

Effectiveness of Behavioral Attitude Assessment Instruments in Structural Leadership Training

Eris Yustiono

Department of Human Resource Management Apparatus, Politeknik STIA LAN Bandung, Indonesia

Abstract

This study examines the effectiveness of behavioral attitude assessment instruments used in structural leadership training programs by analyzing how instrument- and assessor-related conditions shape their validity, reliability, objectivity, and practicality. Employing a quantitative approach enriched with qualitative feedback, data were collected from 41 facilitators (widyaiswara in Indonesian terminology) from the PKASN Training Center and other government training institutions. The findings indicate that while the instrument is practical and easy to apply, its effectiveness is constrained by weaknesses in validity, reliability, and objectivity. These limitations are associated with inconsistent interpretation of indicators, the presence of subjective bias, and the absence of standardized observation guidelines, all of which reduce inter-rater reliability and behavioral accuracy. The study highlights that instrument effectiveness is not solely determined by its design but also by how it is interpreted and implemented in practice. The findings underscore the need for evidence-based evaluation instruments that promote fairness, accuracy, and accountability in leadership development programs and support the development of more objective and behaviorally grounded assessment systems in government training institutions.

Article Info

Keywords:
Behavioral Assessment,
Instrument Effectiveness,
Leadership Training,
Training Evaluation

JEL Classification:
C91, M53, P46

Corresponding Author:
Eris Yustiono
(erisyustiono@gmail.com)

Received: 08-10-2025
Revised: 07-11-2025
Accepted: 15-12-2025
Published: 31-12-2025



1. Introduction

Over the past decade, training has become a strategic instrument employed by both public and private organizations to develop employee competencies and support organizational performance (Phillips & Phillips, 2016; Bates, 2020). In the context of Indonesia's public sector, structural leadership training constitutes a mandatory managerial competency development program for all structural officials according to their respective levels, as mandated by the Regulation of the National Institute of Public Administration (Lembaga Administrasi Negara/ LAN) of the Republic of Indonesia Number 5 of 2022 on the Implementation of Structural Leadership Training. The regulation emphasizes that structural training aims to develop managerial competencies that enable structural officials to perform transformational and adaptive leadership roles within a dynamic bureaucracy.

Evaluation of structural leadership training is conducted across four main dimensions: academic evaluation, field learning evaluation, evaluation of leadership actualization projects, and behavioral attitude evaluation. Among these four dimensions, behavioral attitude evaluation holds a particularly critical position because it directly determines participant graduation. Participants who receive a "poor" qualification in the behavioral attitude dimension are declared not to have passed the training—unlike the other dimensions, which only affect the timing of graduation. This underscores the strategic importance of the behavioral attitude dimension in shaping structural leadership competencies in Indonesia.

However, field observations and preliminary interviews with facilitators and training organizers reveal several issues indicating weaknesses in the implementation of behavioral

attitude evaluation. First, facilitators often complete behavioral assessments at the end of the training program without maintaining systematic observation notes. Second, most facilitators lack clear guidelines regarding observational indicators, leading to subjective assessments. Third, a high tendency toward score homogeneity was identified, with almost all participants receiving similar scores in the behavioral dimension. These conditions indicate potential problems in the validity, reliability, and objectivity of the evaluation instrument used.

In international practice, training evaluation is recognized as a systematic process to measure training effectiveness and its impact on individual and organizational performance (Phillips & Phillips, 2016; Bates, 2020). Effective training evaluation must provide information on the extent to which training objectives are achieved, the quality of the instruments used, and the relevance of training outcomes to improved work competencies (Kirkpatrick & Kirkpatrick, 2016). The most widely cited evaluation model is Kirkpatrick's four-level model, which consists of: (1) reaction, (2) learning, (3) behavior, and (4) results.

In the context of structural leadership training in Indonesia, behavioral attitude evaluation essentially reflects the third level of Kirkpatrick's model—behavior. This dimension focuses on the extent to which participants demonstrate behavioral changes such as cooperation, leadership, discipline, and initiative within the training environment. According to Ajzen (2015), behavioral attitude refers to an individual's tendency to respond positively or negatively to a particular object or situation, based on cognitive, affective, and conative components. This perspective suggests that behavioral attitude assessment is not limited to observable actions but also reflects underlying values, beliefs, and motivations (Fishbein & Ajzen, 2010; Katz & Stotland, 2014).

Furthermore, theories on evaluation instruments emphasize four essential criteria for producing objective and reliable assessments: validity, reliability, objectivity, and practicality (Fraenkel et al., 2019; Arikunto, 2018; Azwar, 2018). Validity refers to the extent to which an instrument measures what it is intended to measure; reliability concerns the consistency of measurement results; objectivity ensures freedom from assessor bias; and practicality relates to the ease of implementation and interpretation. When any of these criteria are not met, evaluation results become less trustworthy and fail to accurately represent participant performance (Mulyasa, 2019).

In the context of structural leadership training, several internal studies indicate that the validity and reliability of behavioral assessment instruments have not been fully established. This is due to the absence of empirical validation mechanisms and the lack of inter-rater reliability testing. These conditions may lead to unfair assessments and reduce the credibility of training evaluation results. Meanwhile, modern bureaucracy requires all forms of performance and competency measurement for civil servants to adhere to evidence-based assessment principles to ensure accountability both scientifically and administratively (OECD, 2019).

Previous studies also highlight the importance of behavioral evaluation as an indicator of leadership training effectiveness. Rahman et al. (2021), for example, found that significant leadership behavior changes occur only when evaluation processes use instruments with high reliability. Similarly, Kibe and Namusonge (2015) demonstrated that non-objective behavioral assessment may create a training-performance gap between expected competencies and actual participant behavior. Therefore, improving the effectiveness of behavioral assessment instruments is an urgent need for institutions administering structural leadership training.

The gap between expectations and actual practice highlights the need for scientific investigation into the effectiveness of the behavioral attitude evaluation instrument used in structural leadership training. This study is essential not only for ensuring alignment between the instrument and training evaluation theories and behavioral attitude theories but also for providing empirical foundations for developing more accountable and behavior-oriented leadership training policies.

Based on this background, this study examines the effectiveness of the behavioral attitude evaluation instrument used in structural leadership training by explaining how instrument-related and assessor-related conditions shape its validity, reliability, objectivity, and practicality. Accordingly, the study aims to provide an integrated analysis of the instrument's effectiveness as it is implemented in practice, rather than treating influencing factors as separate variables.

2. Methods

This study employed a quantitative survey approach because this method allows the researcher to capture systematic and measurable perceptions from facilitators who directly evaluate participants' behavioral attitudes in structural leadership training. To complement these quantitative findings, the study also incorporated semi-open questionnaire items that enable respondents to provide narrative reflections. The inclusion of qualitative elements was considered important to obtain contextual insights that numerical responses alone could not sufficiently explain, thereby strengthening the interpretive depth of the findings.

The study involved facilitators from the PKASN Training Center as the primary population, with additional facilitators from other government training institutions included to broaden perspectives and provide comparative variation. This group was selected because they possess direct experience in using and assessing the behavioral attitude evaluation instrument, making their input highly relevant for evaluating its effectiveness. Respondents were chosen using a random sampling strategy while ensuring that all participants had substantive involvement in the assessment process during training.

Data were gathered using a semi-open questionnaire designed to capture respondents' evaluations of the instrument based on four central criteria: validity, reliability, objectivity, and practicality. This type of instrument was chosen because it offers an efficient way to collect structured quantitative data while still allowing respondents to elaborate on the reasoning behind their choices, which enriches the interpretive value of the dataset—particularly in field conditions where time and access are limited.

The collected data were analyzed descriptively to observe general patterns, tendencies, and variations in respondent perceptions. The analysis involved verifying and organizing the data, examining frequency distributions, and interpreting emerging patterns by integrating respondents' narrative explanations. This analytical approach was selected because it aligns with the study's aim to provide an evidence-based understanding of how the evaluation instrument operates in practice and what factors contribute to—or hinder—its effectiveness.

3. Results and Discussion

This section presents the descriptive findings of the study based on responses from 41 facilitators involved in structural leadership training. The results provide an overview of facilitators' understanding of behavioral attitude components, their interpretation of specific indicators, the timing and objectivity of their assessments, and their perceptions of the effectiveness and suitability of the evaluation instrument. The analysis was conducted quantitatively using descriptive techniques with frequency tabulations, accompanied by narrative interpretations grounded in theory. The findings are organized according to the four dimensions of instrument effectiveness—validity, reliability, objectivity, and practicality—as outlined by Fraenkel et al. (2019), Arikunto (2018), and Phillips & Phillips (2016). Each dimension is explained through empirically examined indicators derived from the responses to the semi-open questionnaire.

3.1. Results

Respondents' Knowledge of Behavioral Attitude Components

Respondents' understanding of the behavioral attitude assessment components is a key prerequisite for ensuring accurate evaluation. Table 1 shows the distribution of responses regarding their understanding of the assessed components in structural leadership training.

Table 1. Respondents' Knowledge of Behavioral Attitude Components

Category	Frequency	Percentage
Yes	37	90.24%
Do Not Know	0	0.00%
Partially Know	4	9.76%

Source: Data Processed (2024)

Most facilitators demonstrated a strong understanding of the behavioral attitude dimensions evaluated in structural leadership training. As shown in Table 1, 90.24% of respondents reported that they understood all aspects being assessed, while only 9.76% indicated partial knowledge. This suggests that facilitators possess sufficient conceptual awareness of the key behavioral components—cooperation, initiative, leadership, and discipline—that the instrument aims to evaluate. Although the overall knowledge level appears adequate, the subsequent results indicate that this awareness does not always translate into consistent operationalization during assessment.

Respondents' Understanding of Cooperation and Initiative

Cooperation and initiative constitute the first dimension of participants' behavioral attitudes. Respondents were asked whether they perceive the two aspects as distinct or similar.

Table 2. Understanding the Difference Between Cooperation and Initiative

Category	Frequency	Percentage
Same	3	7.32%
Different	38	92.68%

Source: Data Processed (2024)

Almost all facilitators distinguished cooperation from initiative, with 92.68% identifying them as different constructs, as presented in Table 2. This finding is notable because the current assessment instrument merges these two behaviors into a single item. While respondents generally recognize the conceptual difference between the two, the merged item may challenge facilitators during the scoring process because it requires evaluating two behaviors simultaneously.

Respondents' Understanding of Leadership

Leadership is the second dimension of participants' behavioral attitudes. Respondents were asked whether they understand the concept of leadership.

Table 3. Understanding the Concept of Leadership

Category	Frequency	Percentage
Yes	41	100.00%
No	0	0.00%

Source: Data Processed (2024)

All respondents reported that they understood the concept of leadership, as shown in Table 3. In general, facilitators perceived leadership as the ability to influence, guide, and motivate others toward achieving shared objectives during training activities. This shared understanding indicates that leadership, as a behavioral dimension, is conceptually familiar to facilitators involved in structural leadership training. However, although respondents agreed on the general meaning of leadership, subsequent findings reveal variations in how specific leadership indicators are interpreted and applied in practice, suggesting potential implications for consistency in behavioral assessment.

Respondents' Understanding of Discipline

Discipline is the third dimension of participants' behavioral attitudes. Respondents were asked whether they understand the concept of discipline.

Table 4. Understanding the Concept of Discipline

Category	Frequency	Percentage
Yes	41	100.00%
No	0	0.00%
Total	41	100%

Source: Data Processed (2024)

Similarly, all respondents indicated that they understood the concept of discipline, as presented in Table 4. Most facilitators associated discipline primarily with punctuality and compliance with time-related rules. While these elements are part of disciplined behavior, this interpretation reflects a relatively narrow understanding of the construct. Broader aspects of discipline, such as responsibility, consistency in task completion, and adherence to organizational norms and values, were less frequently emphasized. This variation in interpretation suggests that although the concept of discipline is widely recognized, differences in how facilitators conceptualize its behavioral indicators may affect the uniformity of assessment outcomes.

Timing of Assessment

The timing of assessment plays an important role in influencing evaluation reliability.

Table 5. Timing of Assessment by Facilitators

Category	Frequency	Percentage
Immediately after teaching (same day)	18	43.90%
1–3 days after teaching	13	31.71%
4–7 days after teaching	3	7.32%
More than 7 days after teaching	7	17.07%
Total	41	100%

Source: Data Processed (2024)

The timing of behavioral assessments varied considerably among respondents. As shown in Table 5, 43.90% conducted assessments on the same day as teaching, while 31.71% conducted them within one to three days. A smaller proportion, 17.07%, submitted scores more than seven days after teaching. These delays are noteworthy because the longer the interval between observation and scoring, the greater the likelihood of memory distortion, especially when detailed observation notes are not maintained.

Respondents' Perceptions of Assessment Objectivity

Objectivity is a fundamental pillar of any evaluation instrument.

Table 6. Respondents' Perceptions of Assessment Objectivity

Category	Frequency	Percentage
Yes	21	51.22%
No	4	9.76%
Sometimes	16	39.02%
Total	41	100%

Source: Data Processed (2024)

Respondents provided mixed views regarding their ability to conduct objective assessments. Table 6 shows that only 51.22% consistently felt able to assess participants objectively, while 39.02% stated that they were sometimes objective and 9.76% reported that they were not objective. These results indicate that almost half of the facilitators encountered difficulties in maintaining objectivity during the evaluation process.

Respondents' Perceptions of Instrument Effectiveness

The effectiveness of the instrument serves as the central focus of this study.

Table 7. Respondents' Perceptions of Instrument Effectiveness

Category	Frequency	Percentage
Effective	23	56.10%
Not Effective	12	29.27%
Do Not Know	6	14.63%
Total	41	100%

Source: Data Processed (2024)

A majority of respondents (56.10%) considered the instrument effective, whereas 29.27% viewed it as ineffective and 14.63% were uncertain as shown in Table 7. Although more than half regard the instrument as effective, the substantial proportion of respondents who rated it ineffective or expressed uncertainty signals potential weaknesses in the instrument's construct clarity or its practical application during training.

Need for Instrument Revision

The majority of respondents indicated that the current instrument requires refinement.

Table 8. Perceptions of the Need to Revise the Assessment Instrument

Category	Frequency	Percentage
Needs Revision	28	68.29%
Does Not Need Revision	11	26.83%
Do Not Know	2	4.88%
Total	41	100%

Source: Data Processed (2024)

The need for instrument revision was strongly indicated by respondents. As reflected in Table 8, 68.29% of facilitators believed the instrument requires improvement, while 26.83% stated it does not, and 4.88% were unsure. This finding suggests a widespread perception among facilitators that the current behavioral assessment instrument does not fully align with the behavioral realities observed during training and requires modification to enhance its accuracy and usability.

3.2. Discussion

The findings of this study reveal several important insights into the effectiveness of the behavioral attitude assessment instrument used in structural leadership training. The discussion in this section interprets these insights in relation to theoretical frameworks and prior research, focusing on four key dimensions of instrument quality: validity, reliability, objectivity, and practicality. This section also synthesizes facilitators' perceptions regarding the instrument's effectiveness and the strong calls for revision expressed by respondents.

Validity

The study shows that although facilitators generally understand the behavioral dimensions being assessed, the structure of the instrument itself contributes to validity concerns. The merging of cooperation and initiative into a single assessment item is a clear example of construct underrepresentation. These two behaviors are conceptually distinct—cooperation arising from interpersonal collaboration and initiative emerging from independent action—yet the current instrument treats them as a single construct. This finding aligns with Organ et al.'s (2018) distinction between organizational citizenship behaviors driven by communal norms and individual proactive behaviors. Because the instrument conflates these constructs, it risks failing to measure each behavior accurately, thereby reducing construct validity.

Furthermore, the discrepancy between facilitators' interpretations of indicators and the official indicators provided in the LAN Head Decree reflects a lack of alignment between operational definitions and assessor understanding. This mismatch suggests potential weaknesses in content validity because indicators intended to capture specific behavioral nuances are not consistently recognized or applied by facilitators. This reinforces Arikunto's

(2018) assertion that validity is not achieved merely by providing indicators but by ensuring that assessors interpret and apply them as intended.

Reliability

The study highlights substantial variations in facilitators' interpretations of behavioral indicators and in their timing of assessments. These findings suggest challenges related to inter-rater reliability. Delays in scoring—especially when facilitators rely on memory rather than contemporaneous notes—introduce recall bias, which is consistent with Fraenkel et al.'s (2019) findings that temporal gaps reduce rating consistency. Moreover, the lack of shared interpretation of indicators contributes to scoring variability, undermining the reliability of measurement outcomes. These issues reflect a broader need for rater calibration, a technique widely recommended in assessment research to enhance inter-rater agreement.

Objectivity

Nearly half of the respondents acknowledged that their assessments were not consistently objective. This is consistent with Mulyasa's (2019) argument that objectivity is threatened when assessors rely on subjective impressions rather than standardized criteria. The absence of detailed observation notes, combined with varying interpretations of indicators, increases the likelihood of assessor bias. The findings suggest that objectivity cannot be ensured unless facilitators have both clear behavioral criteria and systematic observation procedures. Without such safeguards, personal preferences, interpersonal relationships, or selective memory can influence assessment outcomes.

Practicality

Although the results indicate that facilitators generally find the instrument easy to use, practicality alone does not guarantee effective assessment. Practicality must be accompanied by conceptual clarity, consistent interpretation of indicators, and feasible observation procedures. In this case, the high number of indicators per behavioral dimension—typically four to six—creates challenges in real training environments. Facilitators may struggle to observe all indicators simultaneously, particularly in large class settings or asynchronous learning scenarios. These practical constraints affect the usability of the instrument and contribute to inconsistent assessment practices.

Overall Effectiveness and Respondents' Call for Instrument Revision

The combination of validity, reliability, objectivity, and practicality concerns suggests that the instrument is only partially effective. While 56.10% of respondents considered the instrument effective, a significant proportion—nearly one-third—perceived it as ineffective, and an additional 14.63% were unsure. The uncertainty expressed by respondents is particularly meaningful, as it indicates a lack of confidence in the clarity and utility of the instrument. The finding that 68.29% of respondents believe the instrument requires revision strongly reinforces this concern. This suggests a collective recognition among facilitators that the instrument does not adequately capture behavioral realities in structural leadership training and that improvements are needed to enhance its accuracy, clarity, and ease of use.

Theoretical and Practical Implications

Theoretical implications arise from the study's demonstration that the quality of an evaluation instrument is foundational to the effectiveness of training evaluation models. If the behavioral measurement tool lacks validity or reliability, the interpretation of behavior-level outcomes becomes questionable, limiting the accuracy of subsequent assessments at the results level. This supports the argument that instrument quality must be considered an essential prerequisite for meaningful evaluation.

Practically, the findings emphasize the need for clearer indicators, improved rater training, and standardized observation guidelines. Institutions offering structural leadership training may also benefit from multi-rater evaluation systems and digital observation tools that can reduce recall bias and enhance scoring consistency. These improvements would strengthen the

assessment process and ensure that behavioral evaluations reflect actual participant performance more accurately.

4. Conclusion

The findings of this study demonstrate that the behavioral attitude assessment instrument used in structural leadership training has not yet achieved an optimal level of effectiveness. Although the instrument is practical and relatively easy to apply, its effectiveness is constrained by weaknesses in validity, reliability, and objectivity. These limitations are primarily shaped by inconsistent interpretation of behavioral indicators, the absence of standardized observation guidelines, and assessment practices that rely heavily on facilitator subjectivity. To establish a truly performance- and behavior-based evaluation system, the instrument must be transformed into an evidence-based behavioral assessment framework in which each indicator is operationalized through clear criteria and supported by documented empirical observations. Such an approach would not only enhance inter-rater consistency but also strengthen the credibility and fairness of behavioral evaluation outcomes. Overall, this study provides empirical insight into current behavioral evaluation practices in structural leadership training and offers a clear direction for the development of assessment instruments that are more scientific, measurable, and aligned with the evolving demands of structural leadership roles.

References

- Ajzen, I. (2015). *The Theory of Planned Behavior*. dalam *The Oxford Handbook of Human Motivation*. Oxford University Press.
- Arikunto, S. (2018). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Rineka Cipta.
- Azwar, S. (2018). *Reliabilitas dan Validitas*. Pustaka Pelajar.
- Bates, R. (2020). *Training Evaluation in Organizations: Theory and Practice*. Routledge.
- Fishbein, M. dan Ajzen, I. (2010). *Predicting and Changing Behavior: The Reasoned Action Approach*. Psychology Press.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to Design and Evaluate Research in Education 10th Edition*. McGraw-Hill Education.
- Katz, D., & Stotland, E. (2014). *The measurement of attitude*. Dalam *The Psychology of Attitudes*. Routledge.
- Keputusan Kepala Lembaga Administrasi Negara Nomor: 2/K.1/PDP.07/2023 Tentang Pedoman Penyelenggaraan Pelatihan Struktural Kepemimpinan
- Kibe, J. M., & Namusonge, G. (2015). *Effect of Training Evaluation on Employee Performance in Public Sector Organizations*. *International Journal of Academic Research in Business and Social Sciences*, 5(6), 2222–6990.
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2016). *Evaluating Training Programs: The Four Levels*. Berrett-Koehler Publishers.
- Lembaga Administrasi Negara. (2022). *Peraturan LAN Nomor 5 Tahun 2022 tentang Penyelenggaraan Pelatihan Kepemimpinan Struktural*.
- Lembaga Administrasi Negara. (2023). *Keputusan Kepala LAN Nomor 2/K.1/PDP.07/2023 tentang Pedoman Penyelenggaraan Pelatihan Struktural Kepemimpinan*.
- Mulyasa, E. (2019). *Manajemen dan Evaluasi Pendidikan*. Remaja Rosdakarya.
- OECD. (2019). *Public Sector Leadership for the 21st Century*. OECD Publishing.
- Organ, D. W., Podsakoff, P. M., & MacKenzie, S. B. (2018). *Organizational Citizenship Behavior: Its Nature, Antecedents, and Consequences*. Sage.
- Phillips, J. J., & Phillips, P. P. (2016). *Evaluation and the Training Professional: A Practical Guide for the Workplace*. Association for Talent Development.
- Rahman, M., Karim, M., & Islam, M. (2021). Leadership Behavior Development through Training Evaluation. *International Journal of Public Administration*, 44(3), 235–247.