

Article

Impact of Behavioral Change on Performance Improvement: A Longitudinal Analysis Using Personnel Appraisal Data

Masakazu Asano * and Masayuki Goto 

School of Creative Science and Engineering, Waseda University, Tokyo 169-8555, Japan

* Correspondence: m.asano@fuji.waseda.jp

Featured Application

The proposed framework can be applied to organizational appraisal systems to identify behavioral domains associated with performance improvement. By leveraging longitudinal personnel data, organizations can quantitatively prioritize skill development areas and design more effective training and evaluation strategies. This approach supports data-driven decision making in human resource management and talent development.

Abstract

This study examines whether within-person changes in behavioral evaluation factors are associated with performance improvement using longitudinal personnel appraisal data from a Japanese firm (fiscal years 2019–2021, $N = 218$). Fourteen behavioral appraisal items were standardized annually and analyzed using exploratory factor analysis, yielding three factors: Comprehensive Behavior, Job Skills, and Discipline/Values. Behavioral change was defined as the difference in factor scores between FY2021 and FY2019. Hierarchical cluster analysis identified four typical behavioral change patterns. Binary logistic regression results indicate that changes in Job Skills are significantly associated with a higher likelihood of performance improvement (odds ratio = 1.33, 95% confidence interval = 1.02–1.75, $p = 0.037$). Additional sensitivity analyses controlling for baseline performance and baseline factor levels supported the positive association between changes in Job Skills and performance improvement. These findings suggest that improvement in job-related skills, rather than overall behavioral level, is associated with short- to medium-term performance improvement. The study provides a change-based analytical framework that leverages longitudinal appraisal data to identify behavioral domains linked to performance improvement, thereby contributing to data-driven human resource management and employee development.

Keywords: behavioral appraisal; behavioral change; longitudinal data; personnel evaluation; factor analysis; cluster analysis; logistic regression



Received: 22 May 2026

Revised: 17 June 2026

Accepted: 22 June 2026

Published: 24 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Improving employee performance is a critical managerial priority because employee performance is closely related to organizational productivity, service quality, and long-term competitiveness. From the perspective of human resource management (HRM), the practical challenge is not only to evaluate employees but also to identify which behavioral domains should be developed through training, feedback, and performance-management interventions. Personnel appraisal systems provide one of the few regularly collected organizational data sources for this purpose, and the use of such records is consistent with

evidence-based management and HR analytics, which emphasize the systematic use of organizational data for managerial decision making (DeNisi & Pritchard, 2006; Marler & Boudreau, 2017; Murphy & Cleveland, 1995; Rousseau, 2006). At the same time, appraisal scores may be influenced by rater differences, measurement error, and common method bias (Podsakoff et al., 2003; Viswesvaran et al., 1996). Therefore, appraisal data should be analyzed carefully as longitudinal organizational measurement records rather than treated merely as static evaluation outcomes.

Relationships between behavioral characteristics and performance have been widely discussed in appraisal research, organizational behavior, and HRM research. Many prior studies estimate associations between behavioral levels and performance levels using cross-sectional data (Dang et al., 2025; Kahya, 2007; Kuvaas, 2008). Such studies are useful for identifying characteristics of high performers, but they do not directly answer a key managerial question: which behavioral domains actually change within employees and whether such changes are associated with subsequent performance improvement. This distinction matters because development interventions are designed to change employee behavior, not merely to classify employees according to static characteristics. Longitudinal analysis of within-person change can address this issue (Sturman, 2005), yet empirical studies that organize multi-item appraisal records into latent behavioral domains and examine whether changes in those domains are associated with performance improvement remain limited.

Accordingly, the present study addresses the following research question: which within-person changes in behavioral evaluation factors are associated with performance improvement? Using actual data from a longitudinal performance evaluation conducted at a Japanese company, the study (1) extracts relatively stable behavioral factors from 14 behavioral appraisal items, (2) defines behavioral change as the difference in factor scores between FY2019 and FY2021 and summarizes typical change patterns through cluster analysis, and (3) estimates the association between behavioral change and the probability of performance improvement by binary logistic regression. The contribution of this study is threefold. First, it shifts the analytical focus of appraisal research from static behavioral levels to within-person behavioral change. Second, it integrates factor extraction, behavioral change scores, clustering, and regression into a change-based HR analytics framework. Third, it provides practical evidence for prioritizing employee development by identifying behavioral domains that are most closely associated with performance improvement. The purpose is not to claim a universal law from a single-company dataset, but to demonstrate an empirical framework for using longitudinal appraisal data to support evidence-based HRM and employee development.

2. Related Work

2.1. Behavioral Appraisal and Job Performance

Job performance is not a unitary construct; it has been conceptualized as including task performance and contextual performance (Borman & Motowidlo, 1997; Campbell et al., 1993). Organizational citizenship behavior (OCB) has also been discussed as discretionary behavior that contributes to organizational effectiveness beyond formal job requirements (Organ, 1988), and many studies have examined its dimensionality and performance implications (LePine et al., 2002; Podsakoff et al., 2000). When discussing the relationship between behavioral appraisal and performance outcomes, it is therefore necessary to consider which aspects of performance—task or contextual—the appraisal items are likely to capture (Rotundo & Sackett, 2002). In addition, behaviors related to task execution can be interpreted from the perspective of job characteristics theory, which suggests that job design and job characteristics influence behavior through motivation and internal

states (Hackman & Oldham, 1980). This perspective implies that behaviors included in the Job Skills factor, such as planning ability, problem solving, information use, and job-specific skills, are likely to be closely linked to the execution of assigned tasks and may therefore have a more immediate association with performance improvement than broader behavioral orientations.

2.2. Performance Appraisal Systems and Operational Issues

Performance appraisal is influenced not only by formal system design but also by operational factors such as rater standards, measurement error, and employees' acceptance of the process (DeNisi & Pritchard, 2006; Murphy & Cleveland, 1995). Meta-analytic evidence also shows that participation in the appraisal process affects reactions to appraisal and perceived acceptance (Cawley et al., 1998). Moreover, the reliability of appraisal ratings can be affected by variation among raters and rating contexts (Viswesvaran et al., 1996). Thus, when appraisal data are analyzed as research data, comparability across years, interpretability of scores, and possible rater or organizational-context effects become critical issues.

In the Japanese context, perceptions of fairness in performance appraisal and the acceptance of appraisal systems have been shown to influence employees' attitudes and behavior (Hayashi & Takahashi, 2002; Takahashi, 2001). Recent studies have also examined the relationship between internal reputation and personnel appraisal, suggesting that appraisal outcomes are shaped not only by objective measurement but also by organizational interactions and shared perceptions (Hattori et al., 2022). These considerations indicate that appraisal scores should be interpreted in light of their institutional and operational context.

2.3. Longitudinal Research and Within-Person Change

Longitudinal data allow researchers to examine how performance-related attributes change within the same individual over time, enabling the analysis of dynamic processes that are difficult to capture with cross-sectional comparisons. Within-person variability in job performance has been emphasized in prior research (Sturman, 2005). Feedback and development mechanisms have also been examined over time in relation to goal accomplishment, self-evaluation, and performance-related outcomes (London & Smither, 1995). The Ability–Motivation–Opportunity (AMO) framework likewise provides a useful lens for understanding how behavioral change can connect employee attributes to performance outcomes (Appelbaum et al., 2000; Jiang et al., 2012).

However, empirical studies using multi-item organizational appraisal data seldom integrate factor extraction, calculation of factor-score changes, estimation of performance-improvement probabilities, and typology construction through clustering in a single analytical framework. Prior related studies have examined stepwise performance improvement and transition patterns in personnel evaluation data (Asano et al., 2026a, 2026b), but they did not explicitly test whether within-person changes in behavioral evaluation factors are associated with performance improvement. The present study therefore positions behavioral change as the central explanatory construct and examines its relationship with performance improvement at the employee level.

2.4. Research Question and Analytical Expectations

Based on the above literature, this study asks: which within-person changes in behavioral evaluation factors are associated with performance improvement? In line with the AMO framework and job characteristics theory, we expected changes in the Job Skills factor to show the most direct positive association with performance improvement because this factor reflects ability-related behaviors connected to task execution. Changes in the Comprehensive Behavior and Discipline/Values factors were examined as additional behavioral

domains, but their expected relationships with short- to medium-term performance improvement were treated as more exploratory because they may influence performance through broader organizational and contextual mechanisms.

3. Materials and Methods

3.1. Analytical Framework

This study is an empirical investigation of how employee behavioral change is associated with performance change using longitudinal personnel appraisal data. The theoretical logic is grounded in the AMO framework, which posits that employee performance is shaped by ability, motivation, and opportunity expressed through work behavior (Appelbaum et al., 2000; Jiang et al., 2012). In this study, the Job Skills factor is interpreted as an ability- and task-execution-related domain because it includes planning, problem solving, information use, and job-specific skills. The Comprehensive Behavior factor captures proactive and collaborative behavioral orientation, whereas the Discipline/Values factor captures value adherence and rule-following behavior. Based on this logic, the analysis expects changes in the Job Skills factor to show the most direct association with short- to medium-term performance improvement, while the associations of the other factors are examined empirically. The purpose of the study is not to establish a general causal mechanism, but rather to quantify the association between behavioral change and performance improvement observed in archival organizational appraisal data.

3.2. Data Description

This study uses actual data from a longitudinal performance evaluation conducted at Company A, a Japanese firm, for FY2019–FY2021. The dataset was anonymized prior to analysis. The appraisal system consists of three performance evaluation items and fourteen behavioral evaluation items. De-identified employee IDs were used to link records across years, and the analysis sample includes 218 employees who remained with the company for all three years. All items were rated on a five-point absolute scale (1 = markedly insufficient, 5 = exceptionally strong). Behavioral scores were standardized within each year (mean = 0, standard deviation = 1), and factor scores were calculated based on these standardized values. Behavioral change was defined as the difference in factor scores between FY2021 and FY2019. The dataset did not include rater IDs, department, job family, tenure, or individual training histories. Therefore, these potential rater and contextual confounds could not be directly modeled and are treated as important limitations of the study.

3.3. Behavioral Variables and Factor Analysis

To capture employee behavior multidimensionally, we analyzed the 14 behavioral appraisal items included in the company's appraisal form: corporate values, discipline, organizational contribution, proactiveness, planning ability, execution, problem solving, information skills, leadership, talent development, dialogue, and three job-specific skill items. Exploratory factor analysis was conducted separately for FY2019, FY2020, and FY2021. Within each year, the 14 behavioral items were standardized to mean zero and standard deviation one, and the factor structure was extracted using the principal factor method with varimax rotation. Factor scores were estimated using the regression method. The directions of the factor scores were aligned so that higher values indicated stronger expression of the corresponding behavioral domain.

To ensure interpretability of year-to-year differences, factors extracted from different years were matched on the basis of their major loading patterns before change scores were computed. As a reliability check, internal consistency was calculated for the main item groups corresponding to the three factors. Cronbach's alpha values for FY2019, FY2020,

and FY2021 were 0.78, 0.82, and 0.79 for Comprehensive Behavior; 0.61, 0.60, and 0.58 for Job Skills; and 0.57, 0.70, and 0.62 for Discipline/Values, respectively. These values suggest acceptable reliability for the broad Comprehensive Behavior factor and more moderate reliability for the narrower Job Skills and Discipline/Values factors. Detailed factor loadings are reported in Section 4.

3.4. Definition of Behavioral and Performance Change

Behavioral change was defined as the difference in factor scores between FY2021 and FY2019. Let $n = 1, \dots, N$ index employees, where $N = 218$. Let $i = 1, 2$, and 3 index the behavioral factors, and let $t \in \{2019, 2020, 2021\}$ index fiscal years. Let $F_{n,i,t}$ denote the factor score for employee n , factor i , and year t . The three factors correspond to the Comprehensive Behavior factor, the Job Skills factor, and the Discipline/Values factor, respectively.

$$\Delta F_{n,i} = F_{n,i,2021} - F_{n,i,2019}, i = 1, 2, 3. \quad (1)$$

This definition captures whether each behavioral domain improved, remained stable, or deteriorated between FY2019 and FY2021. Factor scores for FY2020 were used to confirm year-to-year patterns but were not included in the calculation of $\Delta F_{n,i}$. This two-point comparison is suitable for capturing the relationship between behavioral change and performance change in a transparent manner.

Let $P_{n,t}$ denote the total performance score for employee n in year t . Performance change for employee n was defined as follows:

$$\Delta P_n = P_{n,2021} - P_{n,2019}. \quad (2)$$

The binary outcome variable Y_n was coded as 1 if $\Delta P_n > 0$ and 0 otherwise. Thus, the analysis dataset for the regression model was defined as $\{(Y_n, \Delta F_{n,1}, \Delta F_{n,2}, \Delta F_{n,3}) \mid n = 1, \dots, 218\}$.

Examination of the distributions of the behavioral change variables showed no extreme distortion. Because behavioral scores were standardized within each year before factor scores were calculated, $\Delta F_{n,i}$ reflects changes in relative standing on the standardized factor scores rather than absolute changes on the original raw scale.

3.5. Cluster Analysis

To explore major patterns of behavioral change, hierarchical cluster analysis was conducted using ΔF_1 , ΔF_2 , and ΔF_3 as clustering variables. Ward's method with Euclidean distance was adopted. The number of clusters was set to four based on the dendrogram, cluster-size balance, and interpretability.

3.6. Logistic Regression Analysis

Because the dependent variable was binary, binary logistic regression was used to estimate the association between behavioral change and performance improvement (Agresti, 2010). The explanatory variables were the three employee-level behavioral change scores defined in Equation (1), denoted for readability as $\Delta F_{n,1}$, $\Delta F_{n,2}$, and $\Delta F_{n,3}$. The model is expressed as follows:

$$\log\left(\frac{p_n}{1-p_n}\right) = \beta_0 + \beta_1\Delta F_{n,1} + \beta_2\Delta F_{n,2} + \beta_3\Delta F_{n,3} \quad (3)$$

where $p_n = P(Y_n = 1 \mid \Delta F_{n,1}, \Delta F_{n,2}, \Delta F_{n,3})$

The estimation results are reported in Table 1 and interpreted in Section 4.4.

Table 1. Effects of Behavioral Change on Performance Improvement: Binary Logistic Regression Results.

Variable	B	SE	Wald χ^2	p-Value	Odds Ratio	95% CI for OR
Constant	−0.482	0.142	11.54	0.001	0.62	0.47–0.82
$\Delta F1$: Comprehensive Behavior Factor	0.140	0.151	0.86	0.353	1.15	0.86–1.55
$\Delta F2$: Job Skills Factor	0.287	0.138	4.34	0.037	1.33	1.02–1.75
$\Delta F3$: Discipline/Values Factor	−0.206	0.173	1.42	0.233	0.81	0.58–1.14

Note: The dependent variable is performance improvement (1 = the total performance score increased from FY2019 to FY2021; 0 = otherwise). The explanatory variables are $\Delta F1$, $\Delta F2$, and $\Delta F3$, defined as employee-level factor-score differences between FY2021 and FY2019. Parameters were estimated by maximum likelihood. OR = odds ratio; CI = confidence interval. $n = 218$.

3.7. Methodological Considerations

This study uses real organizational data. Although the sample size is limited, the dataset has the strength of tracking the same individuals over three years. The exploratory factor model was supplemented by confirmatory factor analysis, which yielded acceptable fit indices (CFI = 0.92, RMSEA = 0.05), supporting the adequacy of the factor structure. In addition, yearly standardization was applied before calculating factor scores to reduce the influence of year-specific differences in score distributions and to improve comparability across years. To address concerns about regression to the mean and baseline differences, additional robustness checks were conducted using models that included baseline performance and baseline factor levels. Sensitivity analyses were also conducted using a stricter performance-improvement threshold and a continuous performance-change outcome.

At the same time, some limitations of the measurement approach should be noted. Factor analysis was conducted separately by year, and full measurement invariance was not statistically tested using multi-group confirmatory factor analysis. Accordingly, the behavioral change variables should be interpreted as changes in standardized factor scores rather than absolute changes on the original appraisal scale. The dataset also did not include rater IDs or detailed contextual variables such as department, job family, tenure, promotion, or training participation. Consequently, rater leniency, rater drift, department-specific conditions, and changes in work arrangements during the COVID-19 period could not be directly controlled. These points are explicitly considered when interpreting the results.

4. Results

4.1. Factor Analysis Results

Exploratory factor analysis of the 14 behavioral appraisal items identified three factors with eigenvalues greater than 1.0. Factors with eigenvalues below 1.0 were not retained because they contributed limited additional explanatory power and were difficult to interpret substantively. The same three conceptual factors were obtained across the three fiscal years based on the major loading patterns, although full measurement invariance was not formally tested. Therefore, the factor structure is treated as sufficiently stable for exploratory longitudinal analysis, but not as evidence of strict scale invariance.

Table 2 shows the factor loadings for FY2019 obtained from the factor analysis. Similar three-factor solutions were obtained for FY2020 and FY2021, and the composition of the major loading items was broadly consistent across years. The reliability coefficients reported in Section 3.3 further indicate that the factors had reasonable internal consistency, although the Job Skills and Discipline/Values factors should be interpreted with some caution because they contain fewer and more specific appraisal items.

Table 2. Factor Analysis Results for the Behavioral Appraisal Items in FY2019 (Principal Factor Method with Varimax Rotation).

Behavioral Item	Factor 1 Comprehensive Behavior	Factor 2 Job Skills	Factor 3 Discipline/Values
Proactiveness	0.82	0.18	0.09
Execution	0.79	0.22	0.08
Dialogue	0.77	0.19	0.12
Leadership	0.74	0.26	0.11
Organizational Capability	0.71	0.33	0.10
Talent Development	0.67	0.28	0.14
Planning Ability	0.63	0.48	0.16
Problem-Solving Ability	0.58	0.61	0.13
Information Skills	0.29	0.72	0.11
Job-Specific Skill 1	0.32	0.70	0.14
Job-Specific Skill 2	0.28	0.68	0.17
Job-Specific Skill 3	0.24	0.65	0.19
Corporate Values	0.12	0.17	0.81
Discipline	0.10	0.15	0.78
Eigenvalue	5.07	3.01	1.81
Variance Explained (%)	36.2	21.5	12.9
Cumulative Variance Explained (%)	36.2	57.7	70.6

Note: Principal factor method with varimax rotation. $n = 218$. Values indicate factor loadings. Similar three-factor solutions were obtained for FY2020 and FY2021.

Factor 1 showed high loadings on proactiveness, execution, dialogue, and leadership, representing proactive and collaborative behavior. This factor was therefore labeled the Comprehensive Behavior factor ($\Delta F1$).

Factor 2 showed high loadings on planning ability, problem solving, and information skills and was interpreted as the Job Skills factor ($\Delta F2$).

Factor 3 showed high loadings on corporate values and discipline and was labeled the Discipline/Values factor ($\Delta F3$).

The contribution rates were 36.2% for Factor 1, 21.5% for Factor 2, and 12.9% for Factor 3, with a cumulative contribution rate of 70.6%, indicating that the three-factor solution summarized the structure of the behavioral appraisal items reasonably well for the purpose of exploratory organizational analysis.

4.2. Distribution of Behavioral Change

Using the behavioral change scores defined in Section 3.4, we examined how employee behavior changed from FY2019 to FY2021. Descriptive statistics for the three change variables are reported in Table 3. Because the behavioral items were standardized within each year before factor scores were computed, the means of the change variables are naturally close to zero. The focus of the analysis is therefore not the group-level average increase or decrease on the original appraisal scale, but the magnitude of within-person changes in relative standing on standardized factor scores.

Among the three factors, $\Delta F2$ (Job Skills) showed the largest standard deviation and the widest range, suggesting substantial heterogeneity across employees in the degree of relative improvement or decline in job-related skills. $\Delta F3$ (Discipline/Values) had a mean near zero but included some employees with a large negative relative change. $\Delta F1$ (Comprehensive Behavior) showed an intermediate degree of dispersion.

Table 3. Descriptive Statistics of Behavioral Change Scores.

Variable	Mean	Standard Deviation	Minimum	Maximum
$\Delta F1$: Comprehensive Behavior Factor	0.00	0.95	−2.97	2.83
$\Delta F2$: Job Skills Factor	0.00	1.08	−3.18	3.30
$\Delta F3$: Discipline/Values Factor	0.00	0.84	−4.27	2.17

Note: For readability, ΔFi denotes the employee-level factor-score difference $\Delta F_{n,i}$ between FY2021 and FY2019. Factor scores were calculated after annual standardization of the 14 behavioral appraisal items. $n = 218$.

4.3. Typology of Behavioral Change Patterns

The cluster analysis yielded four behavioral change patterns. Table 4 reports the number of employees, the number of employees with performance improvement, the cluster composition ratio, and the performance-improvement rate for each cluster. Table 5 reports the mean behavioral change scores within each cluster.

Table 4. Cluster Composition and Performance Improvement Rates.

Cluster	Cluster Size (n)	Number of Employees with Performance Improvement (k)	Composition Ratio (%)	Performance Improvement Rate (%)
1	29	17	13.3	58.6
2	49	18	22.5	36.7
3	25	10	11.5	40.0
4	115	39	52.8	33.9
Total	218	84	100.0	38.5

Note: Cluster analysis was conducted using Ward's method with Euclidean distance. Performance improvement indicates an increase in the total performance score from FY2019 to FY2021. The composition ratio represents the proportion of each cluster in the sample. Percentages may not sum exactly to 100.0 due to rounding.

Table 5. Mean Behavioral Change Scores by Cluster.

Cluster	$\Delta F1$: Comprehensive Behavior Factor	$\Delta F2$: Job Skills Factor	$\Delta F3$: Discipline/Values Factor
1	0.068	1.273	−1.105
2	−0.865	0.922	0.199
3	1.549	0.009	−0.212
4	0.015	−0.716	0.240

Note: For readability, ΔFi denotes the employee-level factor-score difference $\Delta F_{n,i}$ between FY2021 and FY2019. Values represent mean behavioral change scores within each cluster; positive values indicate improvement, whereas negative values indicate decline.

Cluster 1 ($n = 29$) showed a large positive mean value for $\Delta F2$ (mean = 1.273) and had the highest performance-improvement rate (58.6%). Cluster 2 ($n = 49$) also showed a positive mean for $\Delta F2$ but a negative mean for $\Delta F1$ (mean = −0.865), and its performance-improvement rate was 36.7%. Cluster 3 ($n = 25$) showed a large positive mean for $\Delta F1$ (mean = 1.549) and a performance-improvement rate of 40.0%. Cluster 4 ($n = 115$) showed a negative mean for $\Delta F2$ (mean = −0.716) and had a performance-improvement rate of 33.9%.

These results suggest that both behavioral change patterns and performance-improvement rates differ across clusters. However, the omnibus chi-square test for differences in performance-improvement rates across the four clusters was not statistically significant at the 5% level, $\chi^2(3) = 6.07$, $p = 0.108$. Therefore, the cluster analysis should be interpreted primarily as an exploratory typology of behavioral change patterns rather than as a confirmatory test of performance differences. To test the relationship at the individual level, the next subsection estimates the association between behavioral change and performance improvement using logistic regression.

Figure 1 presents a scatter plot of $\Delta F1$ (Comprehensive Behavior) and $\Delta F2$ (Job Skills), illustrating the distribution of employee-level behavioral change scores across clusters. The clusters show distinct patterns in the behavioral change space, particularly along the Job Skills factor and the Comprehensive Behavior factor. This visualization supports the interpretation that the clusters represent qualitatively different behavioral change patterns rather than arbitrary partitions. In particular, Cluster 1 is characterized by a relatively high positive change in the Job Skills factor, whereas Cluster 4 is characterized by a negative mean change in the Job Skills factor.

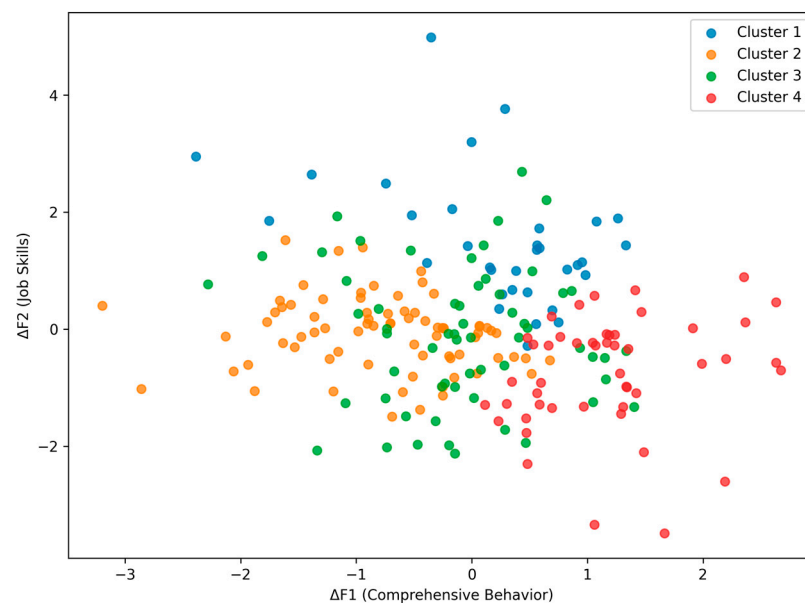


Figure 1. Distribution of Behavioral Change Scores by Cluster.

Figure 2 presents a bar chart of performance-improvement rates across the four behavioral change clusters. The figure shows that Cluster 1 had the highest performance-improvement rate, whereas Cluster 4, characterized by a negative mean change in the Job Skills factor, had a lower performance-improvement rate. This pattern is consistent with the logistic regression result that the change in the Job Skills factor is positively associated with performance improvement.

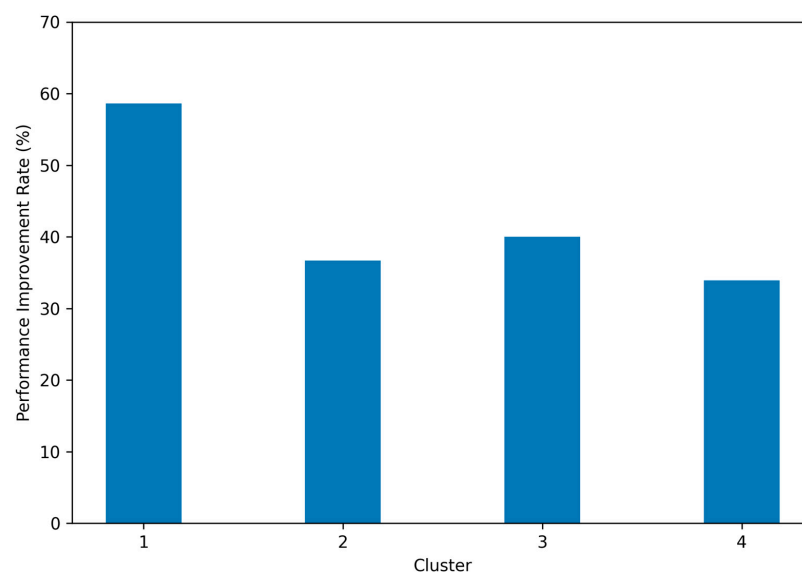


Figure 2. Performance Improvement Rate by Cluster.

4.4. Association Between Behavioral Change and Performance Improvement

Table 1 reports the estimation results of the binary logistic regression. Only $\Delta F2$, the change in the Job Skills factor, showed a statistically significant positive association with performance improvement ($B = 0.287$, odds ratio = 1.33, 95% CI = 1.02–1.75, $p = 0.037$). In substantive terms, holding the other change variables constant, a one-unit increase in $\Delta F2$ increased the odds of performance improvement by approximately 1.33 times. At the mean of the other change variables, the predicted probability of performance improvement increased from approximately 38.2% to 41.6% for a 0.5-unit increase in $\Delta F2$ and to 45.1% for a 1.0-unit increase. Thus, the effect is modest but practically interpretable.

By contrast, neither $\Delta F1$ (Comprehensive Behavior) nor $\Delta F3$ (Discipline/Values) reached statistical significance. Thus, within the present model, changes in these factors were not significantly associated with short- to medium-term performance improvement.

Regarding model fit, the likelihood ratio test was marginally significant ($p < 0.10$), and McFadden's pseudo- R^2 was 0.026. These values indicate that the explanatory power of the parsimonious model is limited. Therefore, the regression result should be interpreted as evidence of a limited but present association between improvement in job-related behavioral skills and performance improvement, not as a strong predictive model or a causal explanation.

4.5. Robustness and Sensitivity Analyses

To address concerns about regression to the mean, outcome dichotomization, and baseline differences, additional robustness checks were conducted. First, a baseline-adjusted logistic model was estimated by adding FY2019 performance and FY2019 factor levels to the three behavioral change variables. Second, the dependent variable was redefined using a stricter threshold, coded as 1 only when the performance score increased by two or more points. Third, a linear regression model was estimated using continuous performance change as the dependent variable. As shown in Table 6, the change in the Job Skills factor remained positive in all robustness specifications and statistically significant in the baseline-adjusted logistic model, the stricter-threshold model, and the continuous-change model. These results support the interpretation that the association between Job Skills improvement and performance improvement is not solely an artifact of the original binary threshold or baseline performance differences.

Table 6. Robustness Checks for the Association Between Job Skills Change and Performance Improvement.

Model	Outcome	Additional Controls	Effect of $\Delta F2$	p -Value	Model Fit
Main logistic model	$\Delta P > 0$	$\Delta F1, \Delta F2, \Delta F3$	OR = 1.33	0.037	McFadden $R^2 = 0.026$
Baseline-adjusted logistic model	$\Delta P > 0$	+ P2019 and baseline factor levels	OR = 2.15	<0.001	McFadden $R^2 = 0.174$
Stricter-threshold logistic model	$\Delta P \geq 2$	+ P2019 and baseline factor levels	OR = 1.82	0.003	McFadden $R^2 = 0.175$
Continuous-change model	Continuous ΔP	+ P2019 and baseline factor levels	B = 0.790	<0.001	$R^2 = 0.352$

Note: Robustness models were estimated using the employee-level longitudinal dataset. Baseline controls include FY2019 performance and FY2019 levels of the three behavioral factors. The continuous-change model uses the raw difference in performance score between FY2021 and FY2019 as the dependent variable. OR = odds ratio; ΔP = performance score difference.

5. Discussion

5.1. Main Findings

This study used longitudinal personnel appraisal data from FY2019 to FY2021 to examine whether changes in behavioral evaluation factors are associated with performance improvement. Exploratory factor analysis of the 14 behavioral appraisal items identified a three-factor structure consisting of Comprehensive Behavior, Job Skills, and Discipline/Values, confirming that the behavioral appraisal data have a multidimensional structure. Cluster analysis of the behavioral change scores organized employees into four typical change patterns. These clusters provided a descriptive visualization of different behavioral change patterns, but the cluster-level differences in performance-improvement rates should be interpreted cautiously because the omnibus test was not statistically significant at the 5% level.

The logistic regression analysis showed that only the change in the Job Skills factor was significantly associated with performance improvement in the main model. This result suggests that improvement in behaviors directly related to job execution may be more closely linked to short- to medium-term performance gains than broader behavioral orientation or discipline/value-related behavior. However, the effect size was modest and the model fit was limited; therefore, the finding should be interpreted as an association observed in this dataset rather than as a strong predictive or causal relationship.

5.2. Theoretical Implications

The findings offer several implications for appraisal research and human resource management research.

First, this study focuses on within-person behavioral change rather than static behavioral levels. Whereas much of the existing literature examines the association between behavioral level and performance level using cross-sectional data, this study shows that appraisal records can be reorganized to examine the improvement process itself. This contributes to HR analytics by demonstrating how routinely collected appraisal data can be used to construct change-oriented indicators for employee development.

Second, the fact that only the change in the Job Skills factor was significant is theoretically consistent with the AMO framework. The Job Skills factor includes planning, problem solving, information use, and job-specific skills, all of which are closely related to ability and task execution. Improvements in these behaviors are therefore more likely to affect performance ratings within a short- to medium-term time horizon than changes in broader collaborative orientation or discipline/value adherence. This interpretation helps explain why $\Delta F2$, rather than $\Delta F1$ or $\Delta F3$, showed the clearest association with performance improvement.

Third, this study demonstrates that existing appraisal records can be used not only as static evaluation results but also as longitudinal data for capturing behavioral change processes. This expands the analytical value of organizational appraisal data and shows how archival HR data can contribute to evidence-based management when interpreted together with measurement limitations and contextual constraints.

5.3. Practical Implications for Human Resource Development and Appraisal

The findings also provide practical implications for personnel development and appraisal-system operation. The Job Skills factor, which was significantly associated with performance improvement, includes behaviors such as planning, problem solving, information utilization, and job-specific skill execution. These results suggest that appraisal feedback should not simply encourage general behavioral improvement but should identify concrete skill-related development goals. For example, employees showing limited

improvement in the Job Skills factor may benefit from targeted support in planning work processes, solving job-specific problems, using information effectively, and acquiring task-specific expertise. At the group level, the cluster-based typology can help HR managers identify employee segments that may require different development interventions, although such decisions should be supplemented by managerial judgment and additional contextual information.

5.4. Limitations and Future Research

This study has several limitations. First, the analysis is based on observational data and estimates association rather than strict causal effects. Second, the data come from a single company, and the sample size is limited to 218 employees. Third, although the factors were matched across years based on major loading patterns, full measurement invariance was not tested using multi-group confirmatory factor analysis. Therefore, the change scores should be interpreted as changes in standardized factor scores rather than absolute changes in the original appraisal scale. Fourth, the dataset did not include rater IDs, department, job family, tenure, promotion, or training participation. As a result, rater leniency, rater drift, department-level conditions, baseline differences in job assignment, and development opportunities could not be directly controlled. Fifth, the period FY2019–FY2021 includes the COVID-19 period, during which work arrangements and appraisal practices may have been affected. The present data did not provide sufficient information to model these policy or context changes directly.

The present study also uses a binary outcome model for performance improvement. While this is practically useful for managerial interpretation, future research could apply more detailed models, such as ordinal logit models, multilevel models, latent change score models, cross-lagged panel models, or dynamic structural equation models, to capture the magnitude, hierarchy, and temporal dynamics of performance change more precisely. In addition, incorporating rater characteristics, department-level context, job category, and training participation would improve understanding of the mechanisms through which behavioral change relates to performance outcomes. Because the latent structure of performance and appraisal ratings may vary across appraisal systems and organizational contexts (Scullen et al., 2000), future studies should examine whether the three-factor structure identified in this study can be replicated using data from other firms or industries.

Despite these limitations, this study makes a meaningful contribution by showing, using longitudinal appraisal data with a focus on behavioral transitions, that improvement in the Job Skills factor is consistently associated with performance improvement across the main and sensitivity analyses.

6. Conclusions and Future Work

This study examined the association between changes in behavioral evaluation factors and performance improvement using personnel appraisal data from a Japanese company collected over FY2019–FY2021. Three factors were extracted from 14 behavioral appraisal items, and changes in factor scores were used to analyze the relationship between behavioral change and performance improvement. The results showed that the change in the Job Skills factor had a statistically significant positive association with performance improvement in the main model, and this association was supported by additional sensitivity analyses that controlled for baseline performance and baseline factor levels. These findings suggest that performance gains may be more closely associated with improvements in behaviors directly related to job execution than with broad improvements in general behavioral orientation, although the explanatory power of the model is limited. Theoretically, the study contributes empirical evidence from a single-company longitudinal dataset by shifting the focus from

static behavioral level to behavioral change. Practically, it suggests that existing appraisal data can be restructured to identify priority domains for employee development in a quantitative manner.

As the study is based on a single-company dataset, caution should be exercised when generalizing the findings. Future research should test the framework using multi-company data, incorporate rater and job-context information, and apply more refined longitudinal models that capture continuous variation in performance change. Nevertheless, the case study demonstrates that accumulating annual performance appraisal data can help identify behavioral changes associated with improved employee performance and support more evidence-based HRM practices.

Author Contributions: Conceptualization, M.A. and M.G.; methodology, M.A. and M.G.; validation, M.A. and M.G.; formal analysis, M.A.; investigation, M.A.; resources, M.A.; data curation, M.A.; writing—original draft preparation, M.A.; writing—review and editing, M.A. and M.G.; supervision, M.G.; project administration, M.G.; funding acquisition, M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partly supported by JSPS KAKENHI Grant Number 24H00370.

Institutional Review Board Statement: Ethical review and approval were waived for this study because the analysis used de-identified archival personnel appraisal data provided by the company and did not involve direct interaction with human participants. The authors had no access to information that could directly identify individuals.

Informed Consent Statement: Informed consent was waived because the study used de-identified archival organizational data and did not involve direct interaction with employees.

Data Availability Statement: The data used in this study are not publicly available due to confidentiality restrictions, as they contain sensitive personnel appraisal records.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Agresti, A. (2010). *Categorical data analysis*. Wiley.
- Appelbaum, E., Bailey, T., Berg, P., & Kalleberg, A. L. (2000). *Manufacturing advantage: Why high-performance work systems pay off*. Cornell University Press.
- Asano, M., Yamagiwa, A., & Goto, M. (2026a). An analytical framework for personnel evaluation data aimed at stepwise performance improvement. *Journal of the Japan Society for Management Information*. Manuscript under review. (In Japanese)
- Asano, M., Yamagiwa, A., & Goto, M. (2026b). Behavioral characteristics and performance relationships based on longitudinal analysis of personnel evaluation data: An empirical examination of factor structures and transition patterns. *Journal of Japan Industrial Management Association*. Manuscript under review. (In Japanese)
- Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance. *Human Performance*, 10, 99–109. [\[CrossRef\]](#)
- Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt, & W. C. Borman (Eds.), *Personnel selection in organizations*. Jossey-Bass.
- Cawley, B. D., Keeping, L. M., & Levy, P. E. (1998). Participation in the performance appraisal process: A meta-analytic review. *Journal of Applied Psychology*, 83, 615–633. [\[CrossRef\]](#)
- Dang, L. H., Vuong, B. N., & Huan, D. D. (2025). How human resource management practices promote job performance through innovative work behavior: Does innovative climate matter? *Cogent Psychology*, 12, 2498245. [\[CrossRef\]](#)
- DeNisi, A. S., & Pritchard, R. D. (2006). Performance appraisal, performance management, and improving individual performance: A motivational framework. *Management and Organization Review*, 2, 253–277. [\[CrossRef\]](#)
- Hackman, J. R., & Oldham, G. R. (1980). *Work redesign*. Addison-Wesley.
- Hattori, Y., Yatera, A., & Arai, K. (2022). The relationship between personal reputation and performance appraisal: An exploration of the influencing factors. *Japanese Journal of Administrative Science*, 34(1–2), 1–19. (In Japanese) [\[CrossRef\]](#)

- Hayashi, Y., & Takahashi, K. (2002). Relationship between performance-based personnel systems and perceived fairness of appraisal. *Keiei Kodo Kagaku*, 16, 15–25. (In Japanese) [[CrossRef](#)]
- Jiang, K., Lepak, D. P., Hu, J., & Baer, J. C. (2012). How does human resource management influence organizational out-comes? A meta-analytic investigation of mediating mechanisms. *Academy of Management Journal*, 55, 1264–1294. [[CrossRef](#)]
- Kahya, E. (2007). The effects of job performance on employee satisfaction. *International Journal of Industrial Ergonomics*, 37, 515–520. [[CrossRef](#)]
- Kuvaas, B. (2008). An exploration of how employee–organization relationships affect HRM outcomes. *Human Resource Management Journal*, 18, 3–18.
- LePine, J. A., Erez, A., & Johnson, D. E. (2002). The nature and dimensionality of organizational citizenship behavior: A critical review and meta-analysis. *Journal of Applied Psychology*, 87, 52–65. [[CrossRef](#)] [[PubMed](#)]
- London, M., & Smither, J. W. (1995). Can multi-source feedback change perceptions of goal accomplishment, self-evaluations, and performance-related outcomes? Theory-based applications and directions for research. *Personnel Psychology*, 48, 803–839. [[CrossRef](#)]
- Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR analytics. *The International Journal of Human Resource Management*, 28, 3–26. [[CrossRef](#)]
- Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal*. Sage Publications.
- Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington Books.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903. [[CrossRef](#)] [[PubMed](#)]
- Podsakoff, P. M., MacKenzie, S. B., Paine, J. B., & Bachrach, D. G. (2000). Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management*, 26, 513–563. [[CrossRef](#)]
- Rotundo, M., & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology*, 87, 66–80. [[CrossRef](#)] [[PubMed](#)]
- Rousseau, D. M. (2006). Is there such a thing as evidence-based management? *Academy of Management Review*, 31, 256–269. [[CrossRef](#)]
- Scullen, S. E., Mount, M. K., & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology*, 85, 956–970. [[CrossRef](#)] [[PubMed](#)]
- Sturman, M. C. (2005). Using within-person variability to understand performance. *Personnel Psychology*, 58, 877–914.
- Takahashi, K. (2001). Fairness of personnel appraisal in employment organizations. *Soshiki Kagaku*, 34, 26–38. (In Japanese)
- Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81, 557–574. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.