary terms, compares training benefits to costs, and calculates ROI percentage using the formula:

$$\textbf{ROI\% = [(Benefits - Costs) / Costs] × 100}$$

The Phillips methodology provides systematic procedures for isolating training effects, converting data to monetary values, tabulating fully-loaded program costs, and calculating financial returns. However, ROI analysis requires substantial expertise, data access, and analytical resources, making it appropriate primarily for high-investment flagship programs rather than universal application (Bothell et al., 2002).

4.1.1.3 Transfer of Learning Theory: Understanding Workplace Application

Transfer of learning refers to the degree to which trainees effectively apply knowledge, skills, and attitudes gained in training to their job contexts (Baldwin & Ford, 1988). Transfer represents training's ultimate objective, yet research consistently documents that only 10-20% of training expenditures result in sustained performance improvement, with the majority of learning failing to transfer to workplace application (Ardondi et al., 2025; Blume et al., 2019). Baldwin and Ford's (1988) seminal transfer framework identifies three categories of factors influencing transfer as follow:

A. Training Design Factors

Include content validity (degree to which training content mirrors actual job requirements), principles of learning incorporated into instruction (practice opportunities, feedback, reinforcement), and sequencing that facilitates retention and application. Research demonstrates that training emphasizing behavioral practice, providing performance feedback, and enabling error-based learning produces superior transfer compared to passive information delivery (Blume et al., 2019).

B. Trainee Characteristics Factors

Encompass ability (cognitive capacity to learn content), personality traits (conscientiousness, openness to experience), and motivation (commitment to learning and applying new skills). Particularly critical is self-efficacy, defined as individuals' confidence in their capability to successfully perform learned behaviors. Higher self-efficacy correlates strongly with transfer attempts and persistence when encountering application challenges (Blume et al., 2019).

C. Work Environment Factors

Prove most influential for transfer yet remain most frequently neglected in training design. Supervisor support (encouragement to apply learning, provision of feedback, removal of barriers), peer support (reinforcement from colleagues, collaborative application opportunities), organizational climate (culture valuing continuous learning and improvement), and opportunity to use learned skills all significantly affect whether training translates to behavioral change (Burke & Hutchins, 2007). When organizational environments fail to support transfer through providing time, resources, encouragement, and consequences for application, even high-quality training with motivated learners fails to produce workplace impact.

Contemporary transfer research has established several key principles. De Jong et al. (2025) found that participants who set specific application goals immediately post-training demonstrated 47% higher transfer rates at 3-month follow-up, and transfer motivation measured through behavioral intentions correlated strongly with actual application behaviors ($r = 0.62$). Hirv-Biene et al. (2025) documented that environmental support factors explained 41% of variance in transfer success, substantially exceeding trainee characteristics (18%) or training design features (23%). Blume et al. (2019) meta-analysis found that training producing significant learning gains frequently failed to generate corresponding behavioral changes with the learning-to-behavior gap attributable primarily to environmental constraints rather than learning deficits.

4.1.1.4 Adult Learning Theory: Andragogical Foundations

Andragogy, the art and science of helping adults learn, recognizes that adult learners differ fundamentally from children in motivations, experiences, and learning preferences (Knowles et al., 2015; Taylor & Hamdy, 2013). Knowles et al. (2015) identified six core andragogical principles shaping how adults learn as follow:

A. Need to Know

Adults need to understand why learning specific content is necessary before engaging seriously with material. They invest energy proportionally to the perceived value and immediate applicability of learning.

B. Self-Concept

Adults possess self-directed learning orientation. They resist situations positioning them as passive recipients of instruction and prefer collaborative approaches respecting their autonomy and experience.

C. Experience

Adults bring accumulated life and work experience serving as rich resource for learning. Training that ignores or devalues this experience alienates learners and wastes valuable knowledge that could enhance collective learning.

D. Readiness to Learn

Adults become ready to learn when content addresses current life tasks and challenges. Training disconnected from immediate application needs generates resistance regardless of content quality.

E. Orientation to Learning

Adults are problem-centered rather than subject-centered in their learning approach. They learn most effectively when new knowledge helps solve real problems they currently face in their work or life.

F. Motivation

While adults respond to external motivators (promotions, salary increases), the most potent motivators are internal: job satisfaction, self-esteem, quality of life. Training appealing to intrinsic motivation generates deeper learning than training relying exclusively on external rewards.

Andragogical principles directly inform evaluation design. Evaluation instruments respecting adult learners treat them as collaborative partners in assessing development rather than passive subjects of testing. Assessment

emphasizes application to authentic work challenges rather than abstract knowledge recall. Feedback frames findings developmentally, supporting continuous professional growth rather than judgmentally categorizing performance (Podoshen, 2024).

4.1.1.5 Critical Perspectives on Kirkpatrick and Alternative Evaluation Models

While adult learning theory explains how managerial participants acquire and apply knowledge during training, organizations must also establish systematic mechanisms to evaluate whether these learning processes translate into behavioral and organizational outcomes. Therefore, several training evaluation frameworks have been developed in the human resource development literature to assess training effectiveness at multiple levels.

Several alternative training evaluation frameworks were considered prior to selecting the Kirkpatrick–Phillips model as the primary theoretical foundation for this study. The CIPP model developed by Stufflebeam (1971) evaluates programs through four dimensions: Context, Input, Process, and Product, and is widely used for formative program audits and curriculum evaluation. However, the model does not explicitly extend its product evaluation to post-training behavioral change or financial outcomes, limiting its suitability for assessing the impact of management development programs on workplace performance and organizational results. Brinkerhoff's (2003) Success Case Method provides rich qualitative insights by identifying highly successful participants and tracing their performance improvements to training interventions. While methodologically valuable for illustrating individual success narratives, its case-based approach is less suitable for systematic program-level evaluation across an entire training portfolio. Holton's (1996) HRD Evaluation Research and Measurement (HRDERM) model expands the Kirkpatrick framework by incorporating motivational variables and environmental transfer factors that influence learning outcomes. Despite its theoretical sophistication, the model is primarily oriented toward research measurement and is less easily translated into practical evaluation instruments for organizational implementation.

The Kirkpatrick–Phillips framework was therefore selected for three context-specific reasons. First, its hierarchical structure directly corresponds to the evaluation gaps identified in GMF's current practices, enabling systematic assessment across reaction, learning, behavior, results, and return on investment levels. Second, the ROI extension introduced by Phillips provides a structured methodology for evaluating financial outcomes, which aligns with the accountability expectations of GMF's BUMN governance environment (Phillips, 2012). Third, the framework has been widely adopted in corporate human resource development practice, offering extensive implementation guidance and established evaluation instruments that can be adapted efficiently to organizational settings (Kirkpatrick & Kirkpatrick, 2016). Although the model has been criticized for oversimplifying causal relationships between evaluation levels (Bates, 2004), it remains one of the most widely applied frameworks for structuring training evaluation systems in organizational contexts.

4.1.2 Empirical Studies

While the preceding section established theoretical frameworks, this section examines empirical research validating these theories and documenting their application across organizational contexts, emphasizing recent studies (2018-2025) employing rigorous methodologies.

A. Validation Studies of the Kirkpatrick-Phillips Framework

Ahmad et al. (2024) conducted structural equation modeling examining relationships among Kirkpatrick levels using data from 384 bank employees. The study found significant positive relationships among all four levels, with reaction positively influencing learning ($\beta = 0.43$, $p < 0.001$), learning influencing behavior ($\beta = 0.51$, $p < 0.001$), and behavior influencing results ($\beta = 0.38$, $p < 0.001$), validating the framework's hierarchical structure. Faisal-E-Alam et al. (2025) revealed that reaction alone explained only 18% of variance in organizational results, but when combined with learning and behavior assessments, explained variance increased to 67%, demonstrating comprehensive multi-level evaluation's superior understanding. Liu et al. (2025) documented significant improvements

across all four levels in standardized nursing training with sustained effects at 12-month follow-up when supported by organizational reinforcement systems.

B. Transfer of Learning: Empirical Evidence

Ardondi et al. (2025) systematic scoping review analyzing 49 studies revealed that while 86-100% of organizations assess reactions and 30-54% measure learning, transfer evaluation remains rare. De Jong et al. (2025) found that participants who set specific application goals post-training demonstrated 47% higher transfer rates at 3-month follow-up compared to control groups ($p < 0.01$), with transfer motivation correlating strongly with actual application ($r = 0.62$, $p < 0.001$). Hirv-Biene et al. (2025) documented that environmental support factors (supervisor encouragement, peer collaboration, organizational rewards) explained 41% of variance in transfer success, substantially exceeding trainee characteristics (18%) or training design (23%). Blume et al. (2019) meta-analysis found training producing significant learning gains ($d = 1.2$ effect sizes) frequently failed to generate corresponding behavioral changes ($d = 0.3$), with the learning-to-behavior gap attributable primarily to environmental constraints rather than learning deficits.

C. Leadership Development Evaluation: Challenges and Innovations

Fortenberry (2025) examining healthcare leadership programs documented that while 95% conducted Level 1 satisfaction assessment, only 38% implemented Level 2, 21% assessed Level 3, and merely 13% tracked Level 4. Harvard Business Review's (2025) survey of 1,847 organizations across 82 countries revealed that companies employing comprehensive multi-level evaluation reported 2.3 times higher leadership development effectiveness compared to satisfaction-only assessment, with 34% higher leadership bench strength and 28% better succession planning readiness.

D. Aviation MRO Training and Competency Development

Güneş et al. (2020) examining aviation maintenance technician training in Turkey identified gaps between training content and actual competency requirements, particularly in safety practices and quality standards. Aircraft-parts (2025) documented that the global MRO industry faces severe workforce

challenges, with 60% of global maintenance workforce approaching retirement, creating critical skills gap threatening operational capacity and safety standards.

4.1.3 Key Concepts

This section defines key concepts essential to understanding comprehensive training development and evaluation systems, spanning the complete training lifecycle from needs assessment through delivery methods to multi-level evaluation approaches.

A. Training Needs Analysis (TNA)

A systematic process for identifying gaps between employees' current capabilities and the knowledge, skills, and abilities required for optimal performance (AIHR, 2024; Nunes et al., 2025). TNA ensures training investments address genuine performance requirements rather than perceived or assumed needs. The classic TNA model examines training needs at three hierarchical levels: organizational level (training aligned with corporate strategy), job/task level (specific competencies for effective performance), and individual level (each employee's capabilities relative to requirements). Strategic TNA provides organizational benefits including alignment with business goals, early identification of skills gaps, data-driven resource allocation, and customized programs maximizing effectiveness.

B. 360-Degree Assessment

Multi-source feedback systematically gathering input from multiple sources surrounding an individual, typically including self-assessment, supervisor ratings, peer evaluations, and direct report perspectives (Church et al., 2019). When administered before training, 360-degree assessment establishes baseline competency levels and identifies specific development needs. When administered after training (typically 3-6 months post-program), it evaluates behavioral change by comparing post-training ratings to pre-training baseline using identical competency frameworks and rater groups.

C. Collaborative Learning and Gamification

Collaborative learning refers to educational approaches wherein participants work together in small groups to solve problems and complete assignments, creating more effective experiences through social interaction (Xperiencify, 2024). Research demonstrates collaborative learning generates superior outcomes, with collaboratively developed courses receiving twice the usefulness ratings and 67% of Gen Z employees favoring social training experiences (360Learning, 2020, 2023). Gamification integrates game design elements (points, badges, leaderboards, challenges) into training contexts. Systematic reviews demonstrate gamification significantly enhances knowledge retention, sharing, and job performance, generating 25-48% increases in employee motivation with 90% indicating productivity enhancement (Chief Learning Officer, 2024; Udeh & Castro, 2025).

D. E-Learning & On-the-Job Training

E-learning delivers instruction through digital platforms enabling flexible, self-paced learning with benefits including accessibility, cost-effectiveness, and scalability, though challenges include technological barriers and reduced social interaction (Coman et al., 2020). On-the-Job Training (OJT) teaches skills in actual work settings through hands-on practice, job rotation, mentoring, or apprenticeships, proving particularly effective for technical skills and contextual knowledge transfer (Jacobs & Park, 2009).

E. Learning Management Systems (LMS)

Software platforms facilitating digital course delivery, progress tracking, assessment administration, and performance reporting. Modern LMS integrate with other organizational systems providing comprehensive learning analytics and personalized learning pathways (LinkedIn, 2023).

4.1.4 Relationships Among Concepts or Variables

These concepts integrate into a comprehensive training development and evaluation system. The process begins with TNA identifying competency gaps aligned with organizational strategy. 360-degree pre-assessment provides detailed

baseline competency profiles for individuals. The design phase translates identified needs into learning objectives and assessment strategies. Development creates learning resources employing collaborative learning, gamification, e-learning, and OJT. Implementation delivers training through multiple modalities managed via LMS platforms.

Evaluation assesses effectiveness across five Kirkpatrick-Phillips levels: enhanced Level 1 capturing transfer readiness predictors, Level 2 verifying learning through pre/post assessment, Level 3 evaluating behavioral change via post-training 360-degree feedback, Level 4 linking training to organizational performance metrics, and Level 5 calculating ROI for flagship programs. Evaluation findings feedback to TNA, creating continuous improvement cycles where data reveals which needs were met, which competencies still gap, and which delivery methods worked.

Transfer of learning connects training to workplace performance. Training design incorporating Baldwin and Ford's principles combined with supportive work environments enables transfer from learning to application. Andragogical principles inform all phases: TNA respecting adults' experience, design emphasizing problem-centered content, development incorporating self-directed elements, implementation treating participants as collaborative partners, and evaluation providing developmental feedback.

This integrated system transforms training from discrete events into strategic organizational capability where evaluation generates knowledge enabling evidence-based refinement, transfer support structures bridge learning and application, and adult learning principles ensure developmental appropriateness for professional learners.

4.2 Conceptual Framework

Having established theoretical foundations, empirical evidence, key concepts, and their interrelationships, this section presents the conceptual framework guiding the development of GMF's comprehensive post-training evaluation system.

4.2.1 Definition of the Framework

The proposed conceptual framework represents a comprehensive multi-level training evaluation system designed specifically for GMF Aero Asia's management development programs (Operational Leadership Program, MRO Management, and MRO Finance). The framework serves multiple interconnected purposes: completing GMF's training development cycle by addressing systematic evaluation gaps; demonstrating training's organizational value through empirical evidence linking programs to strategic objectives; enabling evidence-based program improvement through data revealing curriculum strengths and weaknesses; and building organizational accountability at multiple levels (participants accountable for learning and application, instructors for teaching effectiveness, L&TD for program quality, leadership development for organizational contribution).

4.2.2 Conceptual Model or Approach

The conceptual framework employs an integrated approach combining the ADDIE instructional design model, the Kirkpatrick-Phillips evaluation hierarchy, transfer of learning theory, and andragogical principles within GMF's specific MRO industry and organizational context. The integration of the ADDIE instructional design model and the Kirkpatrick-Phillips evaluation framework is necessary because each model addresses different aspects of training systems. ADDIE structures the training lifecycle through the phases of Analysis, Design, Development, Implementation, and Evaluation, guiding how programs are planned and delivered (Branch, 2009). However, its Evaluation phase provides limited guidance on what outcomes should be measured or how evaluation should be conducted. In contrast, the Kirkpatrick-Phillips framework specifies what should be evaluated across five levels, from participant reaction to return on investment, but it does not indicate where these evaluations should occur within the training development process (Kirkpatrick & Kirkpatrick, 2016; Phillips, 2012).

Integrating the two models addresses these limitations by aligning evaluation levels with stages of the ADDIE cycle. Level 1 (Reaction) is conducted immediately after training during the Implementation phase, Level 2 (Learning)

through pre- and post-assessments within the same phase, and Levels 3 and 4 (Behavior and Results) through follow-up measurements that inform the Analysis phase of the next training cycle. This alignment ensures that evaluation is embedded within GMF's existing training process while creating a feedback loop in which evaluation results inform subsequent program design and improvement.

A. Integration with ADDIE Training Development Cycle

The framework positions evaluation as the culminating component of GMF's complete ADDIE-based training development process:

1) **Pre-Training Phase (Analyze, Design, Develop)**

   GMF conducts Training Needs Analysis identifying development requirements aligned with organizational strategy and individual competency gaps. The L&TD unit designs training programs addressing diagnosed needs. For selected programs, 360-degree pre-assessment establishes baseline leadership competency profiles.

2) **Training Phase (Implement)**

   GMF employs multiple delivery modalities: classroom instruction emphasizing collaborative learning and gamification elements engages participants actively; e-learning components provide flexible access to resources; OJT planning enables workplace application preparation. This multi-modal approach reflects evidence-based training design principles.

3) **Post-Training Phase (Evaluate)**

   The proposed framework implements systematic evaluation across five levels:

   a) **Level 1 (Enhanced Reaction Assessment)**

      Retains GMF's current satisfaction evaluation (instructor quality, material relevance, time adequacy) while adding transfer-predicting constructs: behavioral intention (commitment to apply concepts), learning confidence (self-efficacy regarding mastery), perceived utility (job relevance), and environmental support perception (anticipated barriers and enablers). This enhancement, requiring minimal additional participant time, provides richer insights about transfer readiness.

b) Level 2 (Learning Verification)

Introduces pre-test/post-test designs measuring knowledge acquisition aligned with program-specific learning objectives. OLP assessment emphasizes strategic thinking, operational excellence, innovation, and leadership competencies. MRO Management assessment evaluates cross-functional business knowledge acquisition. MRO Finance assessment verifies financial literacy development for non-finance managers. Learning data identify participants requiring additional support and curriculum areas needing refinement.

c) Level 3 (Behavior Evaluation)

Employs 360-degree feedback 3-6 months post-training, utilizing the same competency framework and rater groups as pre-training assessment where implemented. This approach enables rigorous pre-post comparison revealing training's impact on workplace leadership behaviors. Multi-source ratings (self, supervisor, peers, direct reports) provide comprehensive perspective on behavioral change.

d) Level 4 (Results Tracking)

Links training participation to organizational performance metrics through three strategies: pre-post departmental performance comparison tracking key metrics (operational efficiency, customer satisfaction, employee engagement, safety) before and after training; comparative analysis examining whether departments led by trained managers outperform similar units led by untrained managers; and stakeholder perception surveys gathering senior leaders' assessments of trained participants' organizational impact.

e) Level 5 (ROI Calculation)

Applies Phillips methodology selectively to flagship programs (OLP), converting Level 4 results into monetary terms, tabulating fully loaded program costs, and calculating return on investment. Conservative benefit estimation and transparent acknowledgment of assumptions

provide credible financial justification demonstrating leadership development's profitability.

B. Transfer-Focused Design

The framework incorporates transfer of learning theory at multiple points. Enhanced Level 1 assessment captures transfer-predicting variables enabling early identification of potential transfer barriers and proactive intervention. Level 3 behavioral evaluation directly assesses transfer outcomes through workplace application measurement. Environmental barrier diagnosis enables L&TD to engage with management to build support structures when Level 1 data reveal anticipated lack of supervisor support or resource limitations. Accountability mechanisms signal that behavioral change will be evaluated, creating expectation that application is expected rather than optional.

C. Andragogy-Informed Evaluation

Adult learning principles shape evaluation design throughout. Collaborative assessment treats participants as partners in development assessment rather than passive testing subjects. Relevance emphasis ensures learning assessments emphasize application to authentic work challenges rather than abstract knowledge recall. Self-direction support through 360-degree feedback revealing self-perception versus others-perception gaps enhances self-awareness, enabling participants to direct their own continued development.

4.2.3 Framework Contextualization

The conceptual framework applies theoretical principles within GMF's specific organizational context, adapting generic evaluation approaches to aviation MRO realities, Indonesian BUMN governance requirements, and L&TD's implementation capabilities.

A. Application to GMF's Management Development Programs

As GMF's flagship program targeting senior managers for strategic leadership roles, OLP receives most comprehensive evaluation implementing all five levels. The program's substantial investment, strategic importance, and relatively small cohort size justify rigorous multi-level evaluation including ROI

calculation. MRO Management, with participants spanning multiple directorates learning cross-functional business knowledge, receives Levels 1-3 evaluation. MRO Finance, equipping non-finance department heads with financial literacy, receives Levels 1-2 evaluation given the program's foundational knowledge-building objective. Participants' explicit requests for pre/post assessment makes learning evaluation particularly appropriate.

B. Alignment with BUMN Governance and Strategic Objectives

As a BUMN subsidiary, GMF operates within governance frameworks emphasizing accountability, transparency, and performance measurement. Multi-level evaluation provides empirical evidence that training investments produce measurable outcomes, fulfilling accountability expectations. Results metrics explicitly connect to GMF's 2024 strategic priorities: profitability enhancement (operational efficiency improvements, cost reductions), operational agility and transformation (innovation initiatives, process improvements), and customer-centricity (customer satisfaction increases). This alignment demonstrates training's contribution to organizational strategy rather than operating as disconnected activity. BUMN reform initiatives increasingly emphasize continuous improvement and evidence-based management, which systematic evaluation exemplifies, positioning GMF's training function as sophisticated organizational capability.

CHAPTER V
DISCUSSION

5.1 Theoretical Interpretation of the Evaluation Gap

5.1.1 GMF's Evaluation Practice as an Organizational Pattern

The five problem foundations identified in Chapter I are consistently supported by the findings of Chapter III. GMF's evaluation is limited to Kirkpatrick Level 1, operationalized through a satisfaction survey that consistently yields high scores across all programs. This creates what the research terms an evaluation ceiling paradox, where near-maximum satisfaction scores generate a false sense of evaluative adequacy while the most consequential questions about training effectiveness remain unasked. Bates (2004) cautioned that satisfaction measures are not merely incomplete but potentially misleading, given their weak correlation with learning or behavioral outcomes. Ahmad et al. (2024) further found that reaction measures explained only a moderate portion of variance in learning outcomes, and that learning and behavior data substantially increased explanatory power when added. GMF's exclusive reliance on Level 1 means it collected data that weakly predicted outcomes while systematically omitting data that would illuminate them.

Three theoretical frameworks converge on this conclusion. From the ADDIE perspective, GMF competently executed the first four phases, conducting needs analysis, designing programs, developing multi-modal content, and delivering instruction, but the evaluation phase that closes the feedback loop was functionally absent, producing what Noe et al. (2017) described as a systems failure where sub-systems operate without integration. From the transfer of learning perspective, the absence of Level 3 evaluation left GMF blind to environmental transfer barriers, which Blume et al. (2019) found explain more variance in transfer behavior than training design quality. Moreover, De Jong et al. (2025) demonstrated that anticipated follow-up assessment itself motivates transfer, meaning GMF's evaluation absence may actively suppress the very behavior it cannot observe. From an andragogical perspective, the spontaneous participant requests for pre- and post-

assessment reflect adult learners' need to understand their developmental position, while persistent requests for more hands-on practice signal a systematic mismatch between GMF's instructional approach and experiential learning orientation. The "Too Technical/Complex" theme in MRO Finance further suggests curriculum calibrated to an assumed rather than empirically established baseline, a condition that andragogy predicts impaired both learning and transfer readiness.

The persistence of near-ceiling satisfaction scores across all three programs warrants interpretive scrutiny beyond face value. Three overlapping mechanisms are plausible. First, the evaluation instrument is administered immediately post-session within the training venue, capturing impressions at a moment of peak engagement rather than reflective assessment; ratings submitted in-group and in-venue tend to compress score distributions upward regardless of actual quality variation (survey fatigue and context effects). Second, in organizational settings characterized by high power distance (Hofstede et al., 2010), subordinate participants evaluating programs endorsed by senior leadership may perceive critical feedback as organizationally risky, generating affirmatively skewed responses that reflect deference rather than candid judgment. Third, in high-context Indonesian professional culture, organizational politeness norms (*ewuh pakewuh*) incline participants toward appreciation for instructors' effort rather than evaluative appraisal of content efficacy. These contextual factors collectively suggest that ceiling effects are not simply an artifact of high training quality but a product of measurement conditions, power dynamics, and cultural norms that the redesigned evaluation instruments must account for on both design and score interpretation.

5.1.2 Why a New Evaluation Framework is Theoretically Necessary

The convergent theoretical analysis establishes that GMF's evaluation gap is not a collection of isolated oversights but a structurally coherent organizational posture with the following characteristics: evaluation that terminates at training completion rather than continuing through application; instruments that measure satisfaction without measuring learning, transfer, or impact; and a training lifecycle in which inputs are carefully managed while outputs remain empirically unexamined.

This posture is inconsistent with contemporary evidence-based training governance. Harvard Business Review (2025) found that organizations employing comprehensive multi-level evaluation reported 2.3 times higher leadership development effectiveness compared to organizations limited to satisfaction assessment. Fortenberry (2025) documented that only 13% of healthcare leadership programs reached Level 4 evaluation, yet those that did demonstrated substantially stronger program justification and stakeholder confidence. These findings establish that the framework proposed in this research is not a theoretical luxury but a strategic necessity for an organization seeking to justify, optimize, and advance its leadership development investment.

5.2 The Proposed Evaluation Framework

This section constitutes the primary contribution of the research. It presents the comprehensive post-training evaluation framework developed in direct response to each identified gap, explains the theoretical rationale for every component, and provides visual representations of both the evaluation framework and the broader training framework within which it operates.

5.2.1 Framework Design Rationale

The framework is grounded in three design commitments that distinguish it from generic evaluation templates. First, it is gap-responsive: every component maps directly to a specific problem foundation identified in Chapter I, ensuring that the framework does not add evaluation for its own sake but addresses documented organizational deficiencies. Second, it is theoretically coherent: components are derived from established theoretical frameworks rather than constructed ad hoc, giving the framework both explanatory validity and scholarly defensibility. Third, it is contextually calibrated: the framework acknowledges GMF's resource constraints, technological infrastructure, organizational culture, and MRO industry characteristics, producing a system that is rigorous in principle while implementable in practice.

5.2.2 The Overall Training Framework: Structural Context

Before presenting the evaluation framework in detail, it is necessary to position it within the broader training lifecycle this research proposed for GMF. The following framework, developed as a core contribution of this research, proposed the ADDIE model as the systematic foundation that GMF currently lacks, organizing the organization's existing training methods into an integrated development cycle with comprehensive evaluation as its culminating and feedback-generating phase.

Table 5.1 Overall Training Framework

ADDIE Model as Training Foundation	
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
Feedback Loop: Evaluation informs next TNA	

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.

This overall framework positioned evaluation not as a terminal administrative activity but as the phase that generated the organizational knowledge necessary for continuous improvement. Without evaluation data feeding back into the analysis phase, TNA operated on assumptions rather than evidence, and each

subsequent training cycle repeated the same design decisions regardless of whether previous iterations achieved their objectives.

5.2.3 The Comprehensive Evaluation Framework

The evaluation framework proposed by this research covered all five levels of the Kirkpatrick-Phillips model and was organized by three dimensions: what was measured, how it was measured, and when it was measured. The following matrix presented the complete framework structure.

Table 5.2 Proposed Comprehensive Post-Training Evaluation Framework

Level	Evaluation Focus	Primary 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	360-degree feedback (self, supervisor, peers, direct	3-6 months post-training	Problem 3: No behavior change tracking	Did participants apply what they learned in

Level	Evaluation Focus	Primary Instrument	Timing	Gap Addressed	Key Question Answered
	application of competencies	reports) Competency-based assessment			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 training investment; cost-benefit analysis	Phillips ROI methodology (benefit isolation, monetary conversion, full cost analysis)	12+ months post-training; Applied selectively to key programs	Problem 5: Inability to calculate ROI	Did the training investment generate adequate financial returns?

5.2.4 Level-by-Level Framework Analysis

A. Level 1: Enhanced Reaction Assessment

While measuring satisfaction, the current Level 1 instrument did not evaluate constructs that had been shown to predict transfer. Furthermore Hirv-Biene et al. (2025) showed that self-efficacy is a more powerful predictor of transfer attempts than satisfaction, De Jong et al. (2025) discovered that behavioral intention and perceived environmental support at training completion predict later transfer behavior. In order to transform Level 1 from satisfaction audit to transfer readiness

diagnostic, the updated Level 1 framework kept the GMF's satisfaction items for continuity and instructor feedback while adding five items related to transfer readiness: behavioral intention, learning confidence, perceived job relevance, expected supervisor support, and resource availability. L&TD can then took action before transfer possibilities expired by using low post-training behavioral intention ratings as early warning indicators.

B. Level 2: Learning Assessment

Chapter III identified the absence of systematic knowledge and skill measurement across all three programs, gap also raised spontaneously by participants who requested pre- and post-assessments, consistent with Knowles et al. (2015) on adult learners' need for structured feedback to support self-directed growth. Pre-test and post-test designs that were in line with the learning objectives of each program were introduced in the proposed Level 2 component. Through scenario-based analytical elements, OLP's assessment focused on strategic thinking, operational leadership, innovation, and business negotiation. Cross-functional integration among marketing, supply chain, maintenance operations, quality assurance, and ERP was measured for MRO management. Because pre-test can show baseline financial knowledge and let teachers to calibrate content from Day 1, assessment was particularly important for MRO Finance given the conclusion that content was seen as being overly technical or complex. The theoretical basis drew on Baldwin and Ford (1988), who argued that training outcomes must be measured against stated objectives to establish content validity. Practically, learning gain data confirm effectiveness, item level analysis guides curriculum refinement, and low post test performers can be identified for targeted support before returning to work contexts where knowledge gaps would hinder transfer.

C. Level 3: Behavioral Evaluation via 360-Degree Feedback

Problem 3, the absence of behavior change tracking, was the most consequential gap in GMF's evaluation practice because transfer determined whether training created organizational value or remains limited to satisfaction metrics, as emphasized by Baldwin and Ford (1988) and Ardondi et al. (2025).

GMF currently lacks a mechanism to verify whether OLP graduates demonstrate improved leadership, whether MRO Management participants apply cross functional thinking, or whether developed competencies translate into observable workplace behavior. The proposed Level 3 component introduces 360 degree multi source feedback conducted 3 to 6 months post training to allow sufficient time for behavioral change to emerge while maintaining proximity to the intervention. Using the same competency framework and rater groups as existing pre assessments enables rigorous pre post comparison and maximizes prior investment. The framework evaluates eight leadership dimensions relevant to MRO contexts, each defined through observable behavioral indicators rated by self, supervisors, peers, and direct reports to enhance validity, consistent with Church et al. (2019). Beyond measurement, institutionalizing Level 3 assessment also strengthens transfer motivation, as De Jong et al. (2025) show that anticipated behavioral evaluation increases application, thereby reinforcing the very behavior the system is designed to assess.

D. Level 4: Results Linkage

The lack of a connection between organizational performance outcomes and training attendance is the subject of Problem 4. According to Phillips and Phillips (2016), level 4 evaluation is methodologically challenging since it necessitates elaborate controls to separate training effects from other performance determinants. This study suggests a practical convergent evidence technique that triangulates several data sources rather than striving for methodological perfection. In order to assess trends against organizational norms, departmental performance comparisons first track important indicators including operational turnaround efficiency, safety incidents, engagement ratings, and customer happiness for a full year before and after training. Second, after adjusting for size and operational complexity, comparative analysis compares departments under the direction of skilled managers with comparable units under the direction of untrained managers. Third, structured stakeholder perception surveys collect senior leader evaluations of post training impact to capture qualitative organizational effects. Consistent with Kirkpatrick and Kirkpatrick (2016), causal attribution in organizational settings is probabilistic, not

deterministic. While no single indicator can prove causality, convergent positive trend across independent measures provided defensible level of organizational confidence in training impact, which was substantially stronger than the current absence of evidence.

E. Level 5: Return on Investment

Problem 5 concerned the inability to calculate training ROI, which limited GMF's capacity to engage senior leadership in financial terms. Phillips and Phillips (2016) proposed converting Level 4 resulted into monetary values, subtracting fully loaded program costs, and expressing net benefits as an ROI percentage. Applied to OLP, this required five steps: collecting Level 1 to 4 data, isolating training contribution through conservative attribution estimates, converting performance gains into monetary equivalents using GMF cost and productivity data, calculating total program costs including participant opportunity time, and computing the ROI percentage. This research recommendeded selective ROI application focused on OLP as the flagship leadership investment, consistent with Phillips and Phillips (2016), who advised that ROI studies were most appropriate for high investment programs where financial justification was strategically significant. For MRO Management and MRO Finance, Levels 1 to 3 provided sufficient governance evidence without full ROI analysis. Strategically, Level 5 was primarily communicative. In BUMN governance contexts where training competed with operational expenditures, an ROI figure reframed the discussion from normative importance to quantified financial return per rupiah invested, strengthening L&TD's position as a strategic investment function rather than administrative service unit.

5.3 Implementation Framework and Strategic Implications

5.3.1 Phased Implementation Structure

Comprehensive multi-level evaluation cannot be implemented simultaneously without exceeding L&TD's current capacity. Training evaluation literature recommends phased rather than full-scale deployment (Phillips &

Phillips, 2016; Kirkpatrick & Kirkpatrick, 2016). Phillips and Phillips (2016) advised organizations without dedicated evaluation expertise to build capability incrementally, starting from simpler levels before progressing to more complex ones. Kirkpatrick and Kirkpatrick (2016) suggested allowing adequate time for behavioral change to emerge before conducting Level 3 evaluation, typically 3 to 6 months post-training, and for business results to materialize before Level 4 assessment, at least 6 to 12 months. Change management research further indicated that major organizational process changes require 18 to 24 months for institutionalization (Kotter, 1996). Based on these principles and GMF's organizational context, this research proposed a phased implementation aligned with theoretical priority, methodological complexity, and organizational readiness as outlined in Table 5.2.

Table 5.3 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; multi-level 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	ROI evidence; institutionalized system	Sufficient data for credible ROI; 18-24

Phase	Timeline	Priority Actions	Expected Outcomes	Rationale
		instruments; Document SOPs		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. Timeline recommendations are grounded in training evaluation implementation research (Phillips & Phillips, 2016; Kirkpatrick & Kirkpatrick, 2016) and organizational change theory (Kotter, 1996), contextualized to GMF's operational characteristics. GMF should adapt phasing and duration based on actual implementation experience, resource availability, and strategic priorities

Level 1 (Enhanced Reaction) has high feasibility for immediate implementation. GMF already conducts post-training surveys using Google Forms, so the required technology and participant familiarity already exist. Implementing the enhanced survey only requires revising the current questionnaire to include items on behavioral intention and transfer readiness. No new technology or additional staff capability is required, making Level 1 suitable for Phase 1 implementation with minimal cost.

Level 2 (Learning: Pre/Post Knowledge Assessment) has moderate feasibility. Google Forms can also be used to administer pre-tests and post-tests, meaning the basic infrastructure is available. However, developing valid knowledge assessment questions requires instructional design skills to ensure alignment with learning objectives. L&TD staff may therefore need limited training or consultation to develop appropriate test items. This requirement can be addressed through review and pilot testing during Phase 1.

Level 3 (Behavioral Transfer: 360-Degree Feedback) has moderate-to-high feasibility because GMF already conducts pre-training 360-degree competency assessments for certain leadership programs. This means the organization already has experience with the instrument and multi-rater evaluation process. The current gap is the absence of a post-training assessment conducted three to six months after

training to measure behavioral change. Implementing this level therefore requires scheduling follow-up assessments, adjusting the instrument for comparison, and assigning administrative responsibility within L&TD.

Level 4 (Results: Organizational Impact Tracking) has moderate feasibility but requires coordination across departments. Departmental KPIs are already monitored in GMF's operational systems, meaning the relevant data exists. The main challenge is linking training participation with KPI performance before and after training. Achieving this linkage may require coordination between L&TD, the Human Capital Management Division, and operational units.

Level 5 (ROI Calculation) has the lowest short-term feasibility. The ROI methodology proposed by Phillips (2012) requires converting performance outcomes into monetary values, which demands financial analysis skills and access to cost data that are not currently within the L&TD unit's responsibilities. Implementing ROI evaluation would therefore require collaboration with GMF's Finance Division. For this reason, Level 5 is more suitable for later implementation after Levels 1–4 are already established.

5.3.2 Addressing the Research Questions

The following discussion contextualized each research question within the theoretical frameworks elaborated in Chapter IV, drawing on findings from Chapter III to develop interpretive responses rather than merely descriptive summaries. Definitive conclusive statements were provided in Chapter VI.

RQ 1: What are the current training evaluation practices at GMF, and what gaps exist relative to the Kirkpatrick-Phillips model?

Chapter III documented and Section 5.2 theoretically analyzed GMF's current evaluation: a Level 1-only satisfaction survey generating high but psychometrically limited data, with complete absence of assessment at Levels 2 through 5. The gaps corresponded directly to the five problem foundations: satisfaction without learning verification (Problem 1 and 2), no behavioral tracking (Problem 3), no business results linkage (Problem 4), and no capacity for ROI calculation (Problem 5). Theoretically, this gap pattern reflected an organizational

posture that treated training completion as an evaluative endpoint rather than a developmental milestone, an assumption inconsistent with both Kirkpatrick-Phillips evaluation theory (Kirkpatrick, 2016) and adult learning principles (Knowles et al., 2015).

RQ 2: How can a comprehensive post-training evaluation framework be designed and adapted for GMF's MRO management development programs?

The framework presented in Section 5.2 constituted the direct answer. Each of its five levels was theoretically grounded, gap-responsive, and contextually calibrated to GMF's organizational characteristics. The framework was differentiated by program: MRO Finance received Levels 1-2 given its foundational knowledge-building objective; MRO Management received Levels 1-3 given its leadership competency development focus; OLP received all five levels given its status as the flagship program with the highest strategic and financial investment. The instruments proposed (enhanced survey, pre/post assessments, 360-degree feedback, KPI tracking, ROI methodology) were derived from established theoretical and methodological traditions while being adapted to GMF's available technology infrastructure (Google Forms, Microsoft Excel, Microsoft Forms), resource constraints, and MRO industry context.

RQ 3: What implementation guidelines and practical instruments are needed to operationalize the framework within GMF's constraints?

Section 5.3.1's phased implementation framework provided the operational sequencing. Key implementation principles derived from the analysis were: begin with the level of greatest theoretical priority and practical feasibility (enhanced Level 1 and Level 2, which require no new technology or expertise); build evaluation capability incrementally before advancing to more complex levels; leverage existing GMF practices (the 360-degree pre-assessment) by extending rather than replacing them; and embed evaluation design into training design rather than treating it as a post-hoc addition. The feedback loop structure of the proposed overall training framework (Figure 5.1) provided the architectural principle ensuring that evaluation findings systematically inform subsequent training cycles rather than accumulating in reports without organizational consequence.

5.3.3 Framework Transferability and Contextual Boundaries

The proposed evaluation framework consists of two layers of components that differ in their applicability across organizations and industries. This distinction determines which parts of the framework can be applied directly in other organizations and which parts require adjustment to fit different contexts.

The first layer contains universal components based on the Kirkpatrick-Phillips framework and general evaluation principles. These elements can be applied to any organization implementing management development programs, regardless of industry or governance structure. They include the five-level evaluation hierarchy with its phased implementation logic, the use of pre- and post-tests for Level 2 learning assessment, the recommended three-to-six-month timing for Level 3 behavioral assessment, the use of existing organizational KPIs for Level 4 results measurement, and the selective application of ROI evaluation to high-investment training programs (Kirkpatrick & Kirkpatrick, 2016). These components provide a general structure for building a comprehensive training evaluation system.

The second layer consists of context-specific components that are tailored to GMF's aviation MRO and BUMN environment. These include the leadership competencies used in the Level 3 behavioral assessment, which reflect aviation-related capabilities such as safety culture and regulatory accountability. The Level 4 performance indicators are also specific to the MRO industry, including operational metrics such as aircraft turnaround efficiency and safety incident rates. In addition, the ROI communication approach reflects the financial accountability practices commonly required in Indonesia's BUMN governance system.

Organizations that wish to adopt this framework can apply the universal components directly while adapting the context-specific elements to their own industry and organizational environment. This adaptation should begin with a diagnostic analysis similar to the process conducted in Chapter III, including reviewing existing evaluation practices, identifying gaps based on the five-level evaluation structure, and mapping available organizational data to evaluation requirements. This approach follows the principle of Design and Development

Research, where frameworks developed in a specific context are transferred through adjustment of context-dependent elements rather than direct replication (Richey & Klein, 2007).

5.4 Integrative ADDIE-Kirkpatrick-Phillips Conceptual Model

To address the documented limitation of satisfaction-based evaluation at GMF, this research proposed integrated training and evaluation framework that embeded the Kirkpatrick–Phillips five-level evaluation model within the ADDIE instructional design process. The framework systematically connected Training Needs Analysis, structured program delivery, multi-level outcome measurement, and financial accountability into a continuous improvement system. By linking Reaction, Learning, Behavior, Results, and ROI to organizational strategic objectives, the model transformed training from isolated development activity into evidence-based performance improvement mechanism aligned with GMF's MRO operational and governance context.

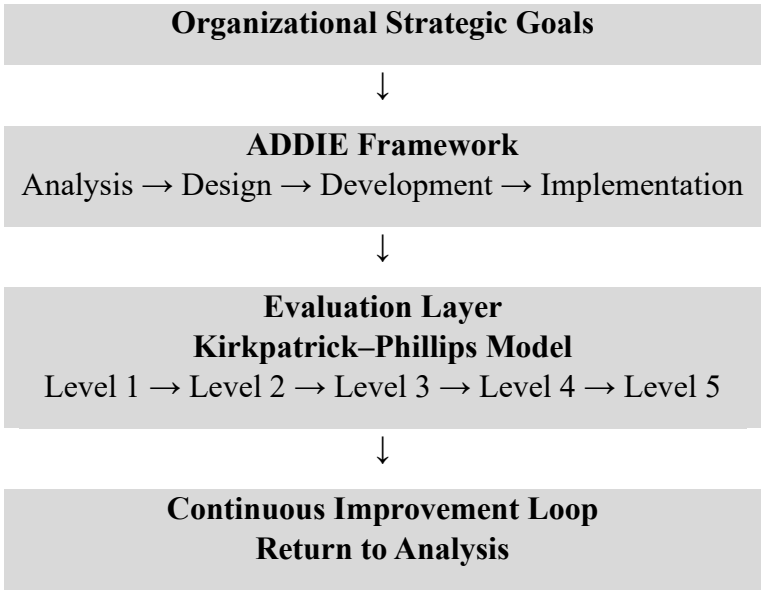


Figure 5.1. Integrated ADDIE and Proposed Kirkpatrick-Phillips Evaluation Framework

A. Analysis Stage

1) Identify competency and performance gaps through Training Needs Analysis.
2) Define expected organizational results (Level 4).
3) Establish baseline data for future ROI calculation (Level 5).

B. Design Stage

1) Formulate measurable learning objectives.
2) Align objectives with Learning (Level 2), Behavior (Level 3), and relevant KPIs (Level 4).

C. Development Stage

1) Develop enhanced Reaction survey (Level 1).
2) Design pre- and post-tests (Level 2).
3) Prepare 360-degree behavior tools (Level 3).
4) Create KPI tracking and ROI framework (Levels 4–5).

D. Implementation Stage

1) Deliver training under the ADDIE structure.
2) Conduct Level 1 and Level 2 assessments.
3) Integrate transfer action planning for behavior application.

E. Post-Implementation Evaluation

1) Measure behavior change (Level 3).
2) Monitor organizational performance impact (Level 4).
3) Calculate ROI for selected programs (Level 5).
4) Use findings to refine future analysis and program design.

5.5 Theoretical Contributions and Limitations

5.5.1 Contributions to Theory

This research made three theoretical contributions to the literature on organizational training evaluation. First, the evaluation ceiling paradox was proposed as a working conceptual lens that builds on Bates' (2004) Level 1 critiqued by highlighting a dynamic mechanism through which consistently high satisfaction

scores may suppress organizational motivation for deeper evaluation. While this concept required empirical validation across broader organizational contexts, its application in the GMF case offered an interpretive framework that complements existing psychometric critiques of Level 1 measures.

Second, the framework's differentiated application across programs (Levels 1-2 for foundational programs, Levels 1-3 for competency programs, Levels 1-5 for flagship investments) contributed a practically grounded approach to the theoretical question of evaluation selectivity. Phillips and Phillips (2016) advocated ROI selectivity on resource grounds; this research extended the argument by tying selectivity to program type and developmental objective rather than cost alone.

Third, the research provided empirical grounding for the ADDIE framework's evaluative deficit hypothesis (Peck, 2025; Educational Technology, 2018) in a non-Western, aviation-industry context, demonstrating that the pattern of strong implementation with weak evaluation was not culturally or industrially specific but represented a recognizable organizational dynamic across contexts.

5.5.2 Limitations

Several limitations affected the scope of the findings. The research was conducted at a single organization during a specific institutional period, limiting direct generalizability to other MRO contexts or BUMN subsidiaries. The qualitative analysis of participant feedback relied on spontaneously generated open-ended responses rather than systematically administered qualitative instruments, meaning the thematic patterns documented in Chapter III represented expressed but not comprehensively elicited participant perspectives. A further limitation pertained to the inference-based derivation of learning objectives in Section 1.3.2 (Specific Objective 2). In the absence of formal learning objective documentation from program designers, learning objectives were inferred from module content, syllabi, and instructional materials using content analysis and Bloom's Taxonomy alignment. While this approach was grounded in established instructional design methodology, the resulting objectives represented the researcher's analytical interpretation rather than formally validated design intent. Accordingly, the Level 2 assessment instruments built upon these inferred objectives should be reviewed and

confirmed by instructional designers or subject matter experts within L&TD prior to operational deployment. Furthermore, the evaluation framework proposed in this research was theoretical and design-oriented: its effectiveness can only be empirically validated through actual implementation, which constituted a natural direction for future research.

CHAPTER VI
CONCLUSION

6.1 Summary of Research Findings

This research investigated PT GMF Aero Asia's post-training evaluation practices and developed a comprehensive evaluation framework based on the Kirkpatrick-Phillips model. Through qualitative-primary analysis of 638 module-level evaluation responses collected across three management development programs and 584 qualitative comments, supported by descriptive quantitative analysis as supplementary data and complemented by document analysis, the research documented GMF's current evaluation landscape and developed a theoretically grounded, contextually calibrated five-level framework addressing identified deficiencies.

A. Current Evaluation Practice

GMF's evaluation was uniformly confined to Level 1 satisfaction assessment across all management development programs, administered via Google Forms immediately post-training. While generating consistently high scores (OLP: $M = 4.91$; MRO Management: $M = 4.86$; MRO Finance: $M = 4.63$), these scores clustered near scale maximum, creating an evaluation ceiling paradox: psychometrically limited data producing false organizational confidence while more consequential evaluation questions remain unasked.

B. Identified Evaluation Gaps

Five systematic gaps corresponded to Chapter I problem foundations: (1) limited evaluation scope prevents determining whether satisfaction translates to learning, behavior change, or business impact; (2) absence of learning measurement means no verification of knowledge acquisition or identification of individuals needing support; (3) no behavior change tracking means evaluation concludes at program completion without workplace application follow-up; (4) lack of business results linkage prevents demonstrating training's contribution to organizational

performance; (5) inability to calculate ROI means no cost-benefit analysis or financial justification for executive decision-making.

C. Participant Perspectives

Qualitative analysis revealed patterns extending beyond satisfaction: spontaneous requests for pre/post assessment (2.2% of comments) reflected adult learners' need to know developmental position; requests for more hands-on practice (4.8%) signaled mismatch between instruction and experiential learning orientation; the "Too Technical/Complex" theme in MRO Finance (4.0%) indicated curriculum calibrated to assume rather than empirically established baseline knowledge.

D. Theoretical Framework

The comprehensive evaluation framework addressed each gap through theoretically grounded components: enhanced Level 1 added transfer-readiness constructs; Level 2 introduced pre/post testing measuring knowledge acquisition; Level 3 employed 360-degree feedback assessing behavioral change 3-6 months post-training; Level 4 linked training to organizational performance through convergent evidence strategies; Level 5 applied Phillips ROI methodology selectively to flagship programs. The framework was differentiated by program type with phased 24-month implementation enabling capability building without overwhelming L&TD capacity.

6.2 Conclusive Answers to Research Questions

The following provided concise and evidence-based conclusions to each research question as derived from the integrated analysis presented in Chapters III-V.

RQ 1: What are the current training evaluation practices at PT GMF Aero Asia, and what gaps exist compared to comprehensive evaluation frameworks based on the Kirkpatrick-Phillips model?

GMF's current practice consists exclusively of Level 1 satisfaction surveys administered immediately post-training via Google Forms, capturing instructor quality, material quality, and time adequacy. Compared to the Kirkpatrick-Phillips

framework, systematic gaps exist at Levels 2 through 5: no pre/post knowledge testing, no workplace application follow-up, no linkage to organizational performance metrics, and no capacity for ROI calculation. These gaps are documented in detail in Section 3.5 and theoretically interpreted in Section 5.1.

RQ 2: How can a comprehensive post-training evaluation framework based on the Kirkpatrick-Phillips model be designed and adapted for GMF's MRO management development programs?

The proposed framework was designed according to three principles: gap-responsiveness, theoretical coherence, and contextual calibration. Each level addresses a documented deficiency: Level 1 is enhanced with five transfer-readiness items; Level 2 introduces pre/post testing calibrated to program-specific objectives; Level 3 employs 360-degree multi-source feedback 3 to 6 months post-training using eight MRO-relevant leadership competencies; Level 4 links training to organizational performance through three convergent strategies; and Level 5 applies Phillips ROI methodology selectively to OLP only. Differentiated application across programs (Levels 1 to 2 for MRO Finance, Levels 1 to 3 for MRO Management, and Levels 1 to 5 for OLP) reflects each program's investment level, developmental objective, and evaluation feasibility. The full framework design rationale and level-by-level specifications are presented in Section 5.2.

RQ 3: What implementation guidelines and practical instruments are needed to operationalize the framework within GMF's constraints?

Operationalization of the framework requires six implementation elements: phased sequencing that builds evaluation capability incrementally without overwhelming L&TD capacity; leverage of existing technology infrastructure (Google Forms, Microsoft Excel) to avoid requiring new investments; clear role definition distributing responsibilities across L&TD, instructors, HR analytics, line managers, and senior leadership; integration with rather than replacement of existing evaluation practices; stakeholder communication and change management to secure commitment; and standard operating procedures ensuring sustainability beyond individual staff. A phased 24-month implementation timeline structures these elements into manageable stages. The practical instruments supporting each

level, including the enhanced reaction survey, pre/post-test templates, 360-degree behavior assessment, KPI tracking template, and ROI calculation worksheet, are presented in Chapter V. Full implementation guidelines are elaborated in Section 5.3.

6.3 Research Contributions

6.3.1 Theoretical Contributions

A. Multi-Theory Integration as Framework Design Principle

The primary theoretical contribution lied in the explicit integration of three distinct but complementary frameworks: the Kirkpatrick-Phillips evaluation hierarchy, Baldwin and Ford's (1988) transfer of learning model, and Knowles et al.'s (2015) andragogical principles. Kirkpatrick's model had been criticized for its limited account of why transfer does or does not occur (Bates, 2004; Blume et al., 2019). This research addressed that gap by treating the frameworks as theoretically complementary: transfer theory identified the conditions enabling Level 3 outcomes that Kirkpatrick described but did not explain, while andragogy shaped how evaluation instruments must be designed to remain credible and motivating for adult professional learners. This synthesis advanced a principle of evaluation framework design as theory integration rather than single-model application, demonstrating that Kirkpatrick's practical strengths and theoretical gaps can be addressed by pairing it with complementary theoretical lenses rather than replacing it altogether.

B. Evaluation Ceiling Paradox

This research proposed the evaluation ceiling paradox as an analytical lens for interpreting GMF's evaluation pattern and suggested a self-reinforcing mechanism by which it may operate. Prior critiques (Bates, 2004) identified Level 1 data's psychometric limitations; this research extended that critique by hypothesizing a second-order organizational effect whereby those limitations might perpetuate themselves through institutional rationalization, cultural score compression, and the availability heuristic. Empirical validation of this concept across other organizational contexts represented a direction for future research.

C. Differentiated Evaluation Application and Contextual Adaptation

While Phillips and Phillips (2016) advocated ROI selectivity on resource grounds, this research extended the argument by tying evaluation depth to program type, developmental objective, and organizational context. The framework's differentiated application (Levels 1-2 for foundational programs, 1-3 for competency programs, 1-5 for flagship investments) offered a practically grounded typology addressing a gap in literature that tended to present Kirkpatrick-Phillips as uniform prescription rather than adaptable framework. Additionally, the dual contextualization of evaluation theory to the aviation MRO industry and the Indonesian BUMN setting advanced a principle that evaluation frameworks were contextually contingent: MRO leadership required instruments sensitive to safety culture and regulatory accountability, while the BUMN governance context shaped how evaluation instruments must be designed and how scores must be interpreted. Together, these contributions demonstrated that training evaluation theory gained analytical power when treated as context-sensitive rather than universally portable.

D. ADDIE Evaluative Deficit

This research provided empirical grounding for theoretical observations about ADDIE implementation gaps (Peck, 2025; Educational Technology, 2018) in non-Western, aviation-industry context, demonstrating that evaluation deficit transcends cultural and industrial boundaries and represented distinct organizational disposition requiring specific cultivation rather than emerging naturally from training maturity.

6.3.2 Methodological Contributions

This research exemplified design and development research in HRD, demonstrating systematic translation of theoretical models into practical instruments while balancing academic rigor with organizational feasibility. The effective integration of quantitative descriptive analysis, qualitative thematic analysis, and document analysis producing richer understanding than any single method provided methodological template for applied HRD research, particularly valuable for emerging scholarship in developing country contexts.

6.3.3 Practical Contributions

GMF receives complete, contextually calibrated evaluation framework with ready-to-implement instruments specifically designed for its three programs, technological infrastructure, resource constraints, organizational culture, and MRO industry characteristics. The framework enabled evidence-based program governance through objective data on learning achievement, behavioral application, organizational impact, and financial returns, supporting data-driven decisions about curriculum revision, instructor selection, program scaling, or discontinuation. For GMF as BUMN subsidiary, the framework provided empirical evidence positioning L&TD as strategic investment function rather than administrative service unit, strengthening budget justification and organizational influence. While designed for GMF, the framework was adaptable by other MRO organizations facing similar evaluation challenges, contributing to industry-wide professionalization of training assessment.

6.4 Research Limitations

A. Single-Organization Focus

Conducted at one organization during specific period (2024-2025), limiting direct generalizability to other MRO contexts, BUMN subsidiaries, or Indonesian organizations broadly. GMF's characteristics might not represent typical organizations.

B. Data Collection Constraints

Qualitative analysis relied on spontaneously generated comments rather than systematically administered instruments (focus groups, interviews), meaning thematic patterns represented expression rather than comprehensively elicited perspectives.

C. Framework Validation

The evaluation framework is theoretical and design-oriented; effectiveness can only be validated through actual implementation. While grounded in established theoretical frameworks and designed for GMF's context, practical

feasibility, stakeholder acceptance, implementation challenges, and actual impact remained to be demonstrated.

D. Temporal and External Validity

Research analyzed one year of data, potentially not capturing longer-term trends or cyclical variations. Transferability to non-Indonesian contexts, non-aviation industries, or private-sector organizations required contextual adaptation accounting for different governance frameworks, cultural norms, resource availability, and strategic priorities.

E. Framework Development without Empirical Validation

This research was explicitly designed and had development research: its contribution is a theoretically grounded framework and associated instruments, not an empirically validated system with demonstrated effectiveness. The proposed instruments had not been pilot tested with GMF participants, had not undergone psychometric validation (reliability analysis, construct validity testing), and had not been tested for practical feasibility under real implementation conditions. Claims regarding the framework's ability to improve training outcomes remained theoretical propositions. Real-world implementation might reveal stakeholder resistance, logistical constraints, or cultural dynamics not fully anticipated by the framework design. Longitudinal validation research should be prioritized as the primary next step (see Section 6.5.3).

F. Absence of Level 3, Level 4, and In-Depth Interview Data

Framework development research conducted within a single internship period cannot generate the behavioral and organizational performance data that the framework proposed to collect. This research contained no Level 3 data on whether trained managers changed their leadership behaviors and no Level 4 data on departmental performance outcomes following training participation. All inferences regarding training impact on workplace performance were theoretical extrapolations from Level 1 patterns and qualitative feedback. The research also relied on spontaneously generated participant comments rather than purposively elicited perspectives from structured in-depth interviews with line managers,

L&TD leadership, or participants, limiting the depth of stakeholder sense-making that interview-based inquiry would provide.

6.5 Recommendations

6.5.1 For GMF Aero Asia

Immediate Actions (0-6 Months)

1. **Enhance Level 1** immediately by adding five transfer-readiness items to existing survey. Required minimal time (2-3 minutes additional), no new technology, produced actionable data identifying transfer risk.
2. **Pilot Level 2 for MRO Finance** given documented complexity concerns. Pre-testing established actual baseline knowledge enabling real-time curriculum calibration; post-testing verified learning and identified individuals needing support.
3. **Conduct stakeholder briefing** for senior leadership, instructors, and HR partners explaining framework rationale, resource requirements, and phased timeline, building support and securing commitments.

Short-Term Actions (6-12 Months)

4. **Scale Level 2** to all programs following successful pilot, involving instructors to ensure alignment with learning objectives and incorporating scenario-based items requiring analytical judgment.
5. **Build L&TD evaluation capability** through external consultant workshops or university partnerships providing training in test development, survey design, 360-degree interpretation, statistical analysis, and ROI methodology.

Medium-Term Actions (12-24 Months)

4. **Implement Level 4 results tracking** for OLP graduates, partnering with HR analytics for performance data focusing on metrics already tracked organizationally.

5. **Conduct ROI study for OLP** following Level 1-4 data collection, employing conservative attribution, transparent assumptions, and sensitivity analysis formatted for executive audience.
6. **Document standard operating procedures** for evaluation administration, data collection, quality assurance, analysis, reporting, and dissemination ensuring consistency and sustainability.

Strategic Actions (Beyond 24 Months)

9. **Integrate evaluation into training design** by specifying learning objectives explicitly, developing assessments concurrently with curriculum, and defining performance metrics programs aim to improve.
10. **Establish feedback loop** from evaluation to program design through annual review meetings examining evaluation data driving specific curriculum modifications and teaching method adjustments.
11. **Extend framework to technical training**, piloting adapted approach for high-priority technical programs with Level 2 emphasizing performance-based testing and Level 4 emphasizing safety and quality metrics.

6.5.2 For Other Organizations

Organizations facing similar challenges should: (1) systematically diagnose existing evaluation practice against comprehensive frameworks identifying specific gaps; (2) prioritize transfer-focused evaluation recognizing that training success is determined by workplace application; (3) adapt rather than adopt generic frameworks to organizational context, program type, and resources; (4) build incrementally through phased implementation managing risk and generating momentum; (5) communicate evaluation's strategic value as capability enabling evidence-based decisions rather than administrative burden.

6.5.3 For Future Research

Six research directions merit investigation: (1) longitudinal validation research tracking GMF's framework implementation examining challenges, stakeholder acceptance, resource requirements, and whether comprehensive evaluation improves training effectiveness; (2) cross-organizational comparison

examining evaluation practices across multiple Indonesian MRO organizations identifying contextual factors associated with evaluation maturity; (3) transfer barrier investigation through qualitative research identifying common obstacles preventing workplace application in aviation MRO contexts; (4) cultural adaptation research investigating appropriate implementation of 360-degree feedback in Indonesian organizational contexts balancing validity with cultural sensitivity; (5) ROI methodology refinement for emerging markets addressing attribution challenges and monetary conversion difficulties; (6) AI-enabled evaluation innovation exploring machine learning applications in automated analysis, predictive modeling, and pattern recognition.

6.6 Concluding Remarks

This research addressed a core question: how can GMF determine whether its management development programs actually produce meaningful learning and organizational impact? The findings revealed that GMF evaluates training primarily through participant satisfaction and leaving deeper questions unanswered. In response, this research developed a comprehensive evaluation framework that translated established models into practical and accessible tools suited to real organizational constraints.

In addition, the research argued that training evaluation should be treated as a strategic capability, not an administrative formality. For GMF, operating within aviation MRO workforce pressures and BUMN accountability standards, systematic evaluation positions L&TD as a value-generating function rather than a support service, with the evidence to justify its standing.

The real barriers for evaluation were not technical but cultural. The proposed instruments were methodologically straightforward; what made evaluation rare was organizational will. This research equipped GMF with the frameworks, tools, and rationale needed. What it had put within reach was the capability GMF previously lacked: to know empirically whether its leadership development investment achieved its purpose and to use that knowledge to improve how training is designed, delivered, and strategically positioned.

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

Ahmadi, H. (2020). ADDIE instructional design model: Development phase in e-learning environments. *Journal of Educational Technology, 17*(2), 45–58.

AIHR. (2024). *Training needs analysis: A step-by-step guide.* Academy to Innovate HR. https://www.aihr.com/blog/training-needs-analysis/

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

Allen, M. W. (2006). *Creating successful e-learning: A rapid system for getting it right the first time, every time.* Pfeiffer.

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

Aviation Pros. (2016). *Leadership requirements in the MRO industry.* Aviation Pros. https://www.aviationpros.com/mro/article/leadership-requirements

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

Bassi, L. J., Benson, G., & Cheney, S. (1996). The top ten trends. *Training & Development, 50*(11), 28–42.

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

Bothell, T. W., & Thompson, C. (2002). Evaluating the return on investment of blended learning. *Educational Technology, 43*(3), 37–45.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Burke, L. A., & Hutchins, H. M. (2007). Training transfer: An integrative literature review. *Human Resource Development Review, 6*(3), 263–296. https://doi.org/10.1177/1534484307303035

Chee, T. S., Lim, C. S., & Ahmad, M. F. (2023). Instructional implementation in corporate training: Managing learner engagement and logistical challenges. *Journal of Workplace Learning, 35*(4), 201–218. https://doi.org/10.1108/JWL-09-2022-0131

Chief Learning Officer. (2024). *Gamification in corporate training: What the research says.* CLO Media. https://www.chieflearningofficer.com/gamification-training

Church, A. H., Bracken, D. W., Fleenor, J. W., & Rose, D. S. (Eds.). (2019). *The handbook of strategic 360 feedback.* Oxford University Press.

Coman, C., Țîru, L. G., Meseșan-Schmitz, L., Stanciu, C., & Bularca, M. C. (2020). Online teaching and learning in higher education during the COVID-19 pandemic: Students' perspective. *Sustainability, 12*(24), 10367. https://doi.org/10.3390/su122410367

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

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

Educational Technology. (2018). *ADDIE model: Strengths, weaknesses, and variations.* Educational Technology. https://educationaltechnology.net/addie-model

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

Harvard Business Review. (2025). *The state of leadership development: Global survey of 1,847 organizations.* Harvard Business Review. https://hbr.org/2025/01/global-leadership-development-survey

Hirv-Biene, J., Leppänen, E., & Väljataga, T. (2025). Environmental determinants of training transfer: A multilevel analysis. *Human Resource Development International, 28*(2), 205–228. https://doi.org/10.1080/13678868.2025.2289741

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind* (3rd ed.). McGraw-Hill.

InstructionalDesign.org. (2018). *ADDIE model.* https://www.instructionaldesign.org/models/addie/

Jacobs, R. L., & Park, Y. (2009). A proposed conceptual framework of workplace learning: Implications for theory development and research in human resource development. *Human Resource Development Review, 8*(2), 133–150. https://doi.org/10.1177/1534484309334269

Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). *Kirkpatrick's four levels of training evaluation.* Association for Talent Development.

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.

Kurt, S. (2017). *ADDIE model: Instructional design.* Educational Technology. https://educationaltechnology.net/addie-instructional-design-model/

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* SAGE Publications.

LinkedIn. (2023). *2023 workplace learning report: Building the agile future.* LinkedIn Learning. https://learning.linkedin.com/resources/workplace-learning-report

Liu, Y., Chen, X., Wang, H., & Zhang, L. (2025). Applying the Kirkpatrick model in standardized nursing training: A longitudinal study with 12-month follow-up. *Nurse Education Today, 135*, 106104. https://doi.org/10.1016/j.nedt.2025.106104

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.

Moore, K. (2023). Overcoming the analysis-skip problem in instructional design: Causes and remedies. *Performance Improvement, 62*(3), 17–25. https://doi.org/10.1002/pfi.22048

Noe, R. A., Hollenbeck, J. R., Gerhart, B., & Wright, P. M. (2017). *Human resource management: Gaining a competitive advantage* (10th ed.). McGraw-Hill Education.

Nunes, M. R., Tomé, E., & Tomé, G. (2025). Training needs analysis in the digital era: A systematic literature review. *European Journal of Training and Development, 49*(3), 201–219. https://doi.org/10.1108/EJTD-07-2024-0098

Papadopoulou, A., Roussaki, I., & Kosmatos, E. (2022). Instructional design blueprinting: Frameworks and best practices for training program development. *Educational Technology Research and Development, 70*(4), 1189–1210. https://doi.org/10.1007/s11423-022-10089-4

Peck, K. (2025). Revisiting ADDIE: Strengths, limitations, and implementation gaps in contemporary instructional design. *TechTrends, 69*(2), 45–54. https://doi.org/10.1007/s11528-025-00943-0

Phillips, J. J. (2003). *Return on investment in training and performance improvement programs* (2nd ed.). Routledge.

Phillips, J. J., & Phillips, P. P. (2016). *Handbook of training evaluation and measurement methods* (4th ed.). Routledge.

Podoshen, J. S. (2024). Adult learning principles and evaluation design: An andragogical framework for training assessment. *Journal of Workplace Learning, 36*(4), 289–305. https://doi.org/10.1108/JWL-02-2024-0021

Rafiola, R. H., & Umar, M. (2024). Development phase in ADDIE-based instructional design: Practices and challenges in Indonesian higher education. *Journal of Education and Learning, 18*(1), 112–122. https://doi.org/10.11591/edulearn.v18i1.21234

Research.com. (2025). *ADDIE model explained: History, phases, and applications.* Research.com. https://research.com/education/addie-model

Richey, R. C., & Klein, J. D. (2007). *Design and development research: Methods, strategies, and issues.* Lawrence Erlbaum Associates.

ROI Institute. (2021). *ROI methodology: The five-level framework for measuring training impact.* ROI Institute. https://www.roiinstitute.net/roi-methodology

Stufflebeam, D. L. (1971). *Educational evaluation and decision making.* Peacock.

Sugiyono. (2019). *Metode penelitian kuantitatif, kualitatif, dan R&D* (2nd ed.). Alfabeta.

Taylor, D. C. M., & Hamdy, H. (2013). Adult learning theories: Implications for learning and teaching in medical education: AMEE Guide No. 83. *Medical Teacher, 35*(11), e1561–e1572. https://doi.org/10.3109/0142159X.2013.828153

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

Udeh, C., & Castro, I. (2025). Gamification in employee training: A meta-analytic review of effects on motivation, knowledge retention, and performance. *Human Resource Development Review, 24*(2), 112–140. https://doi.org/10.1177/15344843251289852

Warr, P., Bird, M., & Rackham, N. (1970). *Evaluation of management training.* Gower.

Xperiencify. (2024). *Collaborative learning: The benefits, strategies, and statistics.* Xperiencify. https://xperiencify.com/collaborative-learning

360Learning. (2020). *The collaborative learning report 2020.* 360Learning. https://360learning.com/reports/collaborative-learning-report

360Learning. (2023). *The future of learning & development 2023.* 360Learning. https://360learning.com/reports/future-of-learning