Estimating the Impact of Managers' Online Training on the Quality of Line Managers-Employee Relationships: Examining Heterogeneity, Mediation, and Moderation Effects

Student Name: Tahir Korma
Supervisor Name: Juraj Medzihorsky

Abstract - Ineffective management plays a major role in reducing productivity within organisations. Evidence shows that about one-third of employees have encountered workplace conflict, with an average of 1.8 hours per week spent addressing these issues, leading to an annual loss of 370 million days in the UK (CIPD, 2020). This study explores the effect of online managerial training on employee relationships with line managers, examining differences in heterogeneity, mediation, and moderation over a one-year period. The primary aim is to assess whether training managers improves these relationships or if changes occur independently of the intervention. Using a randomised controlled trial (RCT) by Bowyer and Urwin (2024), managers in selected workplace units received an online training 'treatment', while others continued with standard practices. Contrary to expectations, the study found a slight decrease in the quality of relationships between employees and line managers following the overall training over time. Similarly, a slight negative impact was observed in the female group, while slight positive changes were noted in males regarding heterogeneity, as well as in mediation and moderation, although none of these changes were significant. The findings suggest that the managerial training provided did not improve line manager-employee relationships across all analyses. This research underscores the need to reassess training programmes and their effectiveness in fostering productive workplace relationships.

Keywords - causal inference, employee-manager relationship, online managerial training, training effectiveness, randomised controlled trial, workplace relationship quality

Introduction

In today's dynamic business environment, effective management is crucial for organisational success. As companies evolve, the importance of developing managers' interpersonal and technical skills has become increasingly evident, highlighting the need for targeted training programmes. Managers are now expected to excel not only in technical expertise but also in fostering positive relationships with their employees. Effective management is defined by the ability to build and sustain strong relationships with employees, which are essential for creating a productive and harmonious workplace (Personalysis, 2023). These relationships significantly impact employee morale, job satisfaction, and overall organisational performance, viewing employees as valuable assets rather than costs (Cook, 2024). No organisation is perfectly equipped with professional human resources. Like other departments, Human Resources require continuous skill development and training to ensure smooth operations. Line managers, who serve as the crucial link between upper management and subordinates, play a vital role in translating strategic directives into operational practices. To enhance organisational productivity, they must be equipped with essential skills and qualities. The quality of interactions between line managers and their subordinates is vital for achieving organisational goals.

This research examined the overall impact of online training interventions on line managers' relationships with employees (Dlamini et al., 2022), with a focus on gender-based variations in measuring the heterogeneity effect in employees (An & Meier, 2021). Additionally, the study explored the mediation effect—how training influences relationships through the ability of line managers to handle conflicts quickly—and the moderation effect, investigating how the relationship between training and manager-employee relationship quality varied based on factors such as employee role (e.g., front vs. back office) (Baron & Kenny, 1986). These questions are particularly relevant given the growing reliance on digital training methods.

The insights gained from this research could assist both practitioners and policymakers in the field of organisational development by providing evidence on how to optimise training programmes. Practitioners might use these findings to refine the design and delivery of training programmes, making them better suited to the needs of managers and their teams. This might include adjusting content to address specific challenges faced by line managers, incorporating interactive elements to increase engagement, or offering ongoing support after training to reinforce learning. For policymakers, the research can offer guidance on structuring and implementing training initiatives that support broader organisational objectives. The study’s examination of gender differences, mediation, and moderation underscores the need to consider various factors when developing training policies. Understanding how these factors affect training effectiveness can help policymakers design more inclusive and adaptable training strategies that accommodate different managerial contexts and employee groups.

In the long term, the findings from this research could contribute to the development of best practices for online training programmes, assisting organisations in investing in interventions that enhance managerial skills and improve workplace relationships. These practices may then be shared across industries, supporting the overall enhancement of management practices.

The theoretical framework for the research is based on literature concerning the effectiveness of managerial training and the dynamics of employee-manager relationships. Research indicates that online training can be as effective as traditional face-to-face methods in delivering content and enhancing skills (Sitzmann et al., 2006; Bernard et al., 2004). However, the success of online training often depends on factors such as the programme's design, participant engagement, and the support available during the training process (Clark & Mayer, 2016). Additionally, literature on employee-manager relationships underscores the importance of effective management in creating positive workplace environments. Strong employee-manager relationships are linked to higher levels of job satisfaction, employee engagement, and organisational commitment (Snyder & Lopez, 2007; Liden et al., 2000). Effective communication, trust, and mutual respect are identified as key components of these relationships (Graen & Uhl-Bien, 1995).

This research adopted a mixed-methods approach, combining quantitative and qualitative data from the Skilled Managers – Productive Workplaces (SMPW) study, which employed a randomised controlled trial (RCT) (Bowyer & Urwin, 2024). Data was gathered through pre- and post-intervention assessments, focusing on managers' competence and confidence, employee experiences and attitudes, the effectiveness of conflict management, and organisational productivity. Furthermore, qualitative insights were collected through interviews and focus groups to investigate productivity improvements and identify potential barriers.

The data was analysed using a causal inference framework with statistical techniques to estimate the differences between those who received training and those who did not. Additionally, thematic analysis was conducted on the qualitative data to reveal underlying themes and insights.

In summary, this research provides a comprehensive evaluation of online training interventions for managers, focusing on gender differences, mediation, and moderation in their effects on line manager-employee relationships. The study concludes by summarising its findings and suggesting avenues for future research on the impact of online managerial training.

Literature Review

The impact of managerial training programs on line managers' relationships with their employees is a significant area of inquiry, especially in the context of growing reliance on digital training methods. This review synthesises research findings on the effectiveness of such training, with a focus on heterogeneity, mediation, and moderation effects. The objective is to understand how online training interventions affect these relationships, with particular attention to gender-based variations and the role of contextual factors.

Effectiveness of Managerial Training Programs

Managerial training programs have been extensively studied, revealing their influence on organisational outcomes. Busso et al. (2023) conducted a meta-analysis of 44 studies and 68 programmes, highlighting key factors contributing to their effectiveness. Their research indicated that training in human resources, soft skills, marketing, and finance/accounting generally improved firm performance, including management practices, productivity, and survival rates. The study employed rigorous methods like randomised controlled trials (Jadad & Enkin, 2007) and regression techniques to standardise effect sizes, showing that programmes delivered by local organisations and involving diverse participants were especially effective. However, Busso et al. acknowledged limitations in their study, such as focusing only on quantitative measures of organisational growth and performance, potential publication bias, and variability in outcome measures. Their findings suggest that while training programmes can significantly enhance firm performance, further research is needed to explore long-term impacts, measure qualitative factors regarding manager-employee training effects, understand the mechanisms driving effectiveness, and examine the role of contextual factors in moderating training success.

Powell and Yalcin's (2010) meta-analysis of studies from 1952 to 2002 offers further insights into the effectiveness of managerial training programmes. Their review of 62 studies and 85 interventions revealed that overall effectiveness remained moderate, with no substantial improvement over time. While participants might acquire knowledge, translating this into behavioural change or measurable results, such as improved relationships with line managers, is more complex. Programs focusing on learning outcomes had more significant effects than those targeting behavioural or results-based outcomes. This indicates that despite persistent efforts in managerial training, substantial gains in changing managers' behaviour or improving employees' perceptions of managers remain elusive. Powell and Yalcin's research highlights the importance of considering study design and outcome measures when assessing training effectiveness.

Variations in Training Effectiveness

The effectiveness of managerial training varies across different managerial levels and contexts. Tamzid (2022) investigated how training impacts differ among top, middle, and first-line managers. Analysis of data from 181 managers in Dhaka, Bangladesh, using One-Way ANOVA (Heiberger & Neuwirth, 2009), revealed that first-line managers benefited most from training, followed by middle and top-level managers. This underscores the need for tailored training programmes that address the specific roles and responsibilities of each managerial level. However, the study's geographical and industry limitations, along with its reliance on selfreported data, point to the need for further research on adapting training to diverse contexts and managerial roles.

Van Leeuwen et al. (2023) conducted a pilot study on an online training programme aimed at enhancing career-oriented people management behaviours among line managers in academic hospitals. Using the Intervention Mapping protocol (Bartholomew et al., 2016), the study found that the training effectively improved participants' self-reflection and motivation. However, the study faced limitations such as a small, non-randomised sample and potential self-selection bias. These issues suggest that future research should investigate the long-term effects of such training, its impact on team outcomes, and its applicability across different settings and occupational groups.

An and Meier (2021) examined gender-based variations in training effectiveness through a field experiment in Denmark, focusing on the impact of transformational leadership training on male and female leaders. They found that both genders benefited from the training, but women showed more significant improvements in transformational leadership behaviours, whereas men experienced greater gains from transactional training (Bass, 1985). The study, which used surveys from 368 leaders and 4,352 employees analysed via a difference-in-difference approach (Lechner, 2011), provides valuable insights. However, its focus on Denmark and specific government functions may limit the generalisability of its findings.

Mediation and Moderation Effects

Dlamini et al. (2022) investigated how the relationship between managers and employees affects performance in a financial organisation in Durban, South Africa. They found that positive employee-manager relationships significantly enhance motivation and performance, whereas negative relationships lead to decreased productivity. Using a quantitative approach with a census method, the study highlighted the importance of positive managerial relationships for employee performance. However, the study's small sample size, focus on a single organisation, and exclusive measurement of quantitative performance factors limit its broader applicability. Additionally, it did not analyse qualitative factors related to employees' sense of belonging within the organisation.

Adhvaryu et al. (2023) examined the effects of soft skills training for production line supervisors using a randomised controlled trial (Jadad & Enkin, 2007). Their findings showed that less-recommended supervisors experienced a 12% increase in productivity, while highly recommended supervisors had a 15% lower likelihood of quitting compared to the control group. This study highlights the strategic allocation of training resources, suggesting that less-recommended supervisors benefit more in terms of productivity, whereas highly recommended ones have improved retention rates. The results emphasise the need for targeted training strategies and further research into training effectiveness in various contexts.

Nielsen et al. (2010) used Kirkpatrick's (Kirkpatrick & Kirkpatrick, 2006) training evaluation model which uses four levels: Reaction (participant satisfaction), Learning (knowledge gained), Behavior (application on the job), and Results (impact on organizational goals) to assess manager training in elderly care centres over 18 months. Their longitudinal study found improved job satisfaction and involvement among the intervention group, though effects on team interdependency and autonomy were mixed. The mixed-methods approach offered valuable insights into training's impact on team dynamics and employee outcomes but also noted challenges such as budget constraints and organisational restructuring.

Theoretical and Practical Implications

Arthur et al. (2003) conducted a meta-analysis to evaluate how different training design and evaluation features influence effectiveness. Their findings revealed medium to large effect sizes across various criteria: 0.60 for reaction, 0.63 for learning, 0.62 for behaviour, and 0.62 for results, underscoring the importance of well-designed and evaluated training programmes. The research highlights that, despite criticisms of certain training methods like lectures, they can be effective in specific contexts.

Lusher (1990) explored training process effectiveness using Bion's and Tuckman's frameworks to understand group dynamics in managerial training (Bion, 1961; Tuckman, 1965). By examining concepts like Dependence, Fight-Flight, and Pairing, Lusher provided insights into how managerial teams interact with training and how group cohesion evolves. This understanding is crucial for addressing gender-related differences in training effectiveness, as varying gender-related behaviours and group interactions can significantly impact training outcomes.

Poulet (1986) proposed a multifaceted evaluation framework for training programs, focusing on selection and content efficiency, as well as cost-effectiveness. This approach allows for comparative analysis of training programs but faces challenges in evaluating intangible skills and developmental outcomes. Poulet's framework underscores the need for a balanced approach that considers both quantitative and qualitative factors when assessing training effectiveness.

Sahni (2020) used the Kirkpatrick framework to evaluate managerial training, focusing on reaction and learning levels. The study gathered cross-sectional data from 136 middle-level managers through surveys assessing satisfaction and learning outcomes before and after training. The results showed a 57% increase in participant satisfaction and knowledge outcomes, with practical orientation identified as a key factor in training success. The research highlights the importance of incorporating practical elements into training programme design to enhance effectiveness.

The literature reviewed underscores the complex impact of managerial training programs on line managers' relationships with employees. While training programs generally enhance organisational performance and managerial learning, their effectiveness varies based on managerial levels, gender, and training focus. Mediation and moderation effects, such as conflict-handling skills and contextual factors, are also crucial. This research aims to address gaps by exploring how online training interventions affect manager-employee relationships, focusing on gender-based variations and contextual factors. It will use a longitudinal study and RCT data (Bowyer & Urwin, 2024) across various sectors, and estimate the effects from employees' qualitative responses rather than just organisational outcomes. These methodological enhancements will provide a more comprehensive understanding of the long-term impacts and cross-sector applicability of training programs. Future research should build on these findings to refine the design and implementation of effective online training programs in diverse organisational settings.

Methodology

The research employed thorough methods to assess the average effects of the online managerial training intervention, including how the impact of training varies between males and females (heterogeneity effect). It explored how a variable, q7_s, influences the effect of the treatment group ($D = 1$) on the outcome variable Y (the mediation effect M of q7_s) and how the relationship between $D = 1$ and Y changes across different levels of a stratum (the moderation effect W of stratum) on the response variable q9_s, which asks, "To what extent do you agree with the statement, 'I have a good relationship with my line manager?'" In the context of this analysis, "managerial training" or "intervention" refers specifically to online training. The report consistently uses either variable names or mathematical notations, with details about these variables and other covariates provided in Table 1: SMPW Dataset Metadata in Appendix - A.

Data Collection

The study uses recent data created by Bowyer and Urwin (2024) from the Skilled Managers Productive Workplaces (SMPW) which was primarily collected from two online surveys of staff conducted at the participating organisations in the SMPW study. Additional organisational-level data was recorded from interviews with organisation representatives.

Twenty-four organisations in the UK, including those in tourism, manufacturing, healthcare, retail, charity, and education, etc., from both the public and private sectors, showed interest in the research to carry out a randomised controlled trial (Urwin, 2022). The trial randomly assigned all managers in different workplace units to receive online training as the "treatment," while other units continued with "business as usual" as the control group. The SMPW used a mixed-methods approach to gather data by combining both quantitative and qualitative information collected before and after a 12-month treatment period. This was done to evaluate the impact of online training on ten different qualitative outcomes, with special emphasis on q7_s, which tracks how quickly line managers address team conflict. However, this research examines another important aspect, q9_s, which reflects the quality of the relationship between employees and their line managers.

The study involved a workplace trial of training interventions aimed at improving the conflict resolution skills of line managers in various private and public sector organisations. Two levels of skills development were tested. The first was a one-day workshop focused on building conflict resolution skills for first-line managers. The second level included an additional two-day workshop designed to prepare senior leaders to support and coach their managers through challenging personnel issues.

The impact of each intervention was tracked over 12 months by evaluating managers' competence and confidence, employees' experiences and attitudes, the efficiency of conflict resolution, and measures of organisational productivity (Wood & Wall, 2007). The data was uploaded to the UK Data Service with open access in two CSV files. The first file contains employee survey data with 3,460 observations and 19 columns. This experimental data includes variables 'wave' and 'group' to segment the data, where 'wave' indicates the time of data collection (e.g., wave 1 is pre-intervention, wave 2 post-intervention) and 'group' shows whether the online training was given (treatment) or not (control). Other variables include demographic details and Likert-scale responses (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree, don't know). The second file contains organisational data with 6 columns and 24 observations, providing information about the organisation's structure and type. The variable 'orgid' serves as the unique identifier for organisations and appears in both the employee survey and organisational data, allowing the files to be linked. Refer to Table 1: SMPW Dataset Metadata (Appendix - A) for complete details on all the variables.

Data Pre-processing

As a preliminary step, the two datasets (employee survey and organisational data) were examined for outliers before addressing missing values. This was done to ensure that the imputation of missing values, if necessary, isn't influenced by any high variance caused by outliers. No outliers were found in the dataset except in the los (length of stay) column of employee survey data, which exhibited a high degree of right-skewness and contained 143 outliers. These outliers were handled using a logarithmic transformation to approximate a normal distribution as illustrated in Figure 1 below.

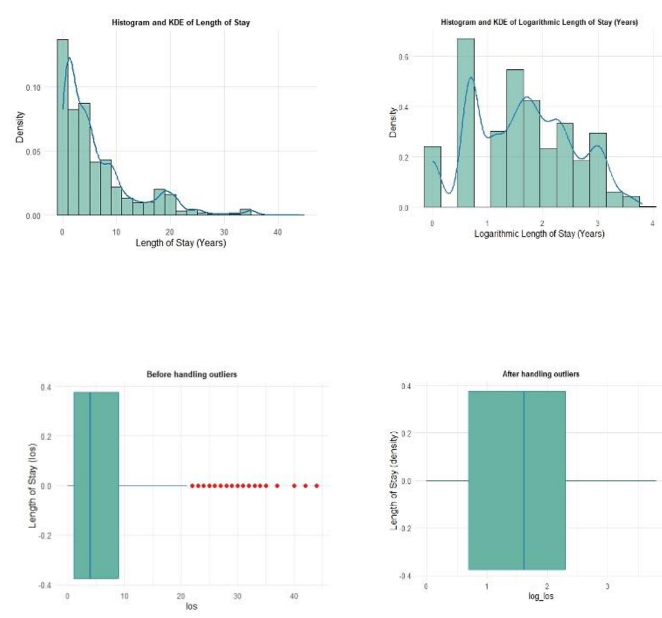

Figure 1: Distribution and Boxplot of LOS Column: Before and After Logarithmic Transformation for Outliers Handling

Following this, missing values were identified in the employee survey data: 210 for gender, 225 for ethnicity, and 840 for los. In the organisational data, 1 missing value was found in org_nor (number of people in the organization). Given that the data is experimental and well-randomised with observations equally distributed across different groups and waves, the analysis uses two approaches to handle the missing data.

The first approach involves excluding missing values for gender and ethnicity, which fall into the MNAR (Missing Not at Random) category. This type of missingness is associated with the unobserved data itself and could result in biased estimates if not properly addressed (Little & Rubin, 2002). For missing values related to numerical columns such as los and the org_nor in the organization, the Predictive Mean Matching (PMM) method for multiple imputation was used. This technique imputes missing values by selecting observed values with similar predicted means, creating multiple imputed datasets. By matching cases with missing values to those with comparable predicted values, PMM helps maintain the data's natural variability and distribution. Refer to Appendix - B, Section 1.1 for the Multiple Imputation Procedure using PMM for handling missing values in los and the org_nor.

The second approach utilized the polynomial regression (Polyreg) method for multiple imputation to handle missing data in categorical variables like gender and ethnicity. This method models relationships between categories and estimate missing values by considering the polynomial connections among categorical predictors. By reflecting the underlying structure of the data, Polyreg helps to account for the uncertainty introduced by missing values. This dual approach allows for comparison of results from both methods, offering a more detailed understanding of the impact of managerial training interventions while reducing bias associated with missing data. Refer to Appendix - B, Section 1.2 for the Multiple Imputation Procedure using Polyreg for handling missing values in gender and ethnicity.

After handling outliers and missing values, the two datasets were merged into separate files: one with dropped missing values and the other with imputed values. Subsequently, all the categorical variables were converted into numeric format by assigning a unique number to each level within the categories. This was done to facilitate modelling and evaluate the various effects of the training interventions.

Econometric Analytical Framework

Difference-in-Differences, also known as DiD (Lechner, 2011), is an effective econometric method used to assess the causal impact of an intervention by comparing changes in outcome variables over time between a treatment group and a control group. This approach is especially useful for analysing the effects of an intervention within an experimental design, as it helps separate the intervention's impact from other time-related factors and confounding variables (Angrist & Pischke, 2009). The DiD method is important for evaluating not only the overall treatment effect but also the effects of heterogeneity, mediation, and moderation.

In this study, the DiD method is used within the framework of a Randomised Controlled Trial (RCT) to assess the impact of an online training intervention on the quality of the relationship between employees and their line managers, as measured by the outcome variable q9_s or Y. The RCT design involved randomly assigning UK-based organisations to either receive the training (treatment or $D = 1$) or continue with business as usual (control or $D = 0$), with data collected at two different time points: before the intervention (wave 1) and after the intervention (wave 2). This setup allows for both within-group comparisons, which track changes over time within each group, and between-group comparisons, which assess differences between groups at specific times (wave 2) to determine the intervention's effects. The value of the RCT here is its ability to provide a clear estimate of the training's effect by removing selection bias and ensuring that any differences in q9_s are due to the intervention itself rather than pre-existing differences or external factors. The randomisation process makes sure that both the treatment and control groups are comparable at the start, allowing any differences observed after the intervention to be confidently linked to the training programme (Fisher, 1935; Rubin, 1974).

DiD uses this RCT setup by comparing changes in Y between the $D = 1$ and $D = 0$ groups at these time points. This method helps to isolate the effect of the training by accounting for time-related factors and trends, giving a clear view of its impact on line manager-employee relationships (Angrist & Pischke, 2009; Imbens & Wooldridge, 2009). While the parallel trends assumption is key in observational DiD studies (which assumes that without treatment, the treatment and control groups would have followed similar trends), the randomisation in this RCT lessens the need for this assumption by ensuring minimal systematic differences between the groups at the start. This supports the reliability of the DiD analysis, even with just one pre-treatment period (Ashenfelter, 1978; Blundell & Costa Dias, 2000).

DiD also enables the exploration of heterogeneous treatment effects through subgroup analyses. In this study, it is used to assess whether the impact of the training differs by demographic factors such as gender. By including interaction terms in the DiD framework (regression), the study can examine how the training's effects vary between male and female employees, offering insights into whether the training is more or less effective for a particular gender (Lechner, 2011).

Additionally, DiD supports mediation and moderation analyses by providing a useful framework for understanding how the training's effects on the outcome variable q9_s might be mediated by factors such as line managers' ability to resolve team conflict quickly, or moderated by factors like stratum, indicating whether the employee works in the front or back office. This approach allows for a more detailed evaluation of how these intermediary and interacting variables influence the training's impact, improving the overall understanding of its effectiveness (Baron & Kenny, 1986; Hayes, 2013).

Finally, DiD is effective for controlling time-varying factors that could affect q9_s. By comparing the differences in changes between the treatment and control groups before and after the intervention, DiD accounts for external factors or trends that might otherwise skew the results. This approach helps ensure that the observed effects are due to the intervention itself, rather than external influences or pre-existing trends (Card & Krueger, 1994; Meyer, 1995)

Directed Acyclic Graphs (DAGs) and Estimands of Interest

Directed Acyclic Graphs (DAGs) are an effective tool for visualising and analysing causal relationships between variables. They consist of nodes (variables) and directed edges (arrows) that indicate causal effects, ensuring that there are no cyclic relationships. By mapping these paths, DAGs help identify both direct and indirect effects, as well as interactions where the relationship between treatment and outcome variables is influenced by another variable. This approach provides a framework for detecting potential confounders and understanding causal dynamics in complex systems (Pearl, 2009).

In this study, DAGs are employed to map and analyse the causal relationships involved in evaluating an online managerial training intervention. They will illustrate how the training's impact varies with heterogeneity, mediation, and moderation, guiding the model setup to ensure that interaction effects are accurately incorporated and the causal structure is correctly represented, which is essential for appropriate statistical analysis.

The study aimed to estimate the following estimands:

(i) Average Treatment Effect (ATE): This reflects the average difference in q9_s between the treatment and control groups that can be attributed to the intervention accounting for the time factor. The below DAG (Figure 2: Average Treatment Effect (ATE) on q9_s) shows several causal paths between treatment, time effects, and the outcome. The direct path from $D = 1 \rightarrow Y$ shows the immediate effect of the treatment on outcome q9_s. Indirectly, $D = 1$ influences the I_1 (group × wave), which then affects Y, reflecting how the treatment's impact varies over time. Likewise, T (wave1 & wave2) affects the I_1 (group × wave), which in turn impacts Y, demonstrating how different periods influence treatment outcome. The direct path from $C \rightarrow Y$ shows the effect of covariates on outcome. As the dataset comes from a randomised controlled trial (RCT), the random assignment of treatment reduces bias from confounding variables and ensures that treatment effects are not skewed by pre-existing group differences. While covariates may influence the outcome, they do not confound the results. The RCT design ensures that observed treatment effects are attributable to the intervention rather than external factors.

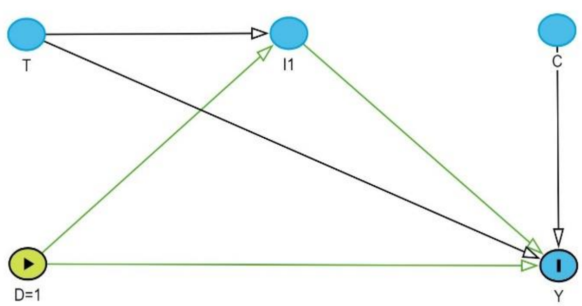

Figure 2: DAG illustrating the Average Treatment Effect (ATE) on q9_s, created using DAGitty (Textor et al., 2017)

ATE Expression

$$\begin{aligned}\mathrm{ATE} &= \mathbb{E}[\,Y \mid D = 1, T\,] - \mathbb{E}[\,Y \mid D = 0, T\,] \\ &= \beta_1 + \beta_3 \mathrm{T}\end{aligned}$$

where;

D: Group indicator (Treatment = 1, Control = 0)
T: Wave indicator (Before = 1, After = 2)
$\mathbb{E}[Y \mid D = 1, T]$: The mean outcome when receiving the treatment ($D = 1$) at wave T
$\mathbb{E}[\,Y \mid D = 0, T\,]$: The mean outcome when not receiving the treatment($D = 0$) at wave T
β_1: Coefficient for the treatment group indicator D. Represents the average treatment effect when $T = 1$ (pre-treatment period)
$\beta_3 T$: Coefficient for the interaction term $D \times T$. Represents the additional effect of the treatment in the post-treatment period (when $T = 2$)

The following conditional independencies are assumed while measuring ATE:

- $D \perp T \mid C$: Treatment assignment is independent of the wave, given covariates. This reflects the random assignment of treatment.
- $D \perp \epsilon \mid T, C$: Treatment assignment is independent of the error term, given the wave and covariates. Ensures no unobserved confounding affects the treatment effect.
- $T \perp \epsilon \mid D, C$: The wave is independent of the error term, given the treatment and covariates. Ensures no unobserved confounding affects the wave effect.
- $D \times T \perp \epsilon \mid D, T, C$: The interaction term is independent of the error term, given treatment, wave, and covariates. Ensures the interaction effect is correctly specified.

(ii) Heterogeneous Treatment Effect: The variation in the ATE for Y across different subgroups, specifically male and female employees, examines whether the training's impact differs between these gender groups. This analysis follows the same model used for estimating the *Average Treatment Effect on q9_s* above, with the only difference being that the dataset is divided into two parts, one for males and one for females, to estimate the treatment effects more precisely. Figure 2: Showing the DAG of ATE on q9_s and the ATE expression, can be used as a reference.

(iii) Mediation Effect: The influence of the mediating variable q7_s on the relationship between the treatment group and the outcome variable q9_s. This estimand investigates how changes in q7_s may affect the training's impact on q9_s. In the given DAG (Figure 3: Average Treatment Effect on q9_s, mediated by q7_s) multiple causal paths emphasise the relationships between the treatment, the mediator, and the outcome. The direct path from $D = 1$ to Y reflects the immediate effect of the treatment on q9_s. The indirect path from $D = 1 \rightarrow$

$M \to Y$ demonstrates how q7_s mediates the relationship between the treatment and q9_s, enabling us to explore how changes in q7_s influence the impact of the training on q9_s. As outlined earlier, the DAG also includes paths where the $D = 1 \to I_1 \to Y$, showing how treatment effects vary over time, and where wave T affects the outcome demonstrating how different periods influence treatment outcome, highlighting temporal influences. Additionally, the direct path from $C \to Y$ captures the covariate's effect on Y. The DAG does not show any biased paths due to confounding variables, as these are addressed by the experimental design. The inclusion of the mediating variable q7_s allows for a deeper understanding of the causal mechanisms, specifically how the treatment's impact on the outcome is partially transmitted through changes in q7_s. By examining these pathways, the analysis provides insights into both the direct and indirect effects of the treatment, offering a comprehensive view of how the intervention influences the outcome variable q9_s.

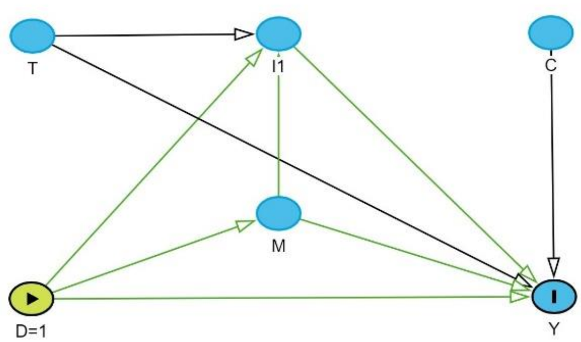

Figure 3: DAG illustrating the Average Treatment Effect on q9_s, mediated by q7_s, created using DAGitty (Textor et al., 2017)

ATE Expression

$$\begin{aligned}\mathrm{ATE} = \big[\mathbb{E}[\,Y \mid D = 1, T, M\,] - \mathbb{E}[\,Y \mid D = 0, T, M\,]\big] \\ + \big(\beta_1 + \beta_3 T \\ + \beta_2(\mathbb{E}[\,M \mid D = 1, T\,] \\ - \mathbb{E}[\,M \mid D = 0, T\,])\big)\end{aligned}$$

where;

D: Group indicator (Treatment = 1, Control = 0)
T: Wave indicator (Before = 1, After = 2)

$[\mathbb{E}[Y \mid D = 1, T, M] - \mathbb{E}[Y \mid D = 0, T, M]]$: This term represents the difference in the outcome Y between the treatment group ($D = 1$) and the control group ($D = 0$). This captures the total effect, including both the direct and indirect effects through M
$\mathbb{E}[\,Y \mid D = 0, T\,]$: The mean outcome when not receiving the treatment($D = 0$) at wave T
β_1: Coefficient for the treatment group indicator $D = 1$. Represents the average direct treatment effect when $T = 1$ (pre-treatment period) and without considering the mediator M
β_3: Coefficient for the interaction term $D \times T$. Represents the additional effect of the treatment in the post-treatment period (when $T = 2$)
β_2: Coefficient for the indirect effect of the treatment through the mediator M
$\mathbb{E}[M \mid D = 1, T] - \mathbb{E}[M \mid D = 0, T]$: The change in the mediator M due to the treatment, capturing how these changes in M affect the outcome Y

In addition to the conditional independencies outlined in the *Average Treatment Effect (ATE)*, one additional assumption is made when measuring the ATE involving the mediator q7_s:

- $M \perp \epsilon \mid D, T, C$: The mediator is independent of the error term, given treatment, time, and covariates. Ensures that the mediator M is not confounded by unobserved factors affecting the outcome. This is crucial for correctly identifying the indirect effect of the treatment through the mediator.

(iv) Moderation Effect: The variation in the treatment effect across different levels of a moderator variable W. This evaluates how the relationship between the treatment and the outcome variable q9_s is affected by different levels of a stratum (e.g., front vs. back office). In the given DAG (Figure 4: Average Treatment Effect on q9_s, moderated by stratum) for estimating the ATE for the outcome variable, multiple causal paths highlight the relationships between treatment, the moderator, and the outcome. The direct path from $D = 1 \to$ Y reflects the immediate effect of the treatment on q9_s. The indirect path from $D = 1 \rightarrow I_2 \rightarrow Y$ and $T \rightarrow I_2 \rightarrow Y$ shows how the interaction term, which combines treatment, time, and moderator (stratum), influences the relationship between treatment and outcome, allowing us to investigate how the combined effects of these variables affect the Y. The path from moderator $W \rightarrow I_2 \rightarrow Y$ demonstrates how different levels of the moderator influence the outcome by interacting with treatment and time. Finally, $C \rightarrow Y$ captures the impact of covariates on outcome. The inclusion of the moderator W allows for a deeper understanding of how the intervention's impact varies across different strata. This offers a comprehensive view of both the direct and moderated effects of the treatment.

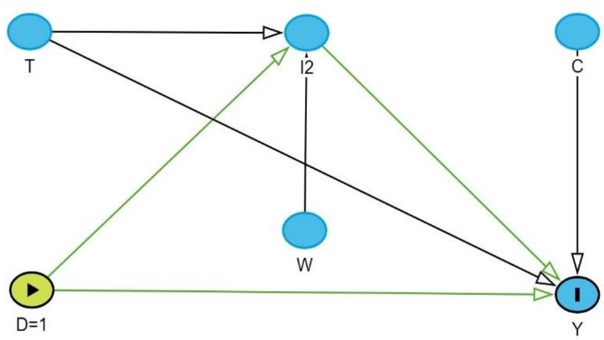

Figure 4: Illustrating the Average Treatment Effect (ATE) on q9_s, moderated by stratum

ATE Expression

$$\text{ATE} = [\beta_1 + \beta_2(\mathbb{E}[\, I_2 \mid D = 1, T, W\,] - \mathbb{E}[\, I_2 \mid D = 0, T, W\,]) + \beta_3 T + \beta_4 W\,]$$

where;

D: Group indicator (Treatment = 1, Control = 0)
T: Wave indicator (Before = 1, After = 2)
β_1: Coefficient for the treatment group indicator $D = 1$. Represents the average direct treatment effect when $T = 1$ (pre-treatment period) and W is at its baseline level
β_2: Coefficient for the interaction term I_2, representing how the combined effect of treatment, time, and moderator impacts the outcome
β_3: Coefficient for the interaction term $D \times T$. Represents the additional effect of the treatment in the post-treatment period (when $T = 2$)
$\beta_4 W$: Coefficient for the moderator. Represents the direct effect of the moderator on the outcome
$\mathbb{E}[I_2 \mid D = 1, T, W]$: Indicates the average value of interaction term I_2 when treatment is received, given specific time and moderator levels
$\mathbb{E}[I_2 \mid D = 0, T, W]$: Indicates the average value of interaction term I_2 when treatment is not received, given specific time and moderator levels

In addition to the conditional independencies outlined in the *Average Treatment Effect (ATE)*, one additional assumption is made when measuring the ATE involving the moderator stratum:

- $I_2 \perp \epsilon \mid D, T, W, C$: The interaction term I_2 is independent of the error term given treatment, time, moderator, and covariates.

Modelling with OLS Regression

Ordinary Least Squares (OLS) regression is a fundamental statistical method used to estimate the relationships between a dependent variable and one or more independent variables. Its strength lies in its simplicity and ease of interpretation, which makes it a popular choice in empirical research. The method works by minimising the sum of the squared differences between the observed values and those predicted by the model. These differences, known as residuals, are squared and summed to find the best-fitting line (or hyperplane in the case of multiple regressors) that represents the relationship between the independent variables and the dependent variable (Sharma, 2020). Mathematically, the OLS estimator for the coefficient β in a simple linear regression model $Y = \alpha + \beta X + \epsilon$ is given by:

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

This equation calculates the slope of the regression line, representing the change in the dependent variable Y for a one-unit change in the independent variable X.

OLS has several notable strengths:

1. Interpretability: The coefficients in OLS regression indicate the average change in the dependent variable for a one-unit change in the predictor variable, with all other variables held constant. This clear interpretation makes the results easy to communicate.

2. Flexibility: OLS can handle both continuous and categorical independent variables. It also permits the inclusion of interaction terms, allowing for the exploration of how the relationship between two variables varies depending on the level of a third variable.

3. Computational Efficiency: OLS is computationally efficient as it utilises matrix operations to swiftly estimate parameters, making it suitable for large datasets. Its efficiency is due to its closed-form solution, which eliminates the need for iterative procedures and simplifies calculations.

OLS regression was selected for this research because of its flexibility and effectiveness in meeting the study's objectives. It provides a clear and interpretable estimate of the overall effect of the training intervention on the outcome variable. It also allows for the inclusion of interaction terms to analyse how the intervention's impact differs across various subgroups, such as by gender. However, to more precisely estimate the heterogeneity effect for male and female employees, the dataset is divided into separate male and female groups, making interaction terms unnecessary in this case. For mediation analysis, the `mediation` package in R is used, which fits multiple OLS models to decompose the total effect into direct and indirect pathways. This approach allows for a detailed understanding of how the training intervention influences the outcome through mediators, enhancing insights into the mechanisms behind the intervention's effects. Furthermore, it is well-suited for moderation analysis, enabling the exploration of how contextual factors, such as work stratum, may influence the strength or direction of the treatment effect. Overall, OLS's ability to handle complex interactions and produce clear results makes it well-suited for thoroughly assessing the impact of the training intervention.

Modelling Expressions

The following OLS modeling expressions were implemented in this research.

1. Estimating Average Treatment Effect (ATE) and Heterogenity Effect: $Y = \beta_0 + \beta_1 D + \beta_2 T + \beta_3 (D \times T) + \gamma TC + \epsilon$

2. Estimating Mediation Effect:

(i) Mediator Model: $q7_{s(M)} = \alpha_0 + \alpha_1 D + \alpha_2 T + \alpha_3 (D \times T) + \epsilon_m$

(ii) Outcome Model: $Y = \beta_0 + \beta_1 D + \beta_2 T + \beta_3 (D \times T) + \beta_8 q7_s + +\beta_9 (q7_s \times T) + \epsilon_y$

3. Estimating Moderation Effect: $Y = \beta_0 + \beta_1 D + \beta_2 T + \beta_3 (D \times T) + \beta_4 W + \beta_5 (D \times W) + \beta_6 (T \times W) + \beta_7 (D \times T \times W) + \gamma TC + \epsilon$

where;

Y: Outcome variable q9_s
D: Treatment group indicator $(1 = \text{ treatment}, 0 = \text{control})$
T: Wave indicator $(1 = \text{pre} - \text{treatment}, 2 = \text{post} - \text{treatment})$
$D \times T$: Interaction between treatment and wave, showing how the treatment effect varies over time
α_0/β_0: Intercept, representing the baseline level of the outcome when all predictors are zero
α_1/β_1: Coefficient for treatment indicator D, indicating the direct effect of the treatment on the outcome
α_2/β_2: Coefficient for wave indicator T, showing the direct effect of time on the outcome
α_3/β_3: Coefficient for interaction D x T, representing how the treatment effect changes over time
β_4: Coefficient for W, indicating its direct effect on the outcome
β_5: Interaction term showing how the treatment effect on the outcome varies with W
β_6: Interaction term showing how the effect of time on the outcome varies with W

β_7: Three-way interaction term indicating how the treatment effect over time is moderated by W
β_8: Coefficient for the mediator q7_s, showing its effect on the outcome while controlling for other variables
β_9: Interaction term indicating how the mediator's effect on the outcome varies with time
γTC: Effect of covariates. γ is a vector of coefficients for the covariates C. The term γTC represents the combined effect of all covariates on the outcome.
$\epsilon/\epsilon_m/\epsilon_y$: Error term. It captures the unobserved factors affecting the outcome and is assumed to be normally distributed with mean zero

Model Interpretation

After fitting the model, the coefficients for the treatment group, wave, covariates, and interaction terms, along with their p-values, were analysed to interpret the effects of training. The following metrics were used for this interpretation:

Coefficients: These represent the change in the dependent variable for a one-unit increase in the predictor variable, while holding other variables constant. A positive coefficient suggests that the training has a beneficial impact, increasing the dependent variable, whereas a negative coefficient indicates a less advantageous effect. Interaction term coefficients reveal how the training impact evolves over time and with other contextual factors.

Standard Errors (SEs): Standard errors measure the precision of the coefficient estimates. Smaller standard errors indicate more reliable estimates, enhancing confidence in the coefficient values.

t-values: The t-value is the ratio of the coefficient to its standard error ($\frac{\beta}{SE}$). High t-values suggest that the training effect is notably different from zero, indicating a substantial impact on the outcome. Similarly, high t-values for interaction terms indicate meaningful changes in the training effect over time or with other factors.

p-values (Pr(>|t|)): The p-value indicates the probability of observing a t-value as extreme as, or more extreme than, the one computed under the null hypothesis that the coefficient is zero. A small p-value signifies that the training effect is statistically relevant and unlikely to be due to random chance.

Additional Metrics: R-squared measures the proportion of variance in the dependent variable explained by the model, while Adjusted R-squared accounts for the number of predictors. The F-statistic tests the overall model relevance, and confidence intervals provide a range within which the true coefficient values are likely to fall.

Results

After performing OLS regression modeling on the analysis of various research questions, the following results were observed. The findings are outlined as follows:

1. Average Treatment Effect (ATE)

The results from the model show the average difference in q9_s between the treatment and control groups, considering the time factor, and reveal how various predictors affect the outcome. The intercept, with an estimate of 0.3353, had a high standard error and a high p-value, meaning it was not an important factor in explaining q9_s. Essentially, the baseline level of q9_s, when all predictors are set to zero, did not have a big impact on the outcome. In short, the intercept did not provide much insight into the differences in q9_s between the treatment and control groups.

The treatment effect, shown by $group1$ (estimated effect of being in the treatment group (as opposed to the control group) on the outcome variable with an estimate of -0.0281, and the time factor, shown by $wave2$ (estimated effect of being in the wave2 (post-treatment period) compared to the wave1 (pre-treatment period) on the outcome variable with an estimate of -0.0079, had only a slight negative impact and were not influential. This suggests that neither the treatment nor the passage of time had a strong effect on q9_s. Similarly, the interaction term $group1{:}wave2$, with an estimate of -0.0225, indicated no major change in the treatment effect over time. This means the treatment's impact on q9_s did not vary much with different time points. Overall, these findings suggest that neither the treatment nor the time factor played a major role in changing q9_s.

Among the predictors, q8_s, which measures the effect of talking to the line manager on employee performance, had the most noticeable positive effect on q9_s, with an estimate of 0.2724. This indicates that improvements in this area greatly enhance the outcome. Other predictors such as q1_s, q2_s, q3_s, q4_s, q7_s, q10_s, and los also had positive effects on q9_s, highlighting their importance in influencing the outcome. These findings suggest that, while the treatment and time factors did not have a big impact, certain key predictors, especially q8_s, were crucial in driving changes in q9_s. This implies that focusing on these influential variables could be more effective in improving q9_s. For more details on other estimates, refer to Table 2: Estimation Table of Average Treatment Effect (ATE) in Appendix - A.

The model explained about 19% of the variation in the outcome variable, with a residual standard error that was fairly high. While some variability remains unexplained, the model performed well overall. This is supported by a very high F-statistic, which shows that the model is effective in explaining the outcome variable. For detailed model summary statistics, see Section 2.1: Model Summary Statistics for ATE.

Overall, the finding suggest that the treatment and time factors had a slight negative and unreliable impact on q9_s, with the intercept providing little insight into differences between groups. However, predictors like q8_s, which reflects interactions with the line manager, had a substantial positive effect on q9_s. This suggests that focusing on key predictors, particularly q8_s, could be more beneficial for improving q9_s.

2.1 Heterogeneous Treatment Effect with Imputed Missing Data for Gender and Ethnicity

(i) In Males: The results for males indicate that the intercept was estimated at 0.2218, but the high standard error and p-value suggest that this estimate does not meaningfully affect q9_s, as it is imprecise and unreliable. Similar to the results of Average Treatment Effect (ATE), the treatment effect for $group1$ had an estimate of -0.0993, pointing to a potential decrease in q9_s for the treatment group compared to the control group. However, the high standard error and p-value indicate that this effect is not consistent or particularly reliable. Similarly, the time effect $wave2$ had an estimate of -0.1390, suggesting a possible decrease in q9_s over time, but again, the high standard error and p-value make this effect unreliable. The interaction term $group1{:}\,wave2$, estimated at 0.1810, indicates a slight positive change in q9_s over time for males. Despite this, the high standard error and p-value suggest that this effect is not consistent, meaning the treatment's impact on q9_s does not change much over time.

As with the ATE results, q8_s had the most noticeable positive effect on q9_s, making it a reliable predictor. Other predictors (mentioned in ATE results), except q4_s (which measures the recognition of good performance within the team), also showed positive effects on q9_s, highlighting their importance in influencing the outcome. For a detailed overview of all estimates, refer to Table 3: Estimation Table of Heterogeneous Treatment Effect (Males) in Appendix - A.

The model explained about 15% of the variation in q9_s, with some variability remaining unexplained. Overall, the model performed well, supported by a strong F-statistic, which shows that the model is effective in explaining the outcome variable. For detailed model summary statistics, see Section 2.2: Model Summary Statistics for Males with Imputed Missing Data.

(ii) In Females: The analysis for females showed that the intercept was estimated at 0.1004, but the high standard error and p-value suggest that this estimate does not meaningfully impact q9_s and does not explain the outcome effectively.

For the treatment effect, $group1$ had an estimate of 0.0259, indicating little to no effect of the treatment on q9_s for females, which is confirmed by the high p-value. Similarly, the time effect $wave2$ had an estimate of 0.0913, suggesting that the effect of time on q9_s is minimal and not strong. The interaction term $group1{:}\,wave2$ had an estimate of

-0.1853, suggesting a possible change in the treatment effect over time for females, but this effect is not consistent.

Among the predictors, q8_s again had the strongest positive effect. Other predictors mentioned in ATE, along with stratum2, also showed positive effects, with stratum2 suggesting slight improvement for females in back-office roles compared to front-office roles. However, this effect is not strong. These results highlight the importance of these variables in influencing the outcome. For a detailed overview of all estimates, refer to Table 4: Estimation Table of Heterogeneous Treatment Effect (Females) in Appendix - A.

The model explained about 22% of the variation in q9_s, with some variability remaining unexplained. Overall, the model performed well, as indicated by a strong F-statistic, showing that the model is effective in explaining the outcome variable. For detailed model summary statistics, see Section 2.3: Model Summary Statistics for Females with Imputed Missing Data.

Overall, the heterogeneous effect results suggest that the treatment had a modest impact on q9_s for both males and females, with other workplace-related factors playing a more prominent role. The consistency in the importance of q8_s across both groups highlights the potential value of focusing on management interactions to improve outcomes, rather than relying solely on the treatment itself. Additionally, the differences in how certain roles influenced the outcome for females could indicate the need for more tailored approaches depending on job type or other contextual factors.

2.2 Heterogeneous Treatment Effect with Dropped Missing Data for Gender and Ethnicity

(i) In Males: The results from the model, excluding missing data, showed that the intercept was estimated at 0.3311, which was almost identical to the model with imputed missing data. Both models had high standard errors and p-values, indicating that the intercept did not have a noticeable impact on the outcome in male group. The treatment effect for $group1$ was estimated at -0.0874, which did not differ much from the imputed model. Both estimates were slightly negative, with high standard errors and p-values, suggesting that the treatment had little effect. The time effect $wave2$, estimated at -0.1499, also showed no considerable difference between the two methods. Both estimates indicated a small negative effect over time, but the high p-values suggested there was no clear impact. The interaction term $group1{:}\,wave2$ was estimated at 0.1054, with high standard errors and p-values, showing no distinct interaction effect, consistent with the results from the imputed data model.

Among the predictors, both models identified similar predictors with no substantial differences. In terms of model fit, the model explained approximately 15.7% of the variance, showing a very slight increase. This small improvement suggested that removing missing data might offer a marginal enhancement in model performance. For detailed model summary statistics, see Section 2.4: Model Summary Statistics for Males with Dropped Missing Data

Overall, handling missing data—either through imputation or by dropping cases—led to similar conclusions about effects and predictors in the male group. Both methods yielded comparable insights, with minimal differences, and showed similar results for treatment, time, and interaction effects.

(ii) In Females: The results for females from the model with missing data removed showed that the intercept was estimated slightly higher at 0.2328 compared to the imputed missing data model. While both models indicated that the intercept had little effect on the outcome, the dropped data model provided a somewhat higher estimate. For the treatment effect represented by $group1$, the dropped data model provided an estimate of 0.0643 with a high standard error and p-value. The treatment effect was not notable in either model, although the dropped data model had a slightly higher estimate. Regarding the time factor $wave2$, the estimate of 0.1279, with a high standard error and p-value, suggested that the time effect was not meaningful, although the dropped data model indicated a slightly larger positive effect. For the interaction term $group1{:}\,wave2$, the estimate of

-0.2321 showed a slightly more negative effect compared to the imputed model, but neither model indicated a considerable interaction. Additionally, there was no major difference in predictors compared to the imputed model. This suggests that imputing or dropping missing data did not reveal major changes

In terms of model performance, the imputed data model explained about 22% of the variance in q9_s, while the dropped data model explained approximately 23.5% of the variance. This slight improvement suggested that removing missing data provided a small boost in explaining the outcome variable. For detailed model summary statistics, see Section 2.5: Model Summary Statistics for Females with Dropped Missing Data

Similar to the findings for the male group, both models indicated that treatment, time, and interaction effects were not notable for females. This suggests that excluding missing data marginally enhances model performance.

3. Mediation Effect

The mediation analysis looked into the role of q7_s, which measures how well the line manager handles and resolves conflict within the team, in mediating the relationship between the treatment group and the outcome variable q9_s, while also considering the time effect. The results show the following:

The Average Causal Mediation Effect (ACME) is estimated at 0.0201, with a 95% confidence interval ranging from -0.0040 to 0.0400. This suggests a small positive impact of q7_s on the relationship between the treatment group and q9_s, but the high p-value indicates that this effect is uncertain and not particularly reliable. The Average Direct Effect (ADE) is estimated at -0.0654, with a confidence interval of -0.1911 to 0.0600. This points to a minor negative direct effect of the treatment on the outcome, though the high p-value suggests this effect is not dependable.

The Total Effect of the treatment on the outcome is estimated at -0.0453, with a confidence interval from -0.1717 to 0.0800. This shows a small negative effect, but again, the high p-value indicates that this effect is not consistent. The proportion of the total effect mediated through q7_s is estimated at -0.4443, with a wide confidence interval of -5.413 to 2.63. The high p-value suggests substantial uncertainty, meaning there is no clear evidence of mediation.

Overall, the analysis suggests that while there is a slight indication that the ability of the line manager to handle and resolve conflict might mediate the relationship between the treatment and q9_s, the effects are small and uncertain. The treatment's impact remains unclear, with both the direct and total effects being unreliable. For a detailed overview, refer to Table 5: Estimation Table of Mediation Effect of 'q7_s' in Appendix - A.

4. Moderation Effect

The analysis of how different factors affect the outcome showed that the starting point (intercept) was estimated at 0.3460. However, this estimate had a high standard error and a high p-value, suggesting that the starting point does not greatly influence the outcome, meaning it isn't a major factor in explaining it. For the treatment effect, the $group1$ factor had an estimate of 0.0133, with a high standard error and a high p-value, indicating that the treatment has little effect. Similarly, the time factor $wave2$ had an estimate of -0.0306, with a high standard error and a high p-value, suggesting that the impact of time on the outcome is only slightly reduced but not notable. The $stratum2$ factor had an estimate of 0.0971, with a high standard error and a high p-value, showing it does not have a strong effect on the outcome either. The interactions between these factors also showed minor effects. For instance, the interaction between $group1{:}wave2$ (estimate: -0.0914), between $group1{:}stratum2$ (estimate: -0.1810), and between $wave2{:}\ stratum2$ (estimate: 0.1390) did not have a noticeable impact, as indicated by their high standard errors and p-values. The three-way interaction term $group1{:}wave2{:}stratum2$ (representing the effect of being in the treatment group during the post-treatment period while working in the back office) had an estimate of 0.3275 showed a slight increase in q9_s, but this was also not a major factor. For more details, please refer to Table 6: Estimation Table of Moderation Effect of Stratum.

Overall, the model accounted for about 19% of the variation in the outcome q9_s and performed well, as shown by a strong F-statistic. This means the model explains a fair amount of the outcome, though some variability remains unexplained. For detailed model summary statistics, see Section 2.6: Model Summary Statistics for Moderator-Stratum.

In summary, the analysis across all four areas showed that the treatment and time factors had minimal impact on the outcome variable q9_s, with no notable changes over time or between the treatment and control groups. The key predictors, including q8_s, consistently displayed a strong positive influence on outcome variable. Both methods for handling missing data, whether through imputation or by dropping cases, produced similar results, confirming that the treatment effects were generally unreliable.

Evaluation

The analysis of the impact of online interventions on line managers' relationships with their employees offered detailed insights into the average effectiveness of the treatment, its differing effects on male and female groups, and the roles of mediation, moderation, and other key predictors emerging as important determinants.

The overall findings showed that neither online training nor the time factor alone made a notable improvement in the quality of the employee-line manager relationship. Both had minimal positive and negative effects across all measures, indicating that the impact was not statistically meaningful. This aligns with Powell and Yalcin's (2010) findings that while managerial training can improve knowledge, translating these gains into enhanced relationships or behavioural change is complex and often limited. The interaction between treatment and time was also unremarkable, suggesting that the treatment's effect did not change over time. This result supports the literature indicating that sustained improvements in manager-employee relationships through training are difficult to achieve.

The model explained approximately 19% of the variation, indicating that while it captured some important factors, a significant portion of variability remained unexplained. The high residual standard error indicated unexplained variance, suggesting there may be other important factors influencing the outcome apart from the measured variables in the data, such as organisational culture, leadership style, or external economic conditions. This finding is consistent with research by Busso et al. (2023) and Van Leeuwen et al. (2023), which emphasised the critical role of contextual factors in determining training effectiveness. Future research should consider these additional factors to better understand line manager-employee relationships.

Moreover, the OLS regression used in this research assumes linear relationships and may not effectively handle non-linearities or interactions, as shown by the minor and statistically insignificant effects in interaction terms. It could also be sensitive to outliers and multicollinearity and assumes homoscedastic errors and normal distribution. Exploring alternative models, such as Generalised Linear Models (GLMs), which are well-suited for handling the ordinal nature of the outcome, could provide insights into how predictors influence different levels of the q9_s. Additionally, Ridge and Lasso regressions could be valuable for addressing multicollinearity and high-dimensional data, potentially improving the assessment of treatment effects and predictor relationships.

Although the online training intervention was intended to enhance line managers' conflict resolution skills with the expectation that this would positively influence the line manager-employee relationship, the results showed that it did not lead to a notable improvement in overall relationships between line managers and their employees compared to the control group over the 12-month period (Bowyer & Urwin, 2024). This result suggests that the anticipated improvement in satisfaction did not materialise, indicating a need to reconsider the treatment's design or implementation. Adjustments could involve evaluating the training's dosage, duration, content, delivery method, or adding support mechanisms to better translate training into practice.

Additionally, other factors related to managerial interaction, employee experience, and organisational behaviour (Sparrow et al., 2016) significantly impacted the main research question. Effective and frequent communication between employees and line managers emerged as crucial for fostering better relationships. This suggests that the online training intervention alone was insufficient and other contextual factors played a role in influencing the mentioned outcome. Therefore, interventions should focus on enhancing communication channels, feedback mechanisms, and regular check-ins.

Emerging predictors, such as employee tenure and role type, indicated potential influences on the outcome, though their effects were not statistically meaningful and warrant further investigation. The results suggest that organisations might benefit more from targeted initiatives, such as in-person workshops, blended learning, or one-on-one coaching, rather than relying solely on online training. This aligns with the need for tailored training approaches noted by Tamzid (2022). A more holistic approach, incorporating additional strategies and addressing underlying factors, may be more effective. Given the lack of a meaningful impact over 12 months, a longer-term follow-up might be useful to identify any delayed effects.

The analysis of heterogeneous treatment effects across two datasets—one with imputed missing values and the other with missing values removed—showed almost identical results, with only a minor improvement when missing values were excluded. This implies a few things. Firstly, the data likely came from a randomised controlled trial (RCT), where randomisation may have balanced covariates across treatment groups, potentially reducing the impact of missing data (Bowyer & Urwin, 2024). Moreover, using multiple imputation methods, such as Predictive Mean Matching (PMM) and Polyregression, may have preserved the data's underlying distribution without introducing bias. Variables with missing data, such as gender and ethnicity, seemed to have a minimal effect on the outcome. This is consistent with the Average Treatment Effect (ATE), which showed that gender and ethnicity did not notably affect the outcome variable. This aligns with some studies reviewed, which suggest that gender and ethnicity might not always be key factors in training effectiveness. For example, Busso et al. (2023) found that while diverse participant involvement is important, focusing on broader organisational outcomes rather than specific demographic factors suggests that gender and ethnicity may have less impact in certain contexts.

However, this contrasts with other findings in the literature. An and Meier (2021) found that gender notably influenced training outcomes, with women showing greater improvements in transformational leadership compared to men. This indicates that gender can play a more significant role in training effectiveness under certain conditions, which contradicts the minimal impact observed here. Additionally, the literature review indicates that training programmes, especially those targeting specific managerial behaviours or soft skills, may vary in effectiveness depending on contextual factors and demographic variables.

Statistically, variables like gender and ethnicity, which had minimal impact, likely had low effect sizes or small coefficients in the regression model, reflecting their limited explanatory power. This supports the view that, in this analysis, these variables had little influence. Nonetheless, the heterogeneous treatment effects suggest that the impact of these variables might depend on the context, aligning with literature that shows variability in the influence of contextual factors across different studies. Future research should explore how different contextual factors, including gender and ethnicity, affect training outcomes across various settings and populations to better understand their role in training effectiveness.

The mediation analysis suggested that while there was a hint that q7_s (managers' ability to handle conflict quickly) might mediate the relationship between the training and line manager-employee relationships, the effects were minimal and not statistically reliable. This indicates that q7_s does not have a strong mediating role in this context. Given the weak mediation effect, the online training programme (Bowyer & Urwin, 2024), aimed at improving conflict handling within teams, did not notably improve line managers' conflict management in a way that affected the quality of the relationship between employees and managers. Supporting this, Dlamini et al. (2022) found that while positive manager-employee relationships are important for performance, some training programmes may not sufficiently cover conflict handling, which aligns with the limited mediation effect observed here. Furthermore, Nielsen et al. (2010) reported varied results regarding the impact of training on team dynamics and job satisfaction, suggesting that not all training interventions enhance all aspects of managerial skills. This reinforces the idea that the current training might not have effectively improved conflict handling. However, this contrasts with some literature suggesting targeted training can lead to notable improvements. Adhvaryu et al. (2023) showed that well-targeted soft skills training resulted in clear gains in productivity and retention, indicating that effective training is possible if well-designed. Additionally, Busso et al. (2023) noted that various training programmes have successfully enhanced management practices and firm performance, suggesting that the current programme might not have been optimally designed or implemented. Future research should explore other potential mediators and examine how different training elements might better influence manager-employee relationships.

The moderation analysis showed a slight positive effect, indicating that the combined influence of the treatment, time, and factors related to employees' roles do not play a major part in determining the success of the intervention or its impact over time. Considering other organisational factors as moderators—such as industry sector, proportion of female managers, number of managers, or size of the organisation—could offer more insights into how different conditions affect outcomes. For instance, Van Leeuwen et al. (2023) observed that contextual factors, like the specific setting of an online training programme, might impact its effectiveness, suggesting that additional variables could be important in understanding training success.

The evaluation revealed that online training interventions had negligible impact on improving the relationships between line managers and their employees over a year. The findings suggest that other factors, such as organizational culture and communication practices, play a crucial role and may need to be addressed alongside training. Future efforts should consider more comprehensive and context-sensitive approaches to achieve better outcomes.

Conclusions

The research examined the impact of online managerial training on the quality of relationships between line managers and their employees, with a focus on heterogeneity, mediation, and moderation effects. The study used a randomised controlled trial (RCT) design and employed the Difference-in-Differences (DiD) method to analyse the data. The findings revealed that the online training intervention did not lead to a statistically notable improvement in the overall quality of manager-employee relationships. The heterogeneity analysis showed that the training's impact did not vary notably between male and female employees. The mediation analysis suggested a potential, though not statistically notable, mediating role of conflict resolution skills in the relationship between training and relationship quality. The moderation analysis indicated that the relationship between training and relationship quality was not notably moderated by employee role type.

The study's findings have important implications for both practitioners and policymakers. The results suggest that online training alone may not be sufficient to improve manager-employee relationships. Organisations may need to consider a more comprehensive approach that includes other interventions, such as coaching, mentoring, or on-the-job training. The study also highlights the importance of considering contextual factors, such as employee role type, when designing and implementing training programmes. Future research could explore the impact of online training on other aspects of manager-employee relationships, such as trust, communication, and support. Additionally, future studies could examine the long-term effects of online training on relationship quality and organisational outcomes. The insights gained from this research can contribute to the development of more effective training programmes that enhance managerial skills and foster positive workplace relationships, ultimately leading to improved organisational performance.

References

Adachi, H., Sekiya, Y., Imamura, K., Watanabe, K., & Kawakami, N. (2019). The effects of training managers on management competencies to improve their management practices and work engagement of their subordinates: A single group pre- and post-test study. *Journal of Occupational Health*, *62*(1). https://doi.org/10.1002/1348-9585.12085

Adhvaryu, A., Murathanoglu, E., & Nyshadham, A. (2023). On the Allocation and Impacts of Managerial Training. *Social Science Research Network*. https://doi.org/10.2139/ssrn.4475933

An, S. H., & Meier, K. J. (2021). Gender and the Effectiveness of Leadership Training: Results From a Field Experiment. *Review of Public Personnel Administration*, *41*(4), 747–770. https://doi.org/10.1177/0734371x20932989

Angrist, J. D., & Pischke, J. S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. Retrieved July 27, 2024, from https://jonnyphillips.github.io/FLS6415/Class_3/Angrist%20&%20Pischke.pdf

Arthur, W., Bennett, W., 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

Ashenfelter, O. (1978). Estimating the Effect of Training Programs on Earnings. *Review of Economics and Statistics*, *60*(1), 47–57. https://doi.org/10.2307/1924332

Atiase, V., & Botchie, D. (2018). Does Managerial Training have any impact on the performance of MSE Managers? Empirical evidence from Ghana. *International Journal of Entrepreneurial Behavior & Research*. https://doi.org/10.1108/IJEBR-05-2022-0405

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Bartholomew, L. K., Markham, C. M., Ruiter, R. a. C., & Parcel, G. S. (2016). *Planning Health Promotion Programs: An Intervention Mapping Approach*.

Bass, B. M. (1985). *Leadership and Performance Beyond Expectations*. New York: Free Press.

Bion, W. R. (1961). *Experiences in Groups and Other Papers*. London: Tavistock Publication.

Blundell, R., & Dias, M. C. (2000). Evaluation Methods for Non-Experimental Data. *Fiscal Studies*, *21*(4), 427–468. https://doi.org/10.1111/j.1475-5890.2000.tb00031.x

Bowyer, A., & Urwin, P. (2024, March 26). *Skilled Managers - Productive Workplaces, Survey Data, 2022-2023*. UK Data Service. Retrieved July 25, 2024, from https://doi.org/10.5255/UKDA-SN-857109

Bryson, A., Forth, J., & Stokes, L. (2017). Does employees' subjective well-being affect workplace performance? *Human Relations*, *70*(8), 1017–1037. https://doi.org/10.1177/0018726717693073

Busso, M., Park, K., & Irazoque, N. (2023). *The Effectiveness of Management Training Programs: A Meta-Analytic Review*. https://doi.org/10.18235/0004815

Card, D., & Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania. *American Economic Review*, *84(4)*, 772–993. https://davidcard.berkeley.edu/papers/njmin-aer.pdf

CIPD. (2020). Managing conflict in the modern workplace. In *CIPD*. Retrieved August 10, 2024, from https://www.cipd.org/uk/knowledge/reports/managing-workplace-conflict-report/

Collins, D. (2022). *The effectiveness of managerial leadership development programs: a meta-analysis of studies from 1982-2001*. https://doi.org/10.31390/gradschool_dissertations.2461

Cook, M. (2024, May 26). *The Art of Connection: Elevating Employee Relations*. https://happycompanies.com/blog/the-art-of-connection-elevating-employee-relations

Dlamini, N. P., Suknunan, S., & Bhana, A. (2022). Influence of employee-manager relationship on employee performance and productivity. *Problems and Perspectives in Management*, *20*(3), 28–42. https://doi.org/10.21511/ppm.20(3).2022.03

Fisher, R. A. (1935). *The Design of Experiments*. Oliver & Boyd. Retrieved July 27, 2024, from https://home.iitk.ac.in/~shalab/anova/DOE-RAF.pdf

Hayes, A. F. (2013). Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. *Journal of Educational Measurement*, *51*(3), 335–337.

Heiberger, R. M., & Neuwirth, E. (2009). One-Way ANOVA. In *Springer eBooks* (pp. 165–191). https://doi.org/10.1007/978-1-4419-0052-4_7

Imbens, G. W., & Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, *47*(1), 5–86. https://doi.org/10.3386/w14251

Jadad, A. R., & Enkin, M. (2007). *Randomized Controlled Trials: Questions, Answers and Musings*.

Kirkpatrick, D., & Kirkpatrick, J. (2006). *Evaluating Training Programs: The Four Levels* (3rd ed.). Berrett-Koehler Publishers.

Lechner, M. (2011). The Estimation of Causal Effects by Difference-in-Difference Methods. *Foundations and Trends® in Econometrics*, *4*(3), 165–224. https://doi.org/10.1561/0800000014

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. https://doi.org/10.1002/9781119013563

Lusher, B. (1990). Improving Working Relationships: Group Effectiveness Training. *Journal of European Industrial Training*, *14*(5). https://doi.org/10.1108/03090599010001562

Meyer, B. D. (1995). Natural and Quasi-Experiments in Economics. *Journal of Business & Economic Statistics*, *13*(2), 151–161. https://doi.org/10.1080/07350015.1995.10524589

Nielsen, K., Randall, R., & Christensen, K. B. (2010). Does training managers enhance the effects of implementing team-working? A longitudinal, mixed methods field study. *Human Relations*, *63*(11), 1719–1741. https://doi.org/10.1177/0018726710365004

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Personalysis. (2023, October 16). *Fostering Employee Connection: A Guide to Cultivating a Cohesive Workplace*. PERSONALYSIS. https://www.personalysis.com/how-to-make-employees-feel-connected/#:~:text=A%20cohesive%20workplace%20is%20characterized,to%20contribute%20their%20best%20efforts.

Poulet, R. (1986). Training Process Effectiveness. *Industrial and Commercial Training*, *18*(1), 10–13. https://doi.org/10.1108/eb004022

Powell, K. S., & Yalcin, S. (2010). Managerial training effectiveness: A meta-analysis 1952-2002. *Personnel Review*, *39*(2), 227–241. https://doi.org/10.1108/00483481011017435

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701. https://doi.org/10.1037/h0037350

Sahni, J. (2020). Managerial Training Effectiveness:An Assessment through Kirkpatrick Framework. *TEM Journal*, *9*(3), 1227–1233. https://doi.org/10.18421/tem93-51

Sharma, N. (2020, August 14). Ordinary Least Squared (OLS) Regression. *Medium*. Retrieved August 17, 2024, from https://medium.com/analytics-vidhya/ordinary-least-squared-ols-regression-90942a2fdad5#:~:text=It%20is%20also%20one%20of,more%20advanced%20concepts%20and%20techniques.

Sparrow, P., Brewster, C., & Chung, C. (2016). *Globalizing Human Resource Management*. https://doi.org/10.4324/9781315668611

Tamzid, M. (2022). The Effects of Different Managerial Levels on Training Effectiveness. *International Journal of Social Science and Human Research*, *05*(08). https://doi.org/10.47191/ijsshr/v5-i8-03

Textor, J., Van Der Zander, B., Gilthorpe, M. S., Liśkiewicz, M., & Ellison, G. T. (2017). Robust causal inference using directed acyclic graphs: the R package 'dagitty.'

International Journal of Epidemiology, *45*(6), 1887–1894. https://doi.org/10.1093/ije/dyw341

Tuckman, B. (1965). Developmental sequence in small groups. *Psychological Bulletin*, *63*, 384–399. https://doi.org/10.1037/h0022100

Urwin, P. (2022, October). *Randomised Trial of the Skilled Managers Online Training Intervention, Protocol for Evaluation in Medium to Large Sized Organisations*. Social Science Registry. Retrieved August 14, 2024, from https://www.socialscienceregistry.org/versions/158700/docs/version/document

Van Leeuwen, E. H., Knies, E., Van Rensen, E. L. J., Taris, T. W., Van Den Heuvel, M., & Lammers, J. W. J. (2023). The Systematic Development of an Online Career-Oriented People Management Training for Line Managers of Professionals: A Pilot Field Intervention Study. *Public Personnel Management*, *53*(1), 61–84. https://doi.org/10.1177/00910260231176510

Wood, S. J., & Wall, T. D. (2007). Work enrichment and employee voice in human resource management-performance studies. *International Journal of Human Resource Management*, *18*(7), 1335–1372. https://doi.org/10.1080/09585190701394150

Appendix - A

SMPW Dataset Metadata

Variables	Description	Units/Possible Responses	Mathematical Notation
pulseid	Unique identifier for each survey respondent	n/a	Not used in analysis
orgid	Unique identifier for each organisation	n/a	Not used in analysis
manager	Indicates if the respondent manages other employees	Yes, No	C (Covariate)
q1_s	Likelihood of recommending the organisation to others (1-10 scale)	1 (very unlikely) to 10 (very likely)	C (Covariate)
q2_s	Agreement with the statement about respecting individual differences	Strongly Agree, Agree, Neither agree nor disagree, Disagree, Strongly Disagree, Don’t know	C (Covariate)
q3_s	Agreement with the statement about ease of getting feedback on performance	same as above	C (Covariate)
q4_s	Agreement with the statement about recognition of good performance in the team	same as above	C (Covariate)
q5_s	Agreement with the statement about effectively tackling poor performance in the team	same as above	C (Covariate)
q6_s	Agreement with the statement about experiencing work-related stress impacting health and job performance	same as above	C (Covariate)
q7_s	Agreement with the statement about the line manager helping resolve team conflict quickly	same as above	M (Mediator) & C (Covariate)
q8_s	Agreement with the statement about talking to the line manager helping improve performance	same as above	C (Covariate)
q9_s	Agreement with the statement about having a good relationship with the line manager	same as above	Y (Outcome Variable)
q10_s	Agreement with the statement about being happy in the role and not thinking of leaving	same as above	C (Covariate)
gender	Respondent's gender identity	Male, Female, Non-binary, Other	C (Covariate) & Interaction Variable
ethnic	Respondent's ethnic group	17 different ethnic groups reported	C (Covariate)
los	Years of employment at the organisation	Free text, numbers rounded to nearest year	C (Covariate)
group	Indicates if the respondent was in the treatment or control group	Treatment, Control	D (1 = Treatment, 0 = Control)
wave	Indicates if response was recorded before or after treatment	1,2	T(1=Before, 2 =After)
stratum	Indicates if the respondent was in a front office (1) or back office (2) unit	1, 2	W (Moderator), & C (Covariate)
sector	Text description of the organisation's sector	24 different sectors reported	C (Covariate)
female_mgr	Proportion of managers that are female	Number between 0 and 1, to two decimal places	C (Covariate)

progress	Mean progress managers have made completing the online course component of the intervention	Number between 0 and 1, to two decimal places	C (Covariate)
num_managers	Number of managers in the organisation	Number of individual managers, rounded to nearest 5	C (Covariate)
org_nor	Number of employees in the organisation	Number of individual employees, rounded to nearest 10	C (Covariate)
	Interaction Term: (group × wave)		I_1
	Interaction Term: (group × wave x stratum)		I_2

Table 1: SMPW Dataset Metadata

Estimation Table of Average Treatment Effect (ATE)

Coefficients	**Estimate**	**Std. Error**	**t value**	**p- value**
Intercept: Baseline value of q9_s when all predictors are set to their reference levels	0.335	0.255	1.317	0.188
group1: Difference in q9_s for the treatment group compared to the control group	-0.028	0.080	-0.351	0.725
wave2: Difference in q9_s between Wave 2 and Wave 1	-0.008	0.081	-0.098	0.922
q1_s: Effect on q9_s	0.102	0.013	7.866	0.000
q2_s: Effect on q9_s	0.095	0.017	5.709	0.000
q3_s: Effect on q9_s	0.084	0.018	4.711	0.000
q4_s: Effect on q9_s	0.046	0.018	2.631	0.009
q5_s: Effect on q9_s	-0.027	0.019	-1.401	0.161
q6_s: Effect on q9_s	0.027	0.018	1.518	0.129
q7_s: Effect on q9_s	0.051	0.019	2.665	0.008
q8_s: Effect on q9_s	0.272	0.018	15.378	0.000
q10_s: Effect on q9_s	0.119	0.017	6.899	0.000
manager1: Difference in q9_s for managers who manage employees versus those who do not	0.053	0.063	0.833	0.405
gender: Effect of gender on q9_s	-0.006	0.054	-0.103	0.918

Coefficients	Estimate	Std. Error	t value	p- value
ethnic: Effect of ethnicity on q9_s	0.012	0.012	0.949	0.343
stratum2: Difference in q9_s for being in Back Office compared to Front Office	0.134	0.079	1.710	0.087
sector: Effect of industry sector on q9_s	0.007	0.005	1.498	0.134
female_mgr: Effect of proportion of female managers on q9_s	0.023	0.120	0.188	0.851
progress: Effect of average progress made by managers in completing the online training on q9_s	-0.071	0.222	-0.322	0.748
num_managers: Effect of number of managers in the organisation on q9_s	-0.001	0.001	-0.674	0.500
org_nor: Effect of number of employees in the organisation on q9_s	0.000	0.000	-0.068	0.946
los: Effect of length of service (years) on q9_s	0.112	0.035	3.225	0.001
group1:wave2: Combined effect on q9_s for the Treatment Group in Wave 2 compared to the Control Group in Wave 1	-0.022	0.118	-0.190	0.849

Table 2: Estimation Table of Average Treatment Effect (ATE)

Estimation Table of Heterogeneous Treatment Effect (Males)

Coefficients	Estimate	Std.Error	t value	p-value
Intercept: Baseline value of q9_s when all predictors are set to their reference levels	0.222	0.523	0.424	0.672
group1: Difference in q9_s for the treatment group compared to the control group	-0.099	0.131	-0.760	0.448
wave2: Difference in q9_s between Wave 2 and Wave 1	-0.139	0.132	-1.057	0.291
q1_s: Effect on q9_s	0.106	0.024	4.518	0.000
q2_s: Effect on q9_s	0.111	0.028	3.982	0.000
q3_s: Effect on q9_s	0.065	0.029	2.198	0.028
q4_s: Effect on q9_s	0.032	0.030	1.081	0.280
q5_s: Effect on q9_s	-0.013	0.031	-0.404	0.686
q6_s: Effect on q9_s	0.039	0.028	1.388	0.166
q7_s: Effect on q9_s	0.072	0.031	2.308	0.021
q8_s: Effect on q9_s	0.240	0.029	8.137	0.000
q10_s: Effect on q9_s	0.088	0.028	3.139	0.002

Coefficients	Estimate	Std.Error	t value	p-value
manager1: Difference in q9_s for managers who manage employees versus those who do not	0.066	0.106	0.621	0.534
ethnic: Effect of ethnicity on q9_s	0.013	0.020	0.666	0.506
stratum2: Difference in q9_s for being in Back Office compared to Front Office	-0.081	0.127	-0.636	0.525
sector: Effect of industry sector on q9_s	0.001	0.010	0.138	0.890
female_mgr: Effect of proportion of female managers on q9_s	0.000	0.200	-0.002	0.999
progress: Effect of average progress made by managers in completing the online training on q9_s	0.265	0.494	0.536	0.592
num_managers: Effect of number of managers in the organisation on q9_s	0.001	0.002	0.556	0.578
org_nor: Effect of number of employees in the organisation on q9_s	0.000	0.000	-0.402	0.688
los: Effect of length of service (years) on q9_s	0.108	0.060	1.810	0.070
group1:wave2: Combined effect on q9_s for the Treatment Group in Wave 2 compared to the Control Group in Wave 1	0.181	0.196	0.921	0.357

Table 3: Estimation Table of Heterogeneous Treatment Effect (Males)

Estimation Table of Heterogeneous Treatment Effect (Females)

Coefficients	Estimate	Std.Error	t value	p-value
Intercept: Baseline value of q9_s when all predictors are set to their reference levels	0.100	0.307	0.327	0.744
group1: Difference in q9_s for the treatment group compared to the control group	0.026	0.105	0.247	0.805
wave2: Difference in q9_s between Wave 2 and Wave 1	0.091	0.105	0.874	0.382
q1_s: Effect on q9_s	0.097	0.016	6.059	0.000
q2_s: Effect on q9_s	0.092	0.021	4.339	0.000
q3_s: Effect on q9_s	0.096	0.023	4.167	0.000
q4_s: Effect on q9_s	0.057	0.023	2.511	0.012
q5_s: Effect on q9_s	-0.035	0.025	-1.400	0.162
q6_s: Effect on q9_s	0.015	0.023	0.637	0.524
q7_s: Effect on q9_s	0.043	0.025	1.745	0.081
q8_s: Effect on q9_s	0.285	0.023	12.567	0.000
q10_s: Effect on q9_s	0.139	0.022	6.250	0.000

Coefficients	Estimate	Std.Error	t value	p-value
manager1: Difference in q9_s for managers who manage employees versus those who do not	0.050	0.082	0.609	0.543
ethnic: Effect of ethnicity on q9_s	0.008	0.016	0.519	0.604
stratum2: Difference in q9_s for being in Back Office compared to Front Office	0.274	0.104	2.626	0.009
sector: Effect of industry sector on q9_s	0.010	0.005	1.857	0.064
female_mgr: Effect of proportion of female managers on q9_s	0.111	0.190	0.585	0.559
progress: Effect of average progress made by managers in completing the online training on q9_s	-0.009	0.266	-0.033	0.973
num_managers: Effect of number of managers in the organisation on q9_s	-0.001	0.001	-0.561	0.575
org_nor: Effect of number of employees in the organisation on q9_s	0.000	0.000	0.060	0.952
los: Effect of length of service (years) on q9_s	0.111	0.044	2.501	0.012
group1:wave2: Combined effect on q9_s for the Treatment Group in Wave 2 compared to the Control Group in Wave 1	-0.185	0.151	-1.224	0.221

Table 4: Estimation Table of Heterogeneous Treatment Effect (Females)

Estimation Table of Mediation Effect of ‘q7_s’

Term	Estimate	95% CI.Lower	95% CI.Upper	p-value
ACME	0.0201	-0.0040	0.04	0.074
ADE	-0.0654	-0.1911	0.06	0.352
Total Effect	-0.0453	-0.1717	0.08	0.522
Prop. Mediated	-0.4443	-5.413	2.63	0.564

Table 5: Estimation Table of Mediation Effect of ‘q7_s’

Estimation Table of Moderation Effect of Stratum

Coefficients	Estimate	Std. Error	t value	p-value
Intercept: Baseline value of q9_s when all predictors are set to their reference levels	0.346	0.256	1.352	0.176
group1: Difference in q9_s for the treatment group compared to the control group	0.013	0.090	0.149	0.882
wave2: Difference in q9_s between Wave 2 and Wave 1	-0.031	0.089	-0.346	0.729
stratum2: Difference in q9_s for being in Back Office compared to Front Office	0.097	0.144	0.673	0.501
q1_s: Effect on q9_s	0.103	0.013	7.898	0.000
q2_s: Effect on q9_s	0.095	0.017	5.709	0.000
q3_s: Effect on q9_s	0.083	0.018	4.649	0.000
q4_s: Effect on q9_s	0.046	0.018	2.614	0.009
q5_s: Effect on q9_s	-0.026	0.019	-1.352	0.176
q6_s: Effect on q9_s	0.027	0.018	1.514	0.130
q7_s: Effect on q9_s	0.051	0.019	2.656	0.008
q8_s: Effect on q9_s	0.272	0.018	15.369	0.000
q10_s: Effect on q9_s	0.119	0.017	6.888	0.000
manager1: Difference in q9_s for managers who manage employees versus those who do not	0.048	0.063	0.765	0.444
gender: Effect of gender on q9_s	-0.006	0.054	-0.111	0.911
ethnic: Effect of ethnicity on q9_s	0.012	0.012	0.980	0.327
sector: Effect of industry sector on q9_s	0.007	0.005	1.541	0.123
female_mgr: Effect of proportion of female managers on q9_s	0.022	0.120	0.180	0.857
progress: Effect of average progress made by managers in completing the online training on q9_s	-0.084	0.222	-0.380	0.704
num_managers: Effect of number of managers in the organisation on q9_s	-0.001	0.001	-0.684	0.494
org_nor: Effect of number of employees in the organisation on q9_s	0.000	0.000	-0.069	0.945
los: Effect of length of service (years) on q9_s	0.114	0.035	3.280	0.001

Coefficients	**Estimate**	**Std. Error**	**t value**	**p-value**
group1:wave2: Combined effect on q9_s for the Treatment Group in Wave 2 compared to the Control Group in Wave 1	-0.091	0.131	-0.700	0.484
group1:stratum2: Combined effect on q9_s for the Treatment Group in the Back Office compared to the Control Group in the Front Office	-0.181	0.199	-0.911	0.363
wave2:stratum2: Combined effect on q9_s for Wave 2 in the Back Office compared to Wave 1 in the Front Office	0.139	0.213	0.653	0.514
group1:wave2:stratum2: Combined effect on q9_s for the Treatment Group in Wave 2 within the Back Office, compared to the Control Group in Wave 1 within the Front Office	0.327	0.305	1.075	0.282

Table 6: Estimation Table of Moderation Effect of Stratum

Appendix - B

Section 1.1 - Multiple Imputation Procedure using PMM for handling missing values in los and org_nor

```
# Load the library for multiple imputation
library(mice)

# org_nor imputation
org_data$org_nor <- as.integer(org_data$org_nor)
imputed_data <- mice(org_data, m = 5, method = 'pmm', seed = 123)

# Get the completed data with imputations
completed_data <- complete(imputed_data)
completed_data$org_nor
org_data$org_nor <- completed_data$org_nor

# los imputation
emp_survey$los <- as.integer(emp_survey$los)
los_imputed_data <- mice(emp_survey, m = 5, method = 'pmm', seed = 123)

# Get the completed data with imputations
los_imputed_data1 <- complete(los_imputed_data)
los_imputed_data1$los
emp_survey$los <- los_imputed_data1$los
```

Section 1.2 - Multiple Imputation Procedure using Polyreg for handling missing values in gender and ethnicity

```
# gender imputation
emp_survey$gender <- as.factor(emp_survey$gender)
gender_imputed_data <- mice(emp_survey, m = 5, method = 'polyreg', seed = 123)
gender_imputed_data1 <- complete(gender_imputed_data)
gender_imputed_data1$gender
emp_survey$gender <- gender_imputed_data1$gender

# ethnic imputation
emp_survey$ethnic <- as.factor(emp_survey$ethnic)
ethnic_imputed_data <- mice(emp_survey, m = 5, method = 'polyreg', seed = 123)
ethnic_imputed_data1 <- complete(ethnic_imputed_data)
ethnic_imputed_data1$ethnic
emp_survey$ethnic <- ethnic_imputed_data1$ethnic
```

Section 2.1: Model Summary Statistics for ATE

```
# Fitting the model
model_df <- lm(q9_s ~ group * wave + q1_s + q2_s + q3_s + q4_s + q5_s + q6_s +
                q7_s + q8_s + q10_s + manager + gender + ethnic + stratum + sector +
                female_mgr + progress + num_managers + org_nor + los, data = df)

# Print the model summary
summary(model_df)
```

```
##
## Call:
## lm(formula = q9_s ~ group * wave + q1_s + q2_s + q3_s + q4_s +
##     q5_s + q6_s + q7_s + q8_s + q10_s + manager + gender + ethnic +
##     stratum + sector + female_mgr + progress + num_managers +
##     org_nor + los, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9690 -1.5536  0.2708  1.4315  3.8487
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.353e-01  2.547e-01   1.317  0.18807
## group1       -2.811e-02  7.999e-02  -0.351  0.72536
## wave2        -7.941e-03  8.065e-02  -0.098  0.92157
## q1_s          1.019e-01  1.296e-02   7.866 4.85e-15 ***
## q2_s          9.530e-02  1.669e-02   5.709 1.23e-08 ***
## q3_s          8.405e-02  1.784e-02   4.711 2.56e-06 ***
## q4_s          4.637e-02  1.762e-02   2.631  0.00855 **
## q5_s         -2.674e-02  1.909e-02  -1.401  0.16144
## q6_s          2.674e-02  1.761e-02   1.518  0.12913
## q7_s          5.089e-02  1.910e-02   2.665  0.00774 **
## q8_s          2.724e-01  1.771e-02  15.378  < 2e-16 ***
## q10_s         1.187e-01  1.720e-02   6.899 6.21e-12 ***
## manager1      5.264e-02  6.317e-02   0.833  0.40469
## gender       -5.563e-03  5.383e-02  -0.103  0.91769
## ethnic        1.153e-02  1.214e-02   0.949  0.34253
## stratum2      1.343e-01  7.851e-02   1.710  0.08733 .
## sector        6.783e-03  4.528e-03   1.498  0.13418
## female_mgr    2.257e-02  1.198e-01   0.188  0.85058
## progress     -7.138e-02  2.219e-01  -0.322  0.74767
## num_managers -6.465e-04  9.591e-04  -0.674  0.50030
## org_nor      -1.053e-05  1.547e-04  -0.068  0.94576
## los           1.121e-01  3.477e-02   3.225  0.00127 **
## group1:wave2 -2.246e-02  1.179e-01  -0.190  0.84893
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.719 on 3437 degrees of freedom
## Multiple R-squared:  0.1916, Adjusted R-squared:  0.1864
## F-statistic: 37.02 on 22 and 3437 DF,  p-value: < 2.2e-16
```

Section 2.2: Model Summary Statistics for Males with Imputed Missing Data

```
# Fitting the model for males
model_male <- lm(q9_s ~ group * wave + q1_s + q2_s + q3_s + q4_s + q5_s + q6_s +
                  q7_s + q8_s + q10_s + manager + ethnic + stratum + sector +
                  female_mgr + progress + num_managers + org_nor + los, data = male)

# Print the summary for the male model
summary(model_male)
```

```
##
## Call:
```

```
## lm(formula = q9_s ~ group * wave + q1_s + q2_s + q3_s + q4_s +
##     q5_s + q6_s + q7_s + q8_s + q10_s + manager + ethnic + stratum +
##     sector + female_mgr + progress + num_managers + org_nor +
##     los, data = male)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4589 -1.6825  0.3726  1.5440  3.1716
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.218e-01  5.234e-01   0.424  0.67176
## group1       -9.927e-02  1.307e-01  -0.760  0.44759
## wave2        -1.390e-01  1.315e-01  -1.057  0.29064
## q1_s          1.064e-01  2.354e-02   4.518 6.81e-06 ***
## q2_s          1.114e-01  2.798e-02   3.982 7.20e-05 ***
## q3_s          6.467e-02  2.942e-02   2.198  0.02813 *
## q4_s          3.191e-02  2.952e-02   1.081  0.27994
## q5_s         -1.270e-02  3.139e-02  -0.404  0.68594
## q6_s          3.910e-02  2.818e-02   1.388  0.16551
## q7_s          7.244e-02  3.139e-02   2.308  0.02118 *
## q8_s          2.397e-01  2.946e-02   8.137 9.19e-16 ***
## q10_s         8.824e-02  2.811e-02   3.139  0.00173 **
## manager1      6.561e-02  1.056e-01   0.621  0.53444
## ethnic        1.309e-02  1.965e-02   0.666  0.50557
## stratum2     -8.093e-02  1.272e-01  -0.636  0.52484
## sector        1.359e-03  9.816e-03   0.138  0.88994
## female_mgr   -3.371e-04  2.005e-01  -0.002  0.99866
## progress      2.648e-01  4.943e-01   0.536  0.59229
## num_managers  1.178e-03  2.120e-03   0.556  0.57845
## org_nor      -9.278e-05  2.308e-04  -0.402  0.68770
## los           1.080e-01  5.967e-02   1.810  0.07049 .
## group1:wave2  1.810e-01  1.965e-01   0.921  0.35711
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.772 on 1331 degrees of freedom
## Multiple R-squared:  0.151,  Adjusted R-squared:  0.1376
## F-statistic: 11.27 on 21 and 1331 DF,  p-value: < 2.2e-16
```

Section 2.3: Model Summary Statistics for Females with Imputed Missing Data

```r
# Fitting the model for females
model_female <- lm(q9_s ~ group * wave + q1_s + q2_s + q3_s + q4_s + q5_s + q6_s +
                 q7_s + q8_s + q10_s + manager + ethnic + stratum + sector +
                 female_mgr + progress + num_managers + org_nor + los, data = female)

# Print the summary for the female model
summary(model_female)
```

```
##
## Call:
## lm(formula = q9_s ~ group * wave + q1_s + q2_s + q3_s + q4_s +
##     q5_s + q6_s + q7_s + q8_s + q10_s + manager + ethnic + stratum +
##     sector + female_mgr + progress + num_managers + org_nor +
```

```
##     los, data = female)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9945 -1.4504  0.2189  1.3717  3.9388
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.1003643  0.3071390   0.327  0.74387
## group1       0.0258501  0.1046337   0.247  0.80489
## wave2        0.0913464  0.1045257   0.874  0.38227
## q1_s         0.0970104  0.0160098   6.059 1.63e-09 ***
## q2_s         0.0923422  0.0212844   4.339 1.51e-05 ***
## q3_s         0.0958465  0.0230027   4.167 3.22e-05 ***
## q4_s         0.0565734  0.0225344   2.511  0.01213 *
## q5_s        -0.0345177  0.0246557  -1.400  0.16167
## q6_s         0.0147812  0.0231986   0.637  0.52409
## q7_s         0.0429760  0.0246327   1.745  0.08119 .
## q8_s         0.2854173  0.0227123  12.567  < 2e-16 ***
## q10_s        0.1389875  0.0222364   6.250 4.99e-10 ***
## manager1     0.0498039  0.0818125   0.609  0.54275
## ethnic       0.0083656  0.0161202   0.519  0.60385
## stratum2     0.2739447  0.1043198   2.626  0.00871 **
## sector       0.0097618  0.0052582   1.857  0.06353 .
## female_mgr   0.1113423  0.1902986   0.585  0.55855
## progress    -0.0088335  0.2657511  -0.033  0.97349
## num_managers -0.0007388 0.0013167  -0.561  0.57481
## org_nor      0.0000161  0.0002691   0.060  0.95229
## los          0.1107381  0.0442770   2.501  0.01246 *
## group1:wave2 -0.1853134 0.1513411  -1.224  0.22092
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.689 on 2004 degrees of freedom
## Multiple R-squared:  0.2249, Adjusted R-squared:  0.2167
## F-statistic: 27.68 on 21 and 2004 DF,  p-value: < 2.2e-16
```

Section 2.4: Model Summary Statistics for Males with Dropped Missing Data

```
# Fitting the model for males
male <- subset(df, gender == 1)
model_male <- lm(q9_s ~ group * wave + q1_s + q2_s + q3_s + q4_s + q5_s + q6_s +
                 q7_s + q8_s + q10_s + manager + ethnic + stratum + sector +
                 female_mgr + progress + num_managers + org_nor + los, data = male)

print(summary(model_male))

##
## Call:
## lm(formula = q9_s ~ group * wave + q1_s + q2_s + q3_s + q4_s +
##     q5_s + q6_s + q7_s + q8_s + q10_s + manager + ethnic + stratum +
##     sector + female_mgr + progress + num_managers + org_nor +
##     los, data = male)
##
## Residuals:
```

```
##      Min      1Q  Median      3Q     Max
## -3.5383 -1.6671  0.3677  1.5287  3.2031
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.3311283  0.5521816   0.600 0.548834
## group1        -0.0873506  0.1355300  -0.645 0.519364
## wave2         -0.1499130  0.1363786  -1.099 0.271877
## q1_s           0.1057729  0.0249159   4.245 2.35e-05 ***
## q2_s           0.1073705  0.0291291   3.686 0.000238 ***
## q3_s           0.0610078  0.0305337   1.998 0.045930 *
## q4_s           0.0277709  0.0306340   0.907 0.364827
## q5_s          -0.0186915  0.0324598  -0.576 0.564832
## q6_s           0.0494767  0.0292123   1.694 0.090577 .
## q7_s           0.0816918  0.0325011   2.514 0.012080 *
## q8_s           0.2438365  0.0306158   7.964 3.75e-15 ***
## q10_s          0.0906591  0.0290970   3.116 0.001877 **
## manager1       0.0527786  0.1094974   0.482 0.629886
## ethnic         0.0119230  0.0202584   0.589 0.556273
## stratum2      -0.0635370  0.1334056  -0.476 0.633967
## sector        -0.0001491  0.0104740  -0.014 0.988648
## female_mgr     0.0183212  0.2106904   0.087 0.930719
## progress       0.2228774  0.5255317   0.424 0.671568
## num_managers   0.0010681  0.0022573   0.473 0.636173
## org_nor       -0.0001112  0.0002406  -0.462 0.643987
## los            0.0818833  0.0623201   1.314 0.189119
## group1:wave2   0.1054151  0.2040424   0.517 0.605505
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.771 on 1233 degrees of freedom
## Multiple R-squared:  0.1568, Adjusted R-squared:  0.1425
## F-statistic: 10.92 on 21 and 1233 DF,  p-value: < 2.2e-16
```

Section 2.5: Model Summary Statistics for Females with Dropped Missing Data

```
# Fitting the model for females
female <- subset(df, gender == 0)
model_female <- lm(q9_s ~ group * wave + q1_s + q2_s + q3_s + q4_s + q5_s + q6_s +
                 q7_s + q8_s + q10_s + manager + ethnic + stratum + sector +
                female_mgr + progress + num_managers + org_nor + los, data = female)


print(summary(model_female))

##
## Call:
## lm(formula = q9_s ~ group * wave + q1_s + q2_s + q3_s + q4_s +
##     q5_s + q6_s + q7_s + q8_s + q10_s + manager + ethnic + stratum +
##     sector + female_mgr + progress + num_managers + org_nor +
##     los, data = female)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -4.0070 -1.4227  0.2163  1.3639  3.9425
```

```
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.328e-01  3.217e-01   0.724 0.469305
## group1        6.430e-02  1.082e-01   0.594 0.552253
## wave2         1.279e-01  1.087e-01   1.177 0.239268
## q1_s          9.499e-02  1.708e-02   5.562 3.05e-08 ***
## q2_s          9.669e-02  2.206e-02   4.382 1.24e-05 ***
## q3_s          9.300e-02  2.388e-02   3.894 0.000102 ***
## q4_s          6.197e-02  2.342e-02   2.646 0.008218 **
## q5_s         -3.811e-02  2.554e-02  -1.492 0.135850
## q6_s          1.028e-02  2.426e-02   0.424 0.671802
## q7_s          3.794e-02  2.553e-02   1.486 0.137332
## q8_s          2.937e-01  2.354e-02  12.478  < 2e-16 ***
## q10_s         1.406e-01  2.294e-02   6.127 1.09e-09 ***
## manager1      7.188e-02  8.450e-02   0.851 0.395073
## ethnic        5.297e-03  1.696e-02   0.312 0.754829
## stratum2      2.767e-01  1.097e-01   2.521 0.011784 *
## sector        8.727e-03  5.446e-03   1.603 0.109211
## female_mgr    1.155e-01  2.011e-01   0.574 0.566016
## progress     -1.686e-01  2.759e-01  -0.611 0.541094
## num_managers -9.033e-04  1.379e-03  -0.655 0.512654
## org_nor       3.287e-05  2.875e-04   0.114 0.908999
## los           9.927e-02  4.653e-02   2.133 0.033034 *
## group1:wave2 -2.321e-01  1.565e-01  -1.483 0.138335
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.682 on 1853 degrees of freedom
## Multiple R-squared:  0.2347, Adjusted R-squared:  0.2261
## F-statistic: 27.07 on 21 and 1853 DF,  p-value: < 2.2e-16
```

Section 2.6: Model Summary Statistics for Moderator-Stratum

```
# Fitting the model for moderator
model_df <- lm(q9_s ~ group * wave * stratum + q1_s +  q2_s + q3_s + q4_s + q5_s +
                q6_s +q7_s + q8_s + q10_s + manager + gender + ethnic + stratum +
                sector + female_mgr + progress + num_managers + org_nor + los,
                 data = df)

# Print the model summary
summary(model_df)
##
## Call:
## lm(formula = q9_s ~ group * wave * stratum + q1_s + q2_s + q3_s +
##     q4_s + q5_s + q6_s + q7_s + q8_s + q10_s + manager + gender +
##     ethnic + stratum + sector + female_mgr + progress + num_managers +
##     org_nor + los, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9739 -1.5456  0.2664  1.4369  3.8632
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)           3.460e-01  2.558e-01   1.352  0.17632    
## group1                1.333e-02  8.955e-02   0.149  0.88168    
## wave2                -3.062e-02  8.855e-02  -0.346  0.72948    
## stratum2              9.714e-02  1.443e-01   0.673  0.50079    
## q1_s                  1.026e-01  1.299e-02   7.898 3.79e-15 ***
## q2_s                  9.530e-02  1.669e-02   5.709 1.23e-08 ***
## q3_s                  8.301e-02  1.785e-02   4.649 3.45e-06 ***
## q4_s                  4.607e-02  1.762e-02   2.614  0.00898 ** 
## q5_s                 -2.581e-02  1.909e-02  -1.352  0.17643    
## q6_s                  2.666e-02  1.761e-02   1.514  0.13016    
## q7_s                  5.071e-02  1.909e-02   2.656  0.00794 ** 
## q8_s                  2.723e-01  1.772e-02  15.369  < 2e-16 ***
## q10_s                 1.185e-01  1.720e-02   6.888 6.68e-12 ***
## manager1              4.835e-02  6.318e-02   0.765  0.44415    
## gender               -5.987e-03  5.381e-02  -0.111  0.91142    
## ethnic                1.190e-02  1.214e-02   0.980  0.32714    
## sector                6.980e-03  4.530e-03   1.541  0.12346    
## female_mgr            2.158e-02  1.199e-01   0.180  0.85715    
## progress             -8.435e-02  2.220e-01  -0.380  0.70399    
## num_managers         -6.560e-04  9.590e-04  -0.684  0.49396    
## org_nor              -1.074e-05  1.549e-04  -0.069  0.94472    
## los                   1.141e-01  3.478e-02   3.280  0.00105 ** 
## group1:wave2         -9.142e-02  1.307e-01  -0.700  0.48426    
## group1:stratum2      -1.810e-01  1.987e-01  -0.911  0.36252    
## wave2:stratum2        1.390e-01  2.129e-01   0.653  0.51383    
## group1:wave2:stratum2  3.275e-01  3.046e-01   1.075  0.28240    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.719 on 3434 degrees of freedom
## Multiple R-squared:  0.1928, Adjusted R-squared:  0.1869 
## F-statistic: 32.81 on 25 and 3434 DF,  p-value: < 2.2e-16
```