



Evaluating Indonesia's first-level police staff and leadership education using the Kirkpatrick model

Eko Wirawan¹, Nurhattati Fuad², Matin³

^{1,2,3} Universitas Negeri Jakarta, Jakarta, Indonesia

wiraekowan@gmail.com¹, nurhattati@unj.ac.id², martin@unj.ac.id³

ABSTRACT

Police leadership education is a substantial investment, yet its effectiveness is seldom evaluated beyond participant satisfaction, and evidence on tiered police managerial education remains scarce. This study evaluates Indonesia's First-Level Staff and Leadership School (Sespimma) using Kirkpatrick's four-level model, aiming to assess reaction, learning, behavior, and results, and to identify the factors that enable and hinder the application of the education. It adopts a qualitative, multi-source case study. An in-depth interview with an alumnus, together with observation and performance documents (LAKIP and IKU), was coded thematically in ATLAS.ti 24 using deductive and inductive coding; open-ended questionnaire responses from sixteen alumni and two supervisors were analyzed as summarized themes, while closed Likert items were reported descriptively as supplementary context. The analysis produced twenty-seven codes across the four levels plus enabling and hindering factors and recommendations. Findings show a positive reaction; gains in management knowledge and in leadership, communication, and problem-solving skills; transfer of learning into more systematic managerial practice; and organizational benefits perceived by supervisors, while subordinate readiness and organizational culture condition transfer.

ARTICLE INFO

Article History:

Received: 26 Jan 202226

Revised: 4 Jul 2026

Accepted: 14 Jul 2026

Publish online: 8 Aug 2026

Keywords:

Kirkpatrick model; police leadership education; Sespimma; training evaluation; transfer of training

Open access

Curricula: Journal of Curriculum Development is a peer-reviewed open-access journal.

ABSTRAK

Pendidikan kepemimpinan kepolisian merupakan investasi besar, namun efektivitasnya jarang dievaluasi di luar kepuasan peserta dan bukti mengenai pendidikan manajerial kepolisian berjenjang masih langka. Penelitian ini mengevaluasi Sekolah Staf dan Pimpinan Pertama (Sespimma) Polri menggunakan model Kirkpatrick empat level, dengan tujuan menilai reaksi, pembelajaran, perilaku, dan hasil, serta mengidentifikasi faktor yang mendukung dan menghambat penerapan hasil pendidikan. Penelitian ini menggunakan studi kasus kualitatif multi-sumber. Wawancara mendalam dengan seorang alumni, bersama observasi dan dokumen kinerja (LAKIP dan IKU), dikodekan secara tematik menggunakan ATLAS.ti 24 melalui pengkodean deduktif dan induktif; jawaban terbuka kuesioner dari 16 alumni dan 2 atasan dianalisis sebagai tema rangkuman, sedangkan butir Likert tertutup dilaporkan secara deskriptif sebagai konteks pendukung. Analisis menghasilkan 27 kode yang mencakup 4 level, ditambah faktor pendukung dan penghambat, serta rekomendasi. Temuan menunjukkan reaksi yang positif; peningkatan pengetahuan manajemen serta keterampilan kepemimpinan, komunikasi, dan pemecahan masalah; transfer pembelajaran ke praktik manajerial yang lebih sistematis; serta manfaat organisasi yang dipersepsikan oleh atasan, sementara kesiapan bawahan dan budaya organisasi mengondisikan transfer tersebut.

Kata Kunci: evaluasi pelatihan; model Kirkpatrick; pendidikan kepemimpinan kepolisian; Sespimma; transfer pembelajaran

How to cite (APA 7)

Wirawan, E., Fuad, N., & Matin, M. (2026). Evaluating Indonesia's first-level police staff and leadership education using the Kirkpatrick model. *Curricula: Journal of Curriculum Development*, 5(3), 975-988.

Peer review

This article has been peer-reviewed through the journal's standard double-blind peer review, where both the reviewers and authors are anonymised during review.

Copyright

2026, Eko Wirawan, Nurhattati Fuad, Matin. This an open-access is article distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) <https://creativecommons.org/licenses/by-sa/4.0/>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author, and source are credited.

*Corresponding author: wiraekowan@gmail.com

INTRODUCTION

Human resources are a central asset in law enforcement organizations, and their systematic development is decisive for a professional and accountable police service (Staller et al., 2022). Police work demands a broad set of trained knowledge, skills, and abilities, ranging from operational competence to the management of complex and potentially volatile public interactions (Bennell et al., 2022). Within the Kepolisian Negara Republik Indonesia (Polri), education and training therefore function as a strategic instrument for forming officers who are competent, adaptive, and capable of leadership. Leadership and management education occupies a special place in this effort because rigorous field experiments show that leadership training can improve leaders' behaviors and, under supportive conditions, organizational effectiveness (An & Meier, 2021; Jacobsen et al., 2022). Consequently, the value of leadership education is realized only when its effects reach on-the-job behavior and organizational outcomes, which makes thorough evaluation essential rather than optional.

One strategic tier of leadership and management education within Polri is the Sekolah Staf dan Pimpinan Pertama (Sespimma), which prepares professional officers of integrity to assume managerial positions in regional units, as documented by the Lembaga Pendidikan dan Pelatihan Kepolisian Negara Republik Indonesia (Lemdiklat Polri) in its 18 February 2025 announcement of the opening of the 73rd Sespimma cohort. Because its alumni are expected to lead frontline units, the quality of education at this tier escalates into the quality of leadership and public service, and the substantial investment it requires demands accountability through a comprehensive evaluation. A comprehensive evaluation must extend beyond participant satisfaction to changes in on-the-job behavior and organizational impact. In practice, however, training evaluation tends to stop at the reaction and learning levels. In contrast, the behavior and results levels are the least frequently assessed systematically, so the real impact of programs often remains unknown (Ambu-Saidi et al., 2024).

One of the most widely used frameworks for evaluating training in a tiered manner is the four-level model developed by Kirkpatrick, which assesses programs at the levels of reaction, learning, behavior, and results. This four-level structure follows the foundational framework introduced by Donald L. Kirkpatrick and James D. Kirkpatrick in their classic book *Evaluating Training Programs: The Four Levels*, published in 2006. The framework is relevant for evaluating Sespimma because it allows a coherent assessment of the participants' learning experience and its impact on unit performance. Despite this practical importance, scholarly knowledge of police leadership remains limited, predominantly leader-centric, and insufficiently attentive to the operational context (Filstad et al., 2025). Recent syntheses likewise find that police education and training remain under-researched, with the design, delivery, and evaluation of instruction identified as persistent blind spots (Frade & Walsh, 2025).

Tiered managerial leadership education in the police has rarely been the focus of a full four-level evaluation, and most existing evaluations rely on quantitative instruments or self-report data, leaving the behavior and results levels under-examined. This constitutes a clear contextual and methodological gap. Accordingly, the novelty of this study lies in evaluating

Polri's first-level leadership and management education (Sespimma) across all four Kirkpatrick levels through a qualitative approach with methodological triangulation, namely in-depth interviews, observation, and document study. To the author's knowledge, a Sespimma evaluation with this scope and approach has not been found in prior research. In line with this gap, the research questions are: (a) how do participants react to the delivery of Sespimma education, (b) what are the participants' learning outcomes, (c) how does the on-the-job behavior of alumni change, (d) what is the program's impact on organizational performance, and (e) what factors support and hinder the application of the education. Correspondingly, this study aims to evaluate the Sespimma program using the Kirkpatrick model at the reaction, learning, behavior, and results levels, and to identify the supporting and hindering factors together with recommendations for improvement.

LITERATURE REVIEW

Training Evaluation as a Function of Human Resource Development

Training and development is a core function of human resource development aimed at building the knowledge, skills, and attitudes that improve individual and organizational performance (Bilderback & Miller, 2023). Within this tradition, training evaluation is the systematic process of judging whether a program achieves its goals and grounding decisions for improvement; it also serves to strengthen participants' professionalism and support continuous quality enhancement (Sari & Solichin, 2023). Various evaluation models have been proposed that differ in scope and emphasis, so the choice of model must fit the purpose and context of the program being assessed; tiered, comprehensive models are valued precisely because they connect the immediate training experience to its downstream effects on behavior and organizational results.

This conceptualization of learning outcomes builds upon the foundational framework introduced by Kurt Kraiger, J. Kevin Ford, and Eduardo Salas in 1993, which implies that training evaluations must go beyond measuring knowledge alone to capture cognitive, skill-based, and affective results. This multidimensional view underpins tiered evaluation models that separate participants' immediate reactions from deeper learning and from downstream changes in behavior and organizational results. Because these domains do not move in lockstep, an increase in knowledge does not guarantee a change in on-the-job behavior, so an evaluation that stops at the reaction or learning level can overstate a program's real value (Kraiger et al., 1993).

Transfer of Training and Its Conditions

A central concern in evaluating leadership and management education is whether learning is actually transferred to work. This focus on the transfer of training builds on the seminal review introduced by Timothy T. Baldwin and J. Kevin Ford in 1988, which established that applying and maintaining learned knowledge and skills on the job is a foundational concern that remains only partially understood despite decades of research. Because such transfer does not occur automatically, evaluation that reaches the behavior and results levels is essential if the benefits of leadership education are to be assessed fully rather than stopping at satisfaction. Trainee characteristics, training design, and the work environment,

particularly manager and supervisor support, are established predictors of transfer (de Jong et al., 2023).

Recent studies further confirm that training design, participant motivation, and the work environment shape transfer and that leadership support helps determine whether learning genuinely becomes workplace practice (Hillberg Jarl, 2024; Nafukho et al., 2023). Taken together, because transfer does not occur automatically and its benefits for individuals, teams, and organizations depend on genuine application at work, a valid assessment of tiered leadership education must complement participant perception with observable evidence, such as work-behavior observation and organizational performance documents, so that claims at the behavior and results levels are anchored in more than self-report.

Leadership and Managerial Development and the Kirkpatrick Model

Leadership and management education forms a distinct strand of human resource development whose value is realized only when it changes how leaders act and how their units perform. Field-experimental evidence indicates that leadership training can improve leaders' behavior and, under supportive conditions, organizational effectiveness (An & Meier, 2021; Jacobsen et al., 2022). This makes higher-level evaluation especially important for managerial education, because assessing satisfaction or knowledge alone cannot show whether newly trained leaders actually lead differently or whether their units benefit. The Kirkpatrick four-level model is widely used to structure this kind of evaluation because it connects participants' reactions and learning to subsequent changes in behavior and organizational results, as set out by Donald L. Kirkpatrick and James D.

Kirkpatrick in their classic book *"Evaluating Training Programs: The Four Levels; its New World"* refinement, presented by James D. Kirkpatrick and Wendy Kayser Kirkpatrick in *"Kirkpatrick's Four Levels of Training Evaluation,"* urges evaluators to define the expected results first and design the assessment backward from them. In practice, however, evaluations of leadership and managerial programs frequently stop at the reaction and learning levels. Hence, the behavior and results levels remain the least assessed, and the downstream value of such education stays under-evidenced (Ambu-Saidi et al., 2024). The model's applicability nonetheless extends across sectors, from health-professions education to public-service training (Cordon et al., 2025; Gupta et al., 2025; Jefnee et al., 2025). This gap is the point of departure for the present evaluation of Sespimma.

Leadership and Culture in Police Organizations

Because the object of this evaluation is leadership education inside a police institution, the characteristics of police leadership and organizational culture form an essential part of its literature base. Scholarship on police leadership remains comparatively thin and leader-centric. However, it consistently stresses that leadership in policing is exercised within a distinctive occupational culture rather than in a neutral setting (Filstad et al., 2025). That culture is commonly described through a tension between a *"management"* orientation and a *"street"* orientation, and it can be relatively durable, transmitting shared values, routines, and a preference for established ways of working that shape how reforms and new practices

are received; comparable challenges are reported in contemporary police education reform (Densley et al., 2025; Triola & Chanin, 2023). These cultural features matter directly for the transfer of leadership education because an organizational culture oriented toward established routines can slow the adoption of the managerial practices that such education promotes. Framing police leadership education against this cultural backdrop, therefore, helps explain why favorable learning does not automatically become changed behavior, and it provides the conceptual foundation for interpreting the enabling and hindering factors identified in this study.

METHODS

This study uses a qualitative approach with an evaluative, multi-source case study design to understand in depth the delivery and impact of the Sespimma program from the perspective of alums and their supervisors (Mtisi, 2022). The aim is analytic rather than statistical generalization: the case is used to examine how the four-level logic operates in a real managerial setting, not to produce population estimates. The evaluation is structured by Kirkpatrick's four-level framework, which assesses a program hierarchically at the levels of reaction, learning, behavior, and results, with assessment at a higher level presupposing achievement at the preceding one; in its New World refinement, the model emphasizes the relevance of training to participants' real tasks and designing evaluation backward from the expected results.

Consistent with this qualitative scope, findings at the behavior and results levels are treated as indicative and are triangulated with documentary evidence rather than presented as independently confirmed impact. The research flow proceeds through several stages: developing instruments based on the four Kirkpatrick levels; collecting data through an in-depth interview, a supplementary questionnaire for alumni and supervisors, observation, and document study; transcribing, anonymizing, and preparing the data; thematic coding and analysis; checking trustworthiness; and drawing evaluative conclusions at each level together with the enabling and hindering factors and recommendations. The flow is summarized in **Figure 1**.



Figure 1. Research Flow of the Sespimma Program Evaluation
Source: The author (2026)

Participants were selected purposely as information-rich sources, based on direct involvement in and understanding of the program being evaluated (Campbell et al., 2020). To make the status of each source explicit, the study uses one coded source and two supplementary sources: an in-depth interview with one alumnus, which is the only source entered into ATLAS.ti and the basis of the 27 codes; a questionnaire completed by sixteen alumni addressing the reaction and learning levels; and a questionnaire completed by two direct supervisors of alumni addressing the behavior and results levels. The coded interview involved a Sespimma alumnus who had completed the education and served in a managerial position in a regional unit, providing depth on all four levels. The sixteen alumni questionnaires added breadth on reaction and learning together with barriers, needed improvements, and recommendations. In contrast, the two supervisor questionnaires

provided an independent perspective on the behavior and results levels, which are least accessible to alumni self-report.

The two questionnaire groups were not entered into ATLAS.ti; their open-ended answers were summarized thematically, and their closed Likert items were reported descriptively. The single coded interview and the small number of supervisors are clear boundaries on the claims, addressed openly in the limitations. All identities are anonymized to preserve confidentiality in accordance with the research ethics agreement, and participation was voluntary and based on informed consent. As a case study, this research relies on multiple sources of evidence to provide a more complete understanding of the phenomenon (Mtisi, 2022). Data were collected through four complementary techniques. First, an in-depth semi-structured interview with the alumnus used a guide organized around the four Kirkpatrick levels with probing questions; this interview is the source that was formally coded in ATLAS.ti.

Second, a questionnaire was administered to alumni and to supervisors as a supplementary source; its open-ended items were analyzed as summarized qualitative themes, while its closed Likert items were tabulated as frequencies and used only as descriptive context, not as inferential evidence, given the marked ceiling effect in the responses and the small number of supervisor respondents. Third, observation of the work environment and behavior used an observation guide. Fourth, the document study examined unit performance documents, namely the Government Agency Performance Report (LAKIP) and the Key Performance Indicators (IKU); these documents function as the more objective and verifiable anchor for the behavior and results levels, complementing perception through documentary triangulation. The in-depth interview transcript, the observation notes, and the performance documents were analyzed thematically with the aid of ATLAS.ti 24, following an established approach to reflexive thematic analysis (Byrne, 2022). These sources were loaded as primary documents, and relevant quotations were coded.

Coding proceeded in two directions: deductive coding used the four Kirkpatrick levels as code groups, while inductive coding, including in-vivo codes, captured categories emerging directly from the informant's expressions. This yielded 27 codes organized into six code groups, whose structure is displayed in **Figure 2**. The open-ended questionnaire responses from the sixteen alumni and the two supervisors were not entered into ATLAS.ti; they were analyzed separately as summarized qualitative themes and used to corroborate and extend the interview-based codes, while the closed questionnaire items were tabulated as frequencies for descriptive context only. The analysis moved iteratively between coding, theme development, and verification until the codes and themes were stable. Trustworthiness was maintained through the criteria of credibility, transferability, dependability, and confirmability; credibility was strengthened through source and technique triangulation, while dependability and confirmability were preserved through an audit trail of the coding process and analytic memos in ATLAS.ti 24 (Ahmed, 2024).

RESULTS AND DISCUSSION

The ATLAS.ti thematic analysis of the in-depth interviews, observation notes, and documents produced 27 codes grouped into six code groups corresponding to the four Kirkpatrick levels,

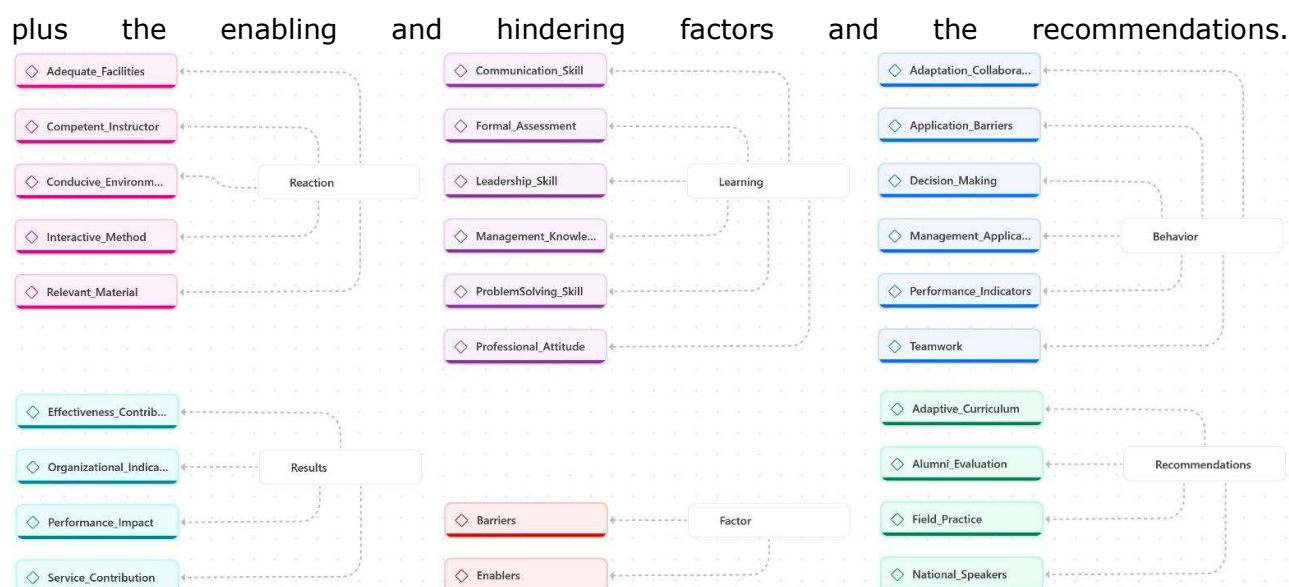


Figure 2. Code Structure of the ATLAS.ti Analysis of the Sespimma Program Evaluation
Source: The author's coding results (2026)

The code structure is visualized in **Figure 2**, while the detailed findings for each code are presented in **Table 1**. The open-ended questionnaire responses from the sixteen alumni and the two supervisors were then examined as summarized themes; they largely reinforced the interview-based codes and are woven into the narrative below. As supplementary context only, the closed questionnaire items showed almost uniform agreement: among the sixteen alumni, agreement or strong agreement was reached about 98.7% on the reaction items and 99.3% on the learning items, while both supervisors agreed or strongly agreed on almost all behavior and results items. This near-uniformity indicates a marked ceiling effect, so these figures are read only as broad corroboration and not as measured differences; the substantive findings below rest on the qualitative material.

Table 1. Coding Results of the Sespimma Program Evaluation Based on the Kirkpatrick Model

Code (Figure 2)	Main Finding	Keywords
REACT_Relevant_Material	Material is relevant to job needs and up to date	Task-relevant · Leadership-management · Up-to-date
REACT_Interactive_Method	Interactive method with discussion and technology	Interactive · Case-discussion · Technology
REACT_Compentent_Instructors	Instructors master the material, supported by expert speakers	Subject-mastery · Expert-speakers · Competent
REACT_Adequate_Facilities	Adequate facilities support the learning process	Dormitory · Wifi-internet · Learning-media
REACT_Conducive_Environment	Conducive learning environment and orderly schedule	Conducive · Pleasant · Structured-schedule
LEARN_Management_Knowledge	Police management knowledge increased markedly	Police-management · Increased · Finance-logistics-HR

Code (Figure 2)	Main Finding	Keywords
LEARN_Leadership_Skill	Leadership skills improved	Leadership · Improved · Cadet-senate
LEARN_ProblemSolving_Skill	Problem-solving developed at a higher level	Problem-solving · NAC-BCS · SKK
LEARN_Communication_Skill	Communication skills improved	Communication · NKP · KKP-FGD
LEARN_Professional_Attitude	Professionalism and motivation strengthened	Professionalism · Improved · Rigorous-selection
LEARN_Formal_Assessment	Formal assessment shows competency gains	Pre-post-test · NKP-Taskap · Physical-test
BEHAV_Management_Application	Applies management principles (POAC) and SWOT at work	POAC · SWOT · Application
BEHAV_Decision_Making	More systematic and multi-aspect decision making	Decision · Systematic · Multi-aspect
BEHAV_Teamwork	Improved teamwork and member empowerment	Teamwork · Team · Empowerment
BEHAV_Adaptation_Collaboration	Adaptation to change and cross-functional collaboration	Adaptation · Flexible · Cross-functional
BEHAV_Performance_Indicators	More planned execution and orderly reporting	Well-planned · Orderly-reporting · LAKIP
BEHAV_Application_Barriers	Barriers from subordinate and environment readiness	Subordinates · Not-yet-adaptive · Barrier
RESULT_Performance_Impact	Positive impact: orderly work, alumni as drivers	Positive-impact · Orderly · Driver
RESULT_Service_Contribution	More professional and humane public service	Public-service · Humane · Discipline
RESULT_Effectiveness_Contribution	Improved effectiveness and work coordination	Effectiveness · Coordination · POAC
RESULT_Organizational_Indicators	Higher target achievement and service quality	Target-achievement · Service-quality · IKU
FACTOR_Enablers	Leadership support, positive culture, facilities/technology	Leadership-support · Positive-culture · Facilities-technology
FACTOR_Barriers	Resistance to change, limited facilities, old culture	Resistance · Limited-facilities · Old-culture
RECOM_Adaptive_Curriculum	Updated material and an adaptive curriculum	Global-material · Cyber-AI · Adaptive
RECOM_Field_Practice	More realistic field practice and simulation	Practice · Simulation · Realistic
RECOM_National_Speakers	National and ministry-level speakers	Speakers · National · Ministry
RECOM_Alumni_Evaluation	Continuous alumni evaluation and external collaboration	Alumni-evaluation · Collaboration · External

Source: the author's coding results (2026). Codes follow the labels in Figure 2

At the reaction level, the alumni assessed the delivery of the education positively across all aspects: the material was relevant and up to date, the method was interactive, the instructors were competent and reinforced by expert speakers, the facilities were adequate, and the learning environment was conducive, with a recurring note being a wish for more national-level and external speakers connected to police duties. At the learning level, alumni reported that their knowledge of police management increased and that leadership, problem-solving, and communication skills improved, with professionalism strengthened; the interview additionally linked these gains to formal assessment, such as the difference between pre-test and post-test results. At the behavior level, drawn primarily from the two supervisors, alumni were reported to apply management principles and structured analysis at work, to make more systematic decisions, to work in teams, and to adapt and collaborate across functions, with more planned execution and more orderly reporting; both supervisors, however, also indicated that application is not always smooth, consistent with alumni statements that field conditions are not as ideal as theory and that some subordinates and organizational routines are not yet adaptive.

At the results level, likewise from the supervisors and supported by the LAKIP and IKU documents, alumni were seen to contribute to more orderly work, more professional public service, better coordination, and organizational performance; because this rests on only two supervisors, it is reported as indicative rather than confirmed. Regarding the fifth question, the enabling factors were leadership support, a positive culture, and adequate facilities and technology. In contrast, the hindering factors were resistance to change, limited facilities, and persistent old work patterns and differing educational backgrounds. The recommendations that recurred across sources were an adaptive curriculum responsive to technology and digitalization, more realistic field practice and simulation, more national-level and external speakers, and continuous alumni evaluation; several respondents additionally called for strengthening integrity in the learning process, which is reported here neutrally as an emergent point for improvement.

Discussion

The uniformly positive reaction is best interpreted as the program satisfying the precondition for higher-level effects rather than as proof of effectiveness in itself. This reading is consistent with public-sector evidence that trainees tend to rate Kirkpatrick-based training positively across all four levels, with learning emerging as especially salient. However, the present case sharpens that point by treating a favorable reaction as an enabler of, not a guarantee for, transfer (Islam et al., 2024). That the learning gains were corroborated by pre-test and post-test differences moves the learning claim beyond perception, echoing evidence that Kirkpatrick-based evaluation can capture genuine competency change in formal training and in non-formal settings where assessment grounds improvement (Iqbal et al., 2021; Wuryanto et al., 2026). The interpretive implication is that, where perception data cluster near the ceiling, formal assessment functions as the most defensible layer of evidence at these lower levels.

The reported application of POAC and SWOT and the more systematic decision-making are best understood as behavioral transfer rather than mere learning. They align with cross-context evidence that four-level evaluation can detect changes in professional practice at the

behavior level, including when behavior is assessed through direct observation (Liu et al., 2025; Ross et al., 2022). Crucially, the readiness barriers among subordinates indicate that transfer is co-produced by the receiving unit rather than determined by the education alone. The present study addresses this by drawing on behavioral evidence from supervisors rather than from alumni alone, echoing the multi-source logic that combines participant and supervisor perspectives in leadership evaluation; the warrant nonetheless remains limited by the small number of supervisors and would be strengthened by a larger group (Alsalamah & Callinan, 2021).

At the results level, the perceived organizational impact must be read with caution, because it rests on only two supervisors and because perceptions of leadership can diverge across observers, as field-experimental work on aligning perceptions of leadership has shown (An et al., 2022). The documented indicators referenced in the LAKIP and IKU are consequently a more credible warrant than perception alone, and the supervisors' account of impact is best treated as a hypothesis to be tested with a larger group of supervisors and independent sources. Reading the fifth research question through the transfer lens clarifies why the same program can yield uneven results. The enabling factors reported here, namely leadership support, a positive unit culture, and adequate facilities and technology, and the hindering factors, namely resistance to change, limited facilities, and entrenched work patterns, are precisely the receiving-environment conditions that determine whether trained behavior survives contact with the workplace. This aligns with police-specific evidence that resistance to change and an entrenched organizational culture are recurrent obstacles to reform in police settings (Colombo et al., 2023). This interpretation reframes the recommendations as levers on those conditions rather than as adjustments to course content alone: an adaptive, technology-aware curriculum and more realistic field practice target the readiness of the trainee, while continuous alumni evaluation and closer supervisory engagement target the readiness of the unit.

The multi-source logic that pairs participant and supervisor perspectives is therefore not only a measurement choice but also a practical route to sustaining transfer, because the supervisors who assess behavior are also the actors who enable it (Alsalamah & Callinan, 2021). Synthesizing across the four levels, the case yields a modest but specific new finding. In the studied setting, the value of Sespimma is realized only when the receiving environment is ready, so that impact is co-produced rather than delivered by training content. This reframes the enabling and hindering factors as the decisive condition for transfer and directly motivates the recommendations, especially continuous alumni evaluation and stronger field practice. Several limitations must temper the claims. First, the coded qualitative analysis rests on a single in-depth interview, so the 27-code structure reflects one alumnus's account; the sixteen alumni and two supervisor questionnaires corroborate and extend it but were not independently re-coded in ATLAS.ti.

Second, the behavior and results levels rest on only two supervisors, so they are reported as indicative rather than confirmed. They are triangulated with the LAKIP and IKU documents as a more objective anchor. Third, the closed questionnaire responses cluster near the ceiling, a pattern consistent with acquiescence and social-desirability bias in a hierarchical setting, which is why they are used only as descriptive context and not as measured evidence. Fourth, the design is cross-sectional, and the alumni came from mixed cohorts, so it captures perceptions at one point rather than change over time. Accordingly, the study

makes no claim to statistical generalization. Future studies should code a broader set of informants in ATLAS.ti, broaden the supervisor base in particular, add pre-test and post-test or documentary measures for the behavior and results levels, and, where possible, adopt comparative, multi-cohort, or longitudinal designs.

CONCLUSION

This study set out to evaluate Sespimma against the four Kirkpatrick levels and to identify the factors that condition the application of the education. Taken as a whole, the evaluation indicates that, in the studied case, the program meets its aims: participants judge its delivery favorably, report meaningful gains in management knowledge and in leadership, communication, and problem-solving skills, and, in their supervisors' assessment, carry these into more systematic managerial practice and perceived organizational benefit. The decisive qualification concerns the fifth question: these benefits materialize only where the receiving unit is ready, so the enabling and hindering factors, rather than the training content alone, determine whether learning becomes results. This answers the five research questions and meets the objective of a full four-level qualitative evaluation, with the qualification that the behavior and results levels are indicative because they rest on a single coded interview and a small number of supervisors.

As suggested, future research should extend the evidence base rather than the interpretation. The priorities are to broaden and to formally code multiple informants, namely alumni, supervisors, and program managers, so that the behavior and results levels no longer rest on a single coded source; to incorporate systematic analysis of performance documents together with pre-test and post-test or other objective measures; and to adopt comparative, multi-cohort, or longitudinal designs. Such work would test whether the patterns observed here hold beyond a single case and would help fill the still-limited evidence on police leadership education at the transfer and results levels.

AUTHOR'S NOTE

The author declares that there is no conflict of interest regarding the publication of this article and confirms that the data and content of this article are free from plagiarism. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors; it was conducted independently and self-funded by the author. The author thanks the key informant for their participation and the Sespimma unit for facilitating the study.

REFERENCES

- Ahmed, S. K. (2024). The pillars of trustworthiness in qualitative research. *Journal of Medicine, Surgery, and Public Health*, 2(1), 1-4.
- Alsalamah, A., & Callinan, C. (2021). Adaptation of Kirkpatrick's four-level model of training criteria to evaluate training programmes for head teachers. *Education Sciences*, 11(3), 116.

- Ambu-Saidi, B., Fung, C.Y., Turner, K., & Lim, A.S.S. (2024). A critical review on training evaluation models: A search for future agenda. *Cognition and Interactive Technologies*, 10(1), 142–170
- An, S. H., & Meier, K. J. (2021). Gender and the effectiveness of leadership training: results from a field experiment. *Review of Public Personnel Administration*, 41(4), 747–770.
- An, S. H., Jensen, U. T., Bro, L. L., Andersen, L. B., Ladenburg, J., Meier, K. J., & Salomonsen, H. H. (2022). Seeing eye to eye: can leadership training align perceptions of leadership?. *International Public Management Journal*, 25(1), 2-23.
- Bennell, C., Jenkins, B., Blaskovits, B., Semple, T., Khanizadeh, A. J., Brown, A. S., & Jones, N. J. (2022). Knowledge, skills, and abilities for managing potentially volatile police–public interactions: a narrative review. *Frontiers in Psychology*, 13(1), 1-16.
- Bilderback, S. L., & Miller, G. J. (2023). Importance of employee development programs in business. *Journal of Management Development*, 42(4), 327-336.
- Byrne, D. (2022). A worked example of Braun and Clarke’s approach to reflexive thematic analysis. *Quality & Quantity*, 56(3), 1391-1412.
- Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S. & Walker, K. (2020). Purposive sampling: complex or simple? Research case examples. *Journal of Research in Nursing*, 25(8), 652-661.
- Colombo, L., Acquadro Maran, D., & Grandi, A. (2023). Attitudes toward organizational change and their association with exhaustion in a sample of Italian police workers. *Frontiers in Psychology*, 14(1), 1-10.
- Cordon, C., Dlugosz, D., Lopez, L., Gelacio, R., Smith-Eivemark, K., Maier, S. & Irwin, T. (2025). Supporting novice nurses’ transition to independent practice: evaluation of the TIPS program using the Kirkpatrick model. *Nursing Reports*, 15(2), 50, 1-26.
- de Jong, B., Jansen in de Wal, J., Cornelissen, F., van der Lans, R., & Peetsma, T. (2023). How to predict transfer of training? Investigating the application of the unified model of task-specific motivation. *International Journal of Training and Development*, 27(2), 242-262.
- Densley, J., MacLean, C. E., & Hilal, S. (2025). Educating and training culturally competent peace officers: challenges and opportunities of the Minnesota model. *Multicultural Learning and Teaching*, 20(2), 183-203.
- Filstad, C., Karp, T., & Rød, A. (2025). What do we know about police leadership? A review of the current status of police leadership research and practice, with suggestions for future research directions. *Policing and Society*, 35(3), 280-293.
- Frade, B., & Walsh, J. P. (2025). Police education and training: a scoping review of the research literature. *The Police Journal*, 1(1), 1-10.
- Gupta, S., Sathe, S., Arya, K., Mitra, B., & Mangeshkar, A. (2025). Evaluation of an integrated teaching approach in physiology, anatomy, and biochemistry for phase I MBBS students using the Kirkpatrick model. *Cureus*, 17(7), 1-9.
- Hillberg Jarl, F. (2024). The impact of leadership on the workplace learning of individuals and teams: a literature review and synthesis. *The Learning Organization*, 31(3), 411-439.

- Islam, M. Z., Abdul Kader Jilani, M. M., & Karim, M. R. (2024). Enhancing post-training evaluation of annual performance agreement training: a fusion of fsQCA and artificial neural network approach. *PloS one*, 19(1), 1-29.
- Iqbal, Z., Anees, M., Khan, R., Wadood, A., & Malik, S. (2021). A comparative analysis of the efficacy of three program-evaluation models–A review on their implication in educational programs. *Humanities & Social Sciences Reviews*, 9(3), 326-336.
- Jacobsen, C. B., Andersen, L. B., Bøllingtoft, A., & Eriksen, T. L. M. (2022). Can leadership training improve organizational effectiveness? Evidence from a randomized field experiment on transformational and transactional leadership. *Public Administration Review*, 82(1), 117-131.
- Jefnee, M. A. I., Kaur, M., Teo, C. H., Goh, S. S. L., Lai, P. S. M., Selvaraj, C. S., & Ramdhan, S. N. (2025). Effectiveness of online asthma training programmes to improve asthma management among school staff: a systematic review using the Kirkpatrick evaluation model. *NPJ Primary Care Respiratory Medicine*, 35(1), 45, 1-10.
- Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, 78(2), 311-328.
- Liu, T., Ren, Y., & Wang, X. (2025). Evaluation of the effectiveness of standardized training for new nurses using the Kirkpatrick model: an observational study. *Medicine*, 104(1), 1-8.
- Mtisi, S. (2022). The qualitative case study research strategy as applied on a rural enterprise development doctoral research project. *International Journal of Qualitative Methods*, 21(4), 1-13.
- Nafukho, F. M., Irby, B. J., Pashmforoosh, R., Lara-Alecio, R., Tong, F., Lockhart, M. E. & Wang, Z. (2023). Training design in mediating the relationship of participants' motivation, work environment, and transfer of learning. *European Journal of Training and Development*, 47(10), 112-132.
- Ross, B., Penkunas, M. J., Maher, D., Certain, E., & Launois, P. (2022). Evaluating results of the implementation research MOOC using Kirkpatrick's four-level model: a cross-sectional mixed-methods study. *BMJ open*, 12(5), 1-8.
- Sari, V. B. M., & Solichin, M. R. (2023). Evaluasi model Kirkpatrick pada program Praktek Pembelajaran Lapangan (PPL) sebagai penguatan profesionalisme mahasiswa keguruan. *Jurnal Ekonomi Pendidikan dan Kewirausahaan*, 11(1), 37-50.
- Staller, M. S., Koerner, S., Bennell, C., & Suss, J. (2022). Police education and training revisited: drawbacks and advances. *Frontiers in Psychology*, 13(1), 1-2.
- Triola, A. M., & Chanin, J. (2023). Police culture, transparency and civilian oversight: a case study of the National City Police Department. *International Journal of Police Science & Management*, 25(1), 81-95.
- Wuryanto, J., Sutisna, A., & Supriyati Saefudin, Y. (2026). Evaluating Indonesia's vocational centre of excellence program with the Kirkpatrick model: linking facilities, resources, and student outcomes. *International Research Journal of Multidisciplinary Scope*, 7(1), 679-697.