DIJMS: Dinasti International Journal of Management Science
E-ISSN: 2686-522X
P-ISSN: 2686-5211
https://dinastipub.org/DIJMS dinasti.info@gmail.com +62 811 7404 455
DOI: https://doi.org/10.38035/dijms.v7i5.6826

Assessing The Adequacy of the Kirkpatrick Model in Evaluating The Performance of Public Sector Apparatus Training: A Post-Learning Evaluative Study of Distance Learning Training at The Tax Education and Training Center

Dito Mahar Putro[1], Iwan Kurniawan Subagja[2], Azis Hakim[3]
[1]Universitas Krisnadwipayana, Jakarta, Indonesia, email. ditomaharpu@gmail.com
[2]Universitas Krisnadwipayana, Jakarta, Indonesia, email. iwankurniawan@unkris.ac.id
[3]Universitas Krisnadwipayana, Jakarta, Indonesia, email. dr_azishakim@yahoo.co.id

Corresponding Author: ditomaharpu@gmail.com

Abstract: This study aims to assess the adequacy of the Kirkpatrick Model in evaluating the performance of distance learning (PJJ)-based public sector apparatus training through a post-learning evaluation approach. The research focuses on four distance learning (PJJ) programs organized by the Tax Education and Training Center (Pusdiklat Pajak). This study uses a model-oriented evaluation approach with an evaluative case study design. The data used are sourced from Level 3 and Level 4 Post-Learning Evaluation (Epaspem) reports, which are analyzed using mixed methods with a limited explanatory approach. Quantitative analysis is used to assess the empirical performance of the Kirkpatrick Model at Level 3 (behavior) and Level 4 (outcomes), while qualitative analysis is conducted to examine the model's limitations in representing evaluation results in the context of distance learning and the public sector. The results show that the Kirkpatrick Model is able to identify changes in work behavior and provide an overview of the short-term work outcomes of alumni in the four distance learning programs. However, this model still has limitations, especially in explaining impact attribution and in representing organizational performance more comprehensively.

Keyword: Kirkpatrick Model, Training Evaluation, Distance Learning, Public Sector Apparatus, Post-Learning Evaluation.

INTRODUCTION

Developing the competency of tax officials is a strategic factor in supporting the performance of public sector organizations, particularly in optimizing state revenue and strengthening tax governance. Within the Ministry of Finance, this function is implemented through education and training units under the Financial Education and Training Agency, including the Tax Education and Training Center (Pusdiklat Pajak). Pursuant to Minister of Finance Regulation Number 124 of 2024 concerning the Organization and Work Procedures of the Ministry of Finance, the Tax Education and Training Center is mandated to provide technical training and knowledge management in taxation to support the implementation of the duties of the Directorate General of Taxes and related stakeholders in the taxation sector, as well as to support the achievement of overall organizational performance.

The implementation of this regulation demonstrates that tax training serves not only as a means of enhancing individual capacity but also as an integral part of the organization's performance management system. Government officials are valuable and potential assets of any government institution. Therefore, providing appropriate competency development not only increases efficiency but also increases productivity (Aziz, Widodo, Subagja: 2021). Therefore, by using appropriate measurement tools, training can be assessed not only for its effectiveness in the learning process but also for its ability to have a tangible impact on employee work behavior and organizational performance. In this context, training evaluation is a crucial tool for ensuring that allocated resource investments generate strategic benefits for public sector organizations.

Kirkpatrick is one of the most widely used training evaluation frameworks in civil servant evaluation practices. Various studies have shown this model to be quite effective in describing changes in work behavior and post-training outcomes. Research by Ritonga, Saepudin, and Wahyudin (2019), for example, demonstrated that applying the Kirkpatrick Model at Levels 3 and 4 was able to identify the transfer of learning to the workplace and its impact on organizational performance. Similar findings were also presented by Khalimi and Sihombing (2021), who demonstrated that evaluations at both levels could describe changes in alumni work behavior and indicate improvements in organizational performance.

Furthermore, a study by Syafitri and Tumanggor (2023) on the evaluation of disaster training for civil servants also confirmed that the Kirkpatrick Model is effective in assessing various aspects of training, from reactions to behavioral changes. However, the study also revealed that the impact at Level 4 is still limited and is heavily influenced by the organizational context and the participants' post-training assignments. This suggests that success at the individual level does not always translate directly to improved overall organizational performance. In the government sector, education and training for government officials are strategic instruments for improving the competence of civil servant resources and supporting the achievement of individual and public sector organizational performance. Therefore, its effectiveness needs to be evaluated and measured through a comprehensive and sustainable evaluation system (Engriyani & Rugaiyah, 2022). Implementing training without follow-up with post-training evaluation has the potential to result in the training being merely administrative in nature, thus not having a real impact on changing individual behavior or organizational performance (Handoko et al., 2022). Beyond its benefits for individuals, in the context of government organizations, the Kirkpatrick Model can be relevant because it provides a holistic reflection of the relationship between the training provided and changes in individual civil servant behavior and their contribution to the organizational performance of their work units (Azmy & Setiarini, 2023).

Most previous studies still use the Kirkpatrick Model normatively and operationally, thus failing to fully critically assess its adequacy. In practice, Level 4 evaluations are often represented solely through individual output or the perceptions of superiors and colleagues, while the complexity of public sector organizational performance—which is influenced by policies, systems, and the work environment—has not been fully accommodated. In fact, several studies have shown that changes in work behavior identified through this model cannot always be directly attributed as the primary cause of improved organizational performance (Ramadhon, 2015; Handalani & Soesanto, 2022).

Although the Kirkpatrick Model of evaluation is known for its ease of use due to its simplicity, systematic nature, and ability to describe a cycle from process to training outcomes through four levels of assessment or evaluation: reaction, learning, behavior, and outcomes (Engriyani & Rugaiyah, 2022), most previous research indicates that evaluation tends to be predominantly focused on levels 1 and 2, namely reaction and learning, while learning evaluation, namely levels 3 and 4, namely evaluation at the behavioral and organizational outcome levels, has not been optimally implemented (Engriyani & Rugaiyah, 2022; Handoko et al., 2022).

In the context of tax apparatus training conducted by the Tax Education and Training Center, these limitations become even more relevant. The performance of tax organizations is influenced by various factors, both structural and external, such as fiscal policy, state revenue targets, tax information systems, and taxpayer compliance levels. Therefore, using a single evaluation model has the potential to oversimplify the contribution of training to organizational performance.

Based on this, this study does not focus on evaluating the success of a particular training program, but rather on assessing the adequacy of the Kirkpatrick Model in evaluating the performance of public sector apparatus training, specifically in the context of tax training at the Tax Education and Training Center. This study emphasizes the ability of the Kirkpatrick Model Level 3 and Level 4 in representing the impact of training on work behavior and organizational performance, as well as identifying the strengths and limitations of the model as a post-learning evaluation instrument. The results of the study are expected to provide methodological contributions in the development of apparatus training evaluation that is more critical, contextual, and relevant to the needs of the public sector, especially in supporting.

METHOD

This research is an evaluative study using a model-oriented evaluation approach. The primary focus of the study is to assess the adequacy of the Kirkpatrick Model in evaluating the training performance of public sector officials, particularly Distance Learning (PJJ)-based training. Therefore, this study is not intended to test the pedagogical effectiveness of PJJ, but rather to evaluate the extent to which the Kirkpatrick Model is able to represent changes in work behavior and organizational outcomes in the PJJ context.

The research approach uses a mixed methods with a limited explanatory design, combining quantitative data from post-learning evaluations with reflective qualitative analysis of the limitations of the evaluation model. Creswell and Plano Clark (2018) state that a mixed methods approach is highly relevant in evaluation research because it allows researchers to explain quantitative findings through a contextual understanding of qualitative data, especially when the data is sourced from existing evaluation documents.

The research design used is an evaluative case study. Yin (2018) explains that an evaluative case study is appropriate when the research aims to examine the application of a model or policy in a real-world context with clear boundaries. In this study:

- Case study: application of the Kirkpatrick Model in post-learning evaluation of distance learning (PJJ),
- Context: distance learning-based public sector personnel training,
- Evaluation focus: adequacy of the Kirkpatrick Model Level 3 (Behavior) and Level 4 (Results).

The case study is limited to four (4) distance learning (PJJ) courses, as listed in the uploaded evaluation report, to maintain depth of analysis while controlling the breadth of discussion.

The research object is the Kirkpatrick Level 3 and Level 4 Models used in post-learning evaluations of four distance learning programs throughout 2025, as follows:

- Tax Bailiff II Distance Learning Program for Fiscal Year 2025,
- Geographic Information Systems (GIS) Distance Learning Program for 2025,
- Tax Analysis and Taxpayer Working Group (KKAWP Group) Distance Learning Program for 2025,

- Regional Control with Internet Media Optimization (PWDOMI) Distance Learning Program for 2025.

These four distance learning programs were selected because they had already undergone Level 4 post-learning evaluations and had relatively uniform distance learning characteristics in terms of evaluation design. The research subjects included:

- alumni staff,
- alumni direct supervisors,
- alumni peers/colleagues.

This multi-source involvement aligns with the recommendations of Kirkpatrick and Kirkpatrick (2007), who emphasize that evaluating behavioral changes and work outcomes requires perspectives beyond those of the training participants to enhance the validity of the findings. This research was limited to:

- the four (4) distance learning programs mentioned above,
- short-term post-learning evaluations (not longitudinal),
- data available in the 2025 Epaspem Level 4 evaluation report,
- the context of distance learning implementation without comparison with traditional learning (taking into account budget efficiencies in traditional learning in 2025).

These limitations were implemented to avoid overgeneralization and maintain the accuracy of the interpretation of the findings, as recommended by Rossi, Lipsey, and Freeman (2004) in public program evaluation research with limited time and resources.

This research used quantitative analysis to assess the empirical performance of the Kirkpatrick Model in the context of the four training programs used in this study. Analytical techniques in the training evaluation report include:

- descriptive statistics on Level 3 and Level 4 evaluation scores,
- paired sample test (paired sample test or Wilcoxon signed-rank test).

Arikunto (2010) states that paired sample test is commonly used in training evaluations to assess changes before and after an intervention. In this study, the results of the statistical tests are not interpreted as causal evidence of the effectiveness of distance learning (PJJ), but rather as indicators of the Kirkpatrick Model's ability to capture post-learning changes.

Qualitative analysis was conducted on a limited basis using content analysis of the narrative findings in the fourth PJJ Epaspem report. Krippendorff (2018) explains that content analysis is effective for identifying the meaning, context, and limitations of quantitative data representation.

This study places the Kirkpatrick Model as the primary focus of the evaluation. Public sector employee training is analyzed through the application of Kirkpatrick Levels 3 and 4. The evaluation results are then used to assess the model's adequacy in the context of training evaluation. Conceptually, this study demonstrates that:

a. Employee training contributes to changes in work behavior and organizational outcomes, as measured using the Kirkpatrick Model.
b. These measurement results are then analyzed to identify the strengths and limitations of the Kirkpatrick Model in addressing evaluation needs in the public sector.
c. The model's adequacy is assessed based on several criteria: indicator completeness, ability to attribute impact, practicality, and alignment with the organization's strategic objectives.
d. The analysis results then serve as the basis for formulating recommendations regarding the development, enrichment, or integration of the Kirkpatrick Model with other evaluation models as proposed in the literature (Faizin & Kusumaningrum, 2023).

The conceptual framework in this study was developed using a model-oriented evaluation approach, focusing on assessing the adequacy and relevance of the evaluation model, rather than solely on the outcomes of the program being evaluated. The assessment of model adequacy is based on program evaluation literature, which emphasizes several important aspects: model comprehensiveness (Chen, 2005; Rossi et al., 2004), ability to explain causal relationships (Rossi et al., 2004), practicality in implementation (Wholey, 1979; Patton, 2008), and alignment with the organization's strategic objectives (OECD, 2019).

Based on this rationale, this study positions the Kirkpatrick Model as the object of evaluation, rather than as a methodological assumption accepted without testing. Public sector personnel training serves as the empirical context to examine the extent to which the Kirkpatrick Model, particularly at Levels 3 and 4, provides an adequate picture for comprehensively evaluating training performance.

The Kirkpatrick Model is based on the assumption that training produces change through a hierarchical and sequential process, starting with participant reactions, learning, behavioral change, and ultimately impacting organizational outcomes (Kirkpatrick & Kirkpatrick, 2006). Within this framework, change at lower levels is seen as a prerequisite for change at higher levels.

However, several evaluation experts emphasize that the relationship between these levels is not always linear. Bates (2004) criticized the Kirkpatrick Model for tending to ignore contextual and systemic factors that influence organizational performance, particularly in complex public sector environments. Furthermore, Phillips and Phillips (2007) stated that organizational-level outcomes are generally influenced by various factors outside of training, making it difficult to attribute Level 4 results solely to the direct impact of training.

Within the conceptual framework of this research, these assumptions are not positioned as final truths, but rather serve as an analytical basis for testing the adequacy of the model's logic when confronted with the empirical conditions of public sector organizations.

RESULTS AND DISCUSSION

Results

1. Empirical Performance of the Kirkpatrick Model in Post-Learning Evaluation

Based on the post-learning evaluation reports analyzed, the application of the Kirkpatrick Model Level 3 and Level 4 empirically generated data indicating post-training changes, both in employee work behavior and in work outcomes, represented by the organizational performance of the relevant unit. The quantitative analysis revealed an increase in the average post-training work behavior score compared to pre-training levels, supported by statistically significant paired-difference test results.

These findings align with the findings of Kirkpatrick and Kirkpatrick (2007), who stated that Level 3 evaluation is effective for identifying learning transfer to the workplace, particularly when data is collected from various sources, such as alumni, direct supervisors, and coworkers. Empirically, the evaluation results indicate that most training alumni have applied the acquired competencies in their daily tasks, particularly in technical aspects of their work and adherence to work procedures. Evaluation at Level 3 also demonstrated changes in staff work behavior after training, particularly in aspects of professionalism, adaptability, and competency application in the workplace. These changes were still influenced by organizational support and the work environment (Azmy & Setiarini, 2023).

At Level 4, post-learning evaluations demonstrated a perception of increased alumni contributions to work unit performance, including improved output quality, timely completion of work, and improved understanding of task substance. This finding is consistent with research by Khalimi and Sihombing (2021) and Ritonga, Saepudin, and Wahyudin (2019), which found that Kirkpatrick Level 4 evaluations were able to provide an indication of the impact of training on organizational performance, albeit as a performance proxy. Evaluation at this level also demonstrated that the impact of training on organizational performance tends to be predictive or indicative and cannot be fully attributed directly to learning outcomes, as it is influenced by various factors external to the organization (Engriyani & Rugaiyah, 2022; Handoko et al., 2022).

However, these empirical results should be interpreted with caution. Phillips and Phillips (2007) emphasized that Level 4 results in the Kirkpatrick Model are more appropriately understood as indicators of impact, not strong causal evidence. Therefore, the empirical performance of the Kirkpatrick Model in this study indicates that the model works operationally, but is not necessarily analytically adequate.

The evaluation results indicate that all training programs had a positive impact on both behavioral levels (Level 3) and organizational outcomes (Level 4). Participants' work behavior improved significantly across all programs, with the highest level of improvement occurring in the Geographic Information Systems (GIS) and Taxpayer Analysis Working Paper training, reaching 100%. At Level 4, the impact on organizational performance also showed strong results, particularly in the GIS and Analysis Working Paper training, which demonstrated consistency between behavioral improvements and work output. Meanwhile, the PWDOMI training demonstrated a fairly strong impact, but it was not fully distributed, necessitating strengthening in the implementation of learning outcomes.

Based on empirical results and reflective analysis, the adequacy of the Kirkpatrick Model was assessed using four evaluative criteria as formulated in the research conceptual framework. The quantitative assessment results indicate that there is variation in the level of adequacy of the evaluation models in each training program. The Group Taxpayer Analysis Working Paper training received the highest score of 4.63, placing it in the strongest category. This score indicates that the evaluation model in this training excels in terms of comprehensiveness, strong causal relationships, and strategic fit, supported by comprehensive indicators and more robust analysis.

Furthermore, the Tax Bailiff II and Geographic Information Systems (GIS) training each received a score of 4.25, also in the strongest category. Both trainings demonstrated good consistency between behavioral changes and performance outcomes, and were supported by indicators that were applicable and easy to implement in an organizational context.

Meanwhile, the Regional Mastery with Internet Media Optimization (PWDOMI) training received a score of 3.75, which is in the very strong category. However, this score is relatively lower than other training programs, particularly in the attribution aspect, indicating that the relationship between behavioral change and performance improvement is not entirely consistent.

Overall, these results indicate that all evaluation models fall within the strong to very strong category, with the main difference being the models' ability to explain the causal relationship between training and performance.

Based on empirical results and reflective analysis, the adequacy of the Kirkpatrick Model was assessed using four evaluative criteria as formulated in the research conceptual framework. The quantitative assessment results indicate that there is variation in the level of adequacy of the evaluation models in each training program. The Group Taxpayer Analysis Working Paper training received the highest score of 4.63, placing it in the strongest category. This score indicates that the evaluation model in this training excels in terms of comprehensiveness, strong causal relationships, and strategic fit, supported by comprehensive indicators and more robust analysis.

Furthermore, the Tax Bailiff II and Geographic Information Systems (GIS) training each received a score of 4.25, also in the strongest category. Both trainings demonstrated good consistency between behavioral changes and performance outcomes, and were supported by indicators that were applicable and easy to implement in an organizational context.

Meanwhile, the Regional Mastery with Internet Media Optimization (PWDOMI) training received a score of 3.75, which is in the very strong category. However, this score is relatively lower than other training programs, particularly in the attribution aspect, indicating that the relationship between behavioral change and performance improvement is not entirely consistent. Overall, these results indicate that all evaluation models fall within the strong to

very strong category, with the main difference being the models' ability to explain the causal relationship between training and performance. The results of the comparative post-learning evaluation can be seen in Table 1.

Table 1. Results of Model Adequacy Comparison

Completeness (Weight 25%)

Indicators	Bailiff II	SIG	KKAWP Group	PWDOMI
Model includes outputs and outcomes	4	4	5	4
Key variables identified	4	4	5	4
Relationships between components are clear	4	4	5	4
Average Completeness	**4.00**	**4.00**	**5.00**	**4.00**

Attribution (Weight 25%)

Indicators	Bailiff II	SIG	KKAWP Group	PWDOMI
Clear cause-and-effect relationship (before-after)	4	4	5	3
Consistency of results (indicating minimal external bias)	4	4	4	3
Average Attribution	**4.00**	**4.00**	**4.50**	**3.00**

Practicality (Weight 25%)

Indicators	Bailiff II	SIG	KKAWP Group	PWDOMI
Data is easily obtained	4	5	4	4
Indicators are measurable	4	5	4	4
Easy to implement by practitioners	4	5	4	4
Average Practicality	**4.00**	**5.00**	**4.00**	**4.00**

Practicality (Weight 25%)

Indicators	Bailiff II	SIG	KKAWP Group	PWDOMI
Aligned with vision/mission	5	4	5	4
Supporting KPIs	5	4	5	4
Relevant to organizational policies	5	4	5	4
Average Strategic Alignment	**5.00**	**4.00**	**5.00**	**4.00**

Aggregate Calculation Per Training

Training	C	A	P	S	Final Score
Tax Bailiff II	4.00	4.00	4.00	5.00	**4.25**
SIG	4.00	4.00	5.00	4.00	**4.25**
KKAWP Group	5.00	4.50	4.00	5.00	**4.63**
PWDOMI	4.00	3.00	4.00	4.00	**3.75**

Discussion

The scoring results indicate that the Group Taxpayer Analysis Working Paper training had the highest model adequacy level with a score of 4.63, followed by the Tax Bailiff II and GIS training with a score of 4.25. Meanwhile, the PWDOMI training received a score of 3.75, indicating that although the model was adequate, there was still room for improvement, particularly in the attribution aspect. Overall, all evaluation models fell into the strong to very strong category.

Overall, the results and discussion indicate that the Kirkpatrick Model is operationally adequate but conceptually limited in evaluating the training performance of public sector officials. This finding reinforces previous studies using the Kirkpatrick Model and provides a new contribution by shifting the focus from training effectiveness to the adequacy of the evaluation model.

Thus, this study supports the view that the Kirkpatrick Model remains relevant as an initial evaluation tool and for monitoring training outcomes, but needs to be enriched or integrated with other evaluation models—such as CIPP or other logic models—to obtain a more comprehensive picture of training performance in the public sector (Faizin & Kusumaningrum, 2023).

CONCLUSION

This study aims to assess the adequacy of the Kirkpatrick Model in evaluating the performance of distance learning (PJJ)-based public sector personnel training through a post-learning evaluation study. Based on the evaluation results and discussion, several key conclusions can be drawn as follows:

1. Empirically, the Kirkpatrick Model Levels 3 and 4 are able to capture changes in work behavior and indicate short-term performance outcomes of alumni in the four PJJ programs. The Level 3 evaluation indicates transfer of learning to the workplace, particularly in technical and procedural aspects relevant to each PJJ's respective tasks. The Level 4 evaluation indicates a perception of alumni contribution to work unit performance, although this contribution is largely represented through individual output and direct superior assessments.
2. Second, from a model adequacy perspective, the Kirkpatrick Model is deemed operationally adequate but conceptually limited. The model's practicality is a key strength, particularly in the context of PJJ, which has limited evaluation time and resources. However, this model is not yet fully capable of representing the complexity of public sector organizational performance, particularly in the aspect of attribution of impact and the influence of systemic factors such as policies, organizational structure, and workload.
3. Third, the main weakness of the Kirkpatrick Model, which consistently emerged across the four distance learning programs (PJJ), is the limited attribution of impact. Identified performance changes cannot be fully attributed to the direct impact of PJJ, as they are influenced by various external factors beyond the training. This situation is further exacerbated by the nature of the evaluation, which was conducted relatively quickly after the PJJ's completion.
4. Fourth, in terms of strategic alignment, the Kirkpatrick evaluation results for the four distance learning programs are more relevant for tactical and operational decision-making, such as improving PJJ design or assigning alumni, but are not yet robust enough to support long-term strategic decisions or assess the impact of comprehensive civil servant training policies.

Overall, this study confirms that the use of the Kirkpatrick Model in evaluating PJJ for civil servants remains relevant and useful, but should be understood as an initial and short-term evaluation tool, not as the sole framework for evaluating public sector training performance.

Based on the research conclusions, recommendations are formulated in three main areas: training evaluation practices and distance learning management policies.

1. Selective and contextual use of the Kirkpatrick Model
   It is recommended that the Kirkpatrick Model continue to be used for short-term distance learning evaluations, particularly to measure changes in work behavior (Level 3) and initial outcomes (Level 4). However, evaluation results should be treated as an indication of performance, not as the sole causal evidence of training success.
2. Strengthening Level 4 indicators based on organizational context

To increase the relevance of evaluations, Level 4 distance learning indicators need to be more closely linked to work processes and the context of the work unit, rather than solely to individual outputs. This is important so that post-learning evaluations provide a more meaningful picture for unit leaders.

3. Utilizing evaluation results to improve distance learning design
   Level 3 and Level 4 evaluation findings need to be systematically utilized as feedback loops to refine distance learning design, particularly regarding the appropriateness of materials to task requirements and supporting the implementation of learning outcomes in the workplace.
4. Integration of the Kirkpatrick Model with Other Evaluation Models
   For strategic evaluation of civil servant training performance, the Kirkpatrick Model is recommended to be combined with other evaluation models, such as the CIPP or logic model, to accommodate the dimensions of context, process, and long-term impact.
5. Clarify the objectives of distance learning evaluation from the planning stage
   The objectives of distance learning evaluation need to be clearly defined from the outset, whether aimed at operational accountability, improving learning quality, or assessing strategic impact. The clarity of these objectives will determine the adequacy of the evaluation model used.
6. Develop a tiered distance learning evaluation system
   It is recommended that distance learning evaluation be conducted in stages, starting with a short-term evaluation (post-training), to mid-term or longitudinal evaluation, to minimize limitations in impact attribution.

REFERENCE

Arikunto, S. (2010). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.

Arikunto, S., & Abdul Jabar, C. S. (2008). *Evaluasi Program Pendidikan: Pedoman Teoritis Praktis Bagi Mahasiswa Dan Praktisi Pendidikan*. Jakarta: Bumi Aksara.

Aziz, DS Widodo, IK Subagja. (2021). Effect of Training and HR Development on Employee Performance with Motivation as an Intervening Variable in PT. Rekasis Gigatama. *Global Journal of Engineering and Technology Advances*. 06(02), 169–181.

Azmy, A., & Setiarini, N. Y. (2023). Kirkpatrick Model as an Evaluation Training Program for Assessors: Case Study of a Government Employee. *International Journal of Management, Accounting and Economics*, 10(9), 675–696.

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

Bowen, G. A. (2009). Document Analysis As a Qualitative Research Method. *Qualitative Research Journal*, 9(2), 27–40. https://doi.org/10.3316/QRJ0902027

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks, CA: SAGE Publications.

Engriyani, E., & Rugaiyah, R. (2022). Evaluasi Program Pendidikan Dan Pelatihan Menggunakan Model Kirkpatrick Pada Institusi Pemerintah Di Indonesia. *Jurnal Ilmiah Wahana Pendidikan*, 8(23), 683–692.

Faizin, A., & Kusumaningrum, H. (2023). Review Model-Model Evaluasi Program Untuk Pendidikan Dan Pelatihan Online. *EduManajerial: Jurnal Manajemen Pendidikan*, 1(1), 42–54.
https://doi.org/10.15408/em.v1i1.32245

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program Evaluation: Alternative Approaches And Practical Guidelines* (4th ed.). Boston, MA: Pearson Education.

Handalani, R. T., & Soesanto, H. (2022). Evaluasi Pelatihan Model Kirkpatrick Dan Model Kemp Menggunakan SEM-PLS: Studi Kasus Pelatihan Bendahara Pengeluaran Di Provinsi DKI Jakarta. MONAS: *Jurnal Inovasi Aparatur*, *4*(2), 453–467.

Huey-Tsyh Chen. (2005). *Practical Program Evaluation: Theory-Driven Evaluation and the Integrated Evaluation Perspective (Thousand Oaks*, CA: Sage Publications).

Joseph S. Wholey, "Evaluability Assessment.". (2010). Handbook of Practical Program Evaluation, ed. Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer (San Francisco: Jossey-Bass).

Khalimi, A., & Sihombing, P. R. (2021). Evaluasi model Kirkpatrick pelatihan persiapan sertifikasi penyuluh antikorupsi pada Pusdiklat PSDM Kementerian Keuangan. *Perspektif: Jurnal Ekonomi dan Manajemen*, 19(2), 184–189.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.). San Francisco, CA: Berrett-Koehler Publishers.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2007). *Implementing The Four Levels: A Practical Guide For Effective Evaluation Of Training Programs*. San Francisco, CA: Berrett-Koehler Publishers.

Krippendorff, K. (2018). *Content analysis: An Introduction To Its Methodology* (4th ed.). Thousand Oaks, CA: SAGE Publications.

Lin, Y. T., & Chuang, H. T. (2011). The application of Kirkpatrick's Four-Level Training Evaluation Model To Evaluate Training Effectiveness. *International Journal of Management and Marketing Research*, 4(2), 17–26.

Michael Quinn Patton. (2008). *Utilization-Focused Evaluation*, edisi ke-4 (Thousand Oaks, CA: Sage Publications).

OECD. (2019). *Better Criteria for Better Evaluation: Revised Evaluation Criteria Definitions and Principles for Use* (Paris: OECD Publishing).

Patton, M. Q. (2015). *Qualitative Research & Evaluation Methods* (4th ed.). Thousand Oaks, CA: SAGE Publications.

Peraturan Menteri Keuangan Republik Indonesia Nomor 124 Tahun 2024 tentang Organisasi dan Tata Kerja Kementerian Keuangan. (2024). *Berita Negara Republik Indonesia Tahun 2024*.

Peter H. Rossi, Mark W. Lipsey, and Howard E. Freeman. (2004). Evaluation: A Systematic Approach, edisi ke-7 (Thousand Oaks, CA: Sage Publications).

Phillips, J. J., & Phillips, P. P. (2007). *Measuring Return On Investment In Training And Performance Improvement Programs*. Burlington, MA: Butterworth-Heinemann.

Pollitt, C., & Bouckaert, G. (2017). *Public Management Reform: A Comparative Analysis – Into The Age Of Austerity* (4th ed.). Oxford: Oxford University Press.

Ritonga, R., Saepudin, A., & Wahyudin, U. (2019). Penerapan Model Evaluasi Kirkpatrick Empat Level Dalam Mengevaluasi Program Diklat Di Balai Besar Pelatihan Pertanian Lembang. *Jurnal Pendidikan Nonformal*, 14(1), 12–19.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A Systematic Approach* (7th ed.). Thousand Oaks, CA: SAGE Publications.

Syafitri, I., & Tumanggor, B. F. (2023). Evaluasi Pelatihan Kebencanaan Bagi Pegawai Negeri Sipil Dengan Model Kirkpatrick Di Badan Nasional Penanggulangan Bencana Republik Indonesia. *Jurnal Administrasi Publik*, xx(x), xx–xx.

Yin, R. K. (2018). *Case Study Research And Applications: Design And Methods* (6th ed.). Thousand Oaks, CA: SAGE Publications.