
Assessing Employee Training Effectiveness: An Empirical Integration of Kirkpatrick Level 1 and Level 2 Evaluations

Frida Aprillia¹, Purwatiningsih², Dhuha Safria³, Alan Budi Kusuma⁴

^{1,2,3,4}Universitas Bina Sarana Informatika, Indonesia

E-mail: Frida.frp@bsi.ac.id¹

Article History:

Received: 01 Juni 2026

Revised: 31 Juli 2026

Accepted: 07 Agustus 2026

Keywords: training effectiveness; Kirkpatrick model; employee development; learning outcomes; training evaluation; private sector.

Abstract: Training evaluation in the private sector is often limited to participant satisfaction, providing insufficient evidence of whether training produces measurable learning outcomes. This study evaluates a basic training program in Jakarta by integrating Kirkpatrick's Level 1 (Reaction) and Level 2 (Learning). A quantitative descriptive design was employed involving 35 employees selected through purposive sampling. Data were collected through Likert-scale questionnaires and pre-test–post-test assessments and analyzed using descriptive statistics and a paired-samples t-test. The findings indicate that participants responded positively to the training, particularly in terms of facilitation quality and material relevance, although overall satisfaction was comparatively lower due to concurrent workload demands. Learning outcomes improved significantly after the training, with post-test scores substantially exceeding pre-test performance and showing greater consistency across participants. These results demonstrate that combining reaction and learning measures provides a more robust assessment of training effectiveness than satisfaction-based evaluation alone. The study highlights the importance of reducing work-related interruptions during training and extending future evaluations to behavioral transfer and organizational performance.

INTRODUCTION

For companies, employee training represents a form of strategic investment aimed at enhancing human resource competencies to remain relevant amidst intense business competition and technological disruption. Organizations are willing to allocate substantial funds with the expectation that these programs will improve employees' knowledge, skills, and work attitudes, ultimately leading to increased organizational productivity (Noe, 2020; Armstrong & Taylor, 2023). Therefore, systematic evaluation is crucial to ensure that the invested capital truly yields a worthwhile return.

Despite substantial budgets, implementing an ideal evaluation remains a major stumbling block in human resource management. Many organizations get caught up in the mere formality of budget absorption, focusing solely on measuring attendance rates or participant satisfaction with facilities and organizers. On the other hand, core essentials such as the achievement of learning

objectives are often overlooked (Arthur et al., 2003; Salas et al., 2012). Consequently, management lacks comprehensive data regarding how effectively the training actually boosts employees' real competencies.

To obtain objective data, the Kirkpatrick Model that originally devised by Kirkpatrick (1959) and updated in 2006, has become the most widely used evaluation framework. This model maps training success into four structured stages: Reaction, which measures participants' perceptions and satisfaction with the training process; Learning, which assesses the level of knowledge absorption, attitude changes, or new skill acquisition; Behavior, which observes performance changes in the workplace; and Results, which looks at the direct impact on organizational business performance. Unfortunately, reality on the ground shows that the majority of organizations only stop at Level 1 (Reaction) and neglect objective measurement at Level 2 (Learning) (Blume et al., 2010; Bates, 2004).

Similar conditions are commonly found within the private sector in DKI Jakarta. As the epicenter of the national economy, corporations in Jakarta demand a highly competitive workforce and routinely conduct training programs. However, their evaluations are often still administrative and subjective, limited merely to satisfaction surveys. This lack of systematic measurement of learning outcomes makes it difficult for companies to formulate evidence-based human resource development policies.

Although the Kirkpatrick Model has been widely applied in the public, healthcare, and education sectors, research focusing on the integration of Level 1 and Level 2 within private companies in DKI Jakarta remains very limited. Most prior studies only present the results of each level separately, without dissecting how participants' impressions (reaction) influence their absorption of the material (learning).

This study aims to evaluate the effectiveness of training programs through the integration of the Kirkpatrick Model Level 1 (Reaction) and Level 2 (Learning) in private companies within the DKI Jakarta region. The output is expected to enrich the human resource management literature while simultaneously serving as a practical guide for corporations in building an objective, systematic, and solution-oriented evaluation system.

LITERATURE REVIEW

Training is a structured and planned mechanism designed by organizations to sharpen employees' knowledge, technical skills, and behaviors so that work operations can run more effectively and efficiently (Noe, 2020). Beyond enhancing individual capacity, this program plays a crucial role in an organization's grand strategy as a form of investment to create a competitive advantage, while simultaneously strengthening human resource capacity to remain prepared for dynamic business environments and accelerating technological growth (Noe, 2020; Armstrong & Taylor, 2023). However, the success of this program should not only be measured by how substantial the presented material is or how competent the facilitators leading the sessions are, but also by the organization's capability to measure its tangible impact. Consequently, training evaluation is an absolute and inseparable part of the human resource development ecosystem to ensure the effectiveness of the programs that have been implemented.

Training evaluation is a structured process crucial for Human Resource Development to assess the extent to which implemented programs have successfully achieved their objectives and to measure the return on training investment (Salas et al., 2012). Through this process, organizations can gather objective data to measure post-training transitions, which encompass aspects of material comprehension, behavioral changes in the workplace, and their tangible

implications for corporate performance (Goldstein & Ford, 2002). Therefore, the essence of this evaluation goes beyond merely assessing participants' subjective satisfaction; instead, it serves as an evidence-based strategic foundation for management in making decisions regarding the continuity, refinement, or design of future human resource development programs.

The Kirkpatrick Model, originally introduced by Kirkpatrick (1959) and further developed with Kirkpatrick (2006), is the most widely utilized framework for measuring training effectiveness through four structured levels. These stages comprise Reaction (Level 1) to assess participants' perceptions and satisfaction; Learning (Level 2) to measure increases in knowledge, skills, confidence, and attitude changes; Behavior (Level 3) to observe behavioral changes within the workplace; and Results (Level 4) to evaluate the financial or operational impacts on the organization's strategic goals. Although this model is highly favored due to its simple structure, ease of application, and capability to present effectiveness data progressively, in reality, many organizations remain stuck on implementing Level 1 alone, leaving data regarding the success of the actual knowledge transfer (learning) process heavily limited (Bates, 2004).

The Kirkpatrick Model operates on the assumption of a causal relationship where a positive reaction from participants toward the training, such as interest in relevant material, competent instructors, and interactive methods, will boost their motivation and engagement, ultimately enhancing learning outcomes. Nevertheless, this assumption is not always absolute, as subjective participant satisfaction does not necessarily correlate directly with a genuine increase in knowledge (Alliger & Janak, 1989). Consequently, to obtain a more comprehensive overview of training effectiveness, organizations must not rely solely on satisfaction indicators, but must integrate them with an objective measurement of actual learning outcomes.

RESEARCH METHOD

This study utilizes a quantitative descriptive approach, aiming to examine the causal relationship between participants' reaction levels (Reaction) and learning outcomes (Learning) based on the Kirkpatrick Model. The population in this study encompasses employees at a private company in the DKI Jakarta region who have participated in the division's basic training program. Sampling was conducted using a non-probability sampling technique, specifically purposive sampling, by establishing specific criteria such as active employee status and involvement in training sessions that evaluated both satisfaction aspects and learning achievements.

The research variables are operationalized into two primary focuses: the independent variable, Reaction (Level 1), which measures participants' perceptions of instructor quality, material, methods, and training facilities; and the dependent variable, Learning (Level 2), which measures objective increases in knowledge, skills, and changes in work attitudes. The primary data collection process was carried out by distributing Likert-scale questionnaires for the Reaction variable, while data for the Learning variable were obtained from the documentation of participants' pre-test and post-test scores.

The data processing in this study involves two software programs: Microsoft Excel to analyze the participants' satisfaction levels or reactions, and SPSS to evaluate the increase in their knowledge.

RESULT AND DISCUSSION

This study involves a sample of 35 individuals consisting of staff members from a division within a private company in DKI Jakarta. Level 1 Evaluation Analysis: Reaction. The evaluation was conducted to determine the participants' reactions during the training. The assessment aspects

comprise 10 (ten) evaluation indicators, which include instructor evaluation consisting of 2 statements, material presentation evaluation consisting of 2 statements, training methods consisting of 2 statements, training facilities with 2 statements, training relevance to the job with 1 statement, and overall training satisfaction with 1 statement.

Table 1. Reaction Result

No	Indicator	Average	Value
Instructor			
1	I1 = The instructor mastered the training material.	3.94285714	Good
	I2 = The instructor delivered the material clearly.	3.8	Good
Training Material			
2	M1 = The training material was easy to understand.	3.71428571	Good
	M2 = The training material aligned with the training objectives.	3.94285714	Good
Training Method			
3	T1 = The training delivery method was engaging.	3.8	Good
	T2 = The training method facilitated the learning process.	4.05714286	Very Good
Training Facilities			
4	F1 = The training room and facilities were adequate.	3.8	Good
	F2 = The training media/equipment functioned properly.	3.82857143	Good
Training Relevance			
5	R1 = The training material is relevant to my job.	4.02857143	Very Good
Overall Satisfaction			
6	S1 = I am satisfied with the overall execution of the training.	3.48571429	Good

The analysis of the Level 1 (*Reaction*) evaluation based on the Kirkpatrick Model for this basic division training indicates generally positive responses, with ratings ranging from "Good" to "Very Good." The training method indicator that facilitates the learning process (T2) and the relevance of the material to the job (R1) ranked highest, with scores of 4.06 and 4.03, respectively. These results confirm the view of Kirkpatrick and Kirkpatrick (2006) that Level 1 evaluation does not merely measure entertainment value, but rather the practical value of the program for the participants' operational needs. Furthermore, the high ratings for the instructor's mastery of the material (I1) and the alignment of the material with the training objectives (M2), both achieving a score of 3.94, demonstrate a strong synergy between facilitator quality and curriculum design. This aligns with the findings of Alliger et al. (1997), which indicate that participants' utility judgments regarding the material have a stronger correlation with driving knowledge transfer than mere emotional satisfaction.

Although the specific components were rated very highly, the overall satisfaction indicator (S1) recorded the lowest score at 3.49. This decline reflects Bates' (2004) theory regarding the vulnerability of Level 1 questionnaires to environmental biases and personal expectations outside the technical classroom context, such as a dense training schedule or the participants' daily workloads. This condition reinforces the argument by Morgan and Casper (2000) on the necessity

of separating participant reaction dimensions into *utility reaction* and *affective reaction*. The data above confirm that while this training is highly superior in terms of functionality (*utility*), management still needs to address macro-level comfort factors (*affective*) to boost overall satisfaction for more efficient HR development programs in the future.

Table 2. Statistic Descriptive

	N	Min	Max	Mean	Std. Deviation
Pre Test	35	20	55	39	9,61
Post Test	35	60	95	76,86	8,67

The results of the descriptive statistical analysis of the 35 respondents indicate a clear surge in competence following the implementation of the division's basic training program. This is demonstrated by the sharp increase in the participants' mean score, which rose drastically from 39.00 on the pre-test to 76.86 on the post-test. This performance improvement is also reflected in the shift of the score boundaries, where the participants' minimum score climbed from 20.00 to 60.00, and the maximum score increased from 55.00 to reach 95.00. Furthermore, the decrease in the standard deviation from 9.61 to 8.67 indicates that the distribution of scores post-training became narrower. Overall, these data indicate that the knowledge transfer process was effective, successfully driving higher and more evenly distributed (homogeneous) comprehension levels among the staff.

Table 3. Paired T Test

	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	dF	Sig. (2-tailed)
				Lower	Upper			
Pair	-37,857	11,523	1,948	-41,815	-33,899	-19,437	34	0,000

Through hypothesis testing using the Paired Samples T-Test method, a t calculated value of -19.437 was found with 34 degrees of freedom (df), and a Sig. (2-tailed) value of 0.000. This significance level, which sits far below the error tolerance threshold ($0.000 < 0.05$), provides a strong basis that there is a significant increase in the staff's knowledge scores after attending the basic division training program compared to before the training. This finding empirically confirms that there is a distinct and genuine difference, as well as an enhancement in competence, between before (pre-test) and after (post-test) the staff participated in the basic division training at this private corporation in DKI Jakarta. The mean difference of -37.857 quantitatively verifies that the educational intervention successfully boosted the employees' cognitive capacity significantly.

From a theoretical perspective, this achievement meets the success criteria for Level 2 Evaluation (Learning) within the framework of the Kirkpatrick Model (1959, 2006). This level is oriented toward measuring how far participants absorb knowledge, sharpen skills, or shift their paradigms post-training. The increase in the mean score from an initial 39.00 to 76.86 indicates that the knowledge transfer mechanism operated optimally within the training room. This result simultaneously proves that the positive impressions recorded in Level 1 (Reaction) previously, particularly regarding the instructor's interactive methods and curriculum quality, did not stop at mere visual satisfaction, but were successfully converted into substantial conceptual comprehension.

This positive correlation aligns with meta-analysis research by Sitzmann et al. (2008) regarding instructional design effectiveness, which states that training designed with a clear workflow is capable of transforming participants' initial cognitive limitations into measurable proficiency. Furthermore, this massive surge in learning outcomes is supported by Blume et al.'s (2010) argument regarding the contextual proximity of the material; when participants perceive that the training topics directly intersect with their actual daily workloads at the office, their internal motivation to deeply master the material during the final exam (post-test) is powerfully stimulated.

In practical terms, the narrowing variation of data in the post-test standard deviation (8.67) compared to the pre-test (9.61) reinforces the cognitive conditioning theory pioneered by Kraiger, Ford, and Salas (1993). This statistical phenomenon indicates that an equalization or standardization of competence has occurred. The training is proven capable of eroding the comprehension gap among staff, making the team's cognitive capacity more uniform (homogeneous) post-program. For corporate management, the validity of these Level 2 results provides an assurance that the allocated human capital investment has successfully formed a mature collective knowledge foundation, a crucial asset before employees are expected to apply it in the form of workplace behavior changes in the field (Kirkpatrick Level 3: Behavior).

CONCLUSION

The evaluation of the division's basic training program demonstrates that the program was effective at both Level 1 (Reaction) and Level 2 (Learning) of the Kirkpatrick Model. Participants responded most positively to the learning methods and the relevance of the training materials, indicating that the program was perceived as useful, applicable, and supportive of their work-related learning needs. This positive reaction is consistent with the substantial improvement in participants' cognitive performance after the training. The increase in post-test scores, accompanied by a narrower distribution of scores, suggests that the program succeeded in improving knowledge while also reducing disparities in competence among participants. The training therefore functioned not merely as a knowledge-transfer activity but also as a mechanism for creating a more consistent level of understanding across the team.

Despite these positive outcomes, overall participant satisfaction remained comparatively lower than other reaction indicators. This finding suggests that participants distinguished between the functional value of the training and their emotional comfort during its implementation. The training could be regarded as relevant and effective while still generating lower affective satisfaction because participants had to manage demanding schedules and unfinished daily workloads during the program. This supports the view that reaction evaluation should be treated as multidimensional, as positive perceptions of instructional quality and usefulness do not automatically translate into equally high levels of overall satisfaction. From a managerial perspective, training effectiveness therefore depends not only on the quality of content and delivery but also on organizational conditions that allow employees to participate without excessive competing work demands.

This study is limited to the first two levels of the Kirkpatrick Model and relies primarily on quantitative data collected immediately after the training. Consequently, the findings cannot determine whether the acquired knowledge was retained over time, transferred into workplace behavior, or ultimately contributed to organizational performance. The use of closed-ended instruments also limits the ability to capture participants' deeper emotional and contextual experiences during the program. Future studies should therefore extend the evaluation to Level 3 (Behavior) and Level 4 (Results), employ longitudinal designs with delayed post-tests, and

incorporate qualitative methods to examine knowledge retention, behavioral transfer, and organizational outcomes more comprehensively. In practical terms, management should reduce competing workloads during training and strengthen post-training supervision by line managers so that improvements in knowledge can be translated into sustained workplace performance.

REFERENCES

- Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology*, 42(2), 331–342. <https://doi.org/10.1111/j.1744-6570.1989.tb00661.x>
- Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, 50(2), 341–358. <https://doi.org/10.1111/j.1744-6570.1997.tb00911.x>
- Armstrong, M., & Taylor, S. (2023). *Armstrong's handbook of human resource management practice* (16th ed.). Kogan Page.
- Arthur, W., Jr., Bennett, W., Jr., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology*, 88(2), 234–245. <https://doi.org/10.1037/0021-9010.88.2.234>
- 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., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management*, 36(4), 1065–1105. <https://doi.org/10.1177/0149206309352880>
- Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation* (4th ed.). Wadsworth.
- Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of the American Society of Training Directors*, 13(11), 3–9.
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.). Berrett-Koehler Publishers.
- 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. <https://doi.org/10.1037/0021-9010.78.2.311>
- Morgan, R. B., & Casper, W. J. (2000). Examining the factor structure of participant reactions to training: A multidimensional approach. *Human Resource Development Quarterly*, 11(3), 301–317. [https://doi.org/10.1002/1532-1096\(200023\)11:3%3C301::AID-HRDQ7%3E3.0.CO;2-P](https://doi.org/10.1002/1532-1096(200023)11:3%3C301::AID-HRDQ7%3E3.0.CO;2-P)
- Noe, R. A. (2020). *Employee training and development* (8th ed.). McGraw-Hill Education.
- Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The science of training and development in organizations: What matters in practice. *Psychological Science in the Public Interest*, 13(2), 74–101. <https://doi.org/10.1177/1529100612436661>
- Sitzmann, T., Brown, K. G., Casper, W. J., Ely, K., & Zimmerman, R. D. (2008). A review and meta-analysis of the nomological network of trainee reactions. *Journal of Applied Psychology*, 93(2), 280–295. <https://doi.org/10.1037/0021-9010.93.2.280>