Coupling Occupational Safety and Labour Productivity in Construction: A Multi-City Empirical Analysis and Predictive Optimization Framework

Subash Thanappan[1*], *Shalin* Prince[2], *Adams* Joe[3], *Chiranjeevi* Rahul Rollakanti[3], *Reena* M R[4]

[1]Professor & Head, Department of Civil Engineering, KAAF University, Accra, Ghana
[2]Programme Manager, Senior Lecturer, Department of Civil and Mechanical Engineering, Middle East College, Muscat, Oman.

[3]Senior Lecturer, Department of Civil and Mechanical Engineering, Middle East College, Muscat, Oman.

[4]Professor, Department of Midwifery, KAAF University, Accra, Ghana.

Abstract. Construction Safety Management and Labour Productivity are frequently viewed as mutually exclusive in construction management. However, there is an increasing body of research indicating that occupational safety contributes to operational stability and better employee performance. This paper assesses the quantitative association between safety and labour productivity by using a multi city dataset which comprises 1275 observation for construction projects across six metropolitan areas. Indicators for both safety and operationally were evaluated to determine how they contributed to the labour productivity through use of key indicators including Personal Protective Equipment (PPE) compliance, hours of safety training, frequency of accidents, number of near misses, number of hours worked at overtime, hours lost due to work stoppages, and number of skilled workers. The predictive modeling approach utilized included Quantile Regression, Recursive Least Squares, Generalized Estimating Equations, and Weighted Least Squares with five fold cross validation to test for robustness. The results show that all safety compliance variables positively correlated with productivity, specifically PPE adherence and training intensity. The models indicated that disruption related variables such as accident frequency, hours worked overtime, and hours lost due to work stoppages negatively correlated with productivity. Overall, the models explained around 72-74% of the variation in productivity in unseen data and therefore have significant predictive capabilities.

---

[*] Corresponding Author: thanappansubash@gmail.com

1 Introduction

The construction sector is among the most labor-intensive globally. As well, it is heavily influenced by the behavior of workers, workplace safety culture, and the way construction projects are managed. Although productivity improvements continue to be at the center of discussions related to the construction industry and public policy, occupational safety continues to be an important priority due to the number of accidents occurring in the construction sector and the economic loss caused by lost time injury and near miss events. In spite of significant advancements in regulations, construction still experiences a considerable number of incidents, near misses, and work stoppages, which each impact project performance. Furthermore, productivity improvements and safety management have been traditionally viewed as two independent management domains; however, they should be viewed as parts of a larger system.

There has been a longstanding assumption in construction management that improving safety will negatively affect productivity [3,4]. This assumption is based on the belief that the requirement for compliance with regulations, training, and the use of personal protective equipment (PPE), etc. will delay the progress of construction activities and increase costs. Nevertheless, recent behavioral and empirical research challenges this assumption. Workplaces where unsafe behaviors occur lead to increased amounts of rework, absenteeism, fatigue, delays, damaged materials, and risk to a company's reputation [5], which ultimately result in decreased levels of productivity. Consequently, safety initiatives not only serve to decrease the number of injuries that occur but they may also help to maintain operational output. Therefore, the issue is no longer if safety is important, but rather how safety affects productivity, and to what degree [6].

Although research has provided some insights into the relationships between safety and productivity, much of the existing body of knowledge provides only limited information about these relationships. A number of prior studies have analyzed specific aspects of safety and productivity including the effect of accident frequency, fatigue, and overtime exposure [7]. Other studies have investigated the effect of safety climate and worker behavioral compliance. However, three main gaps exist. First, many of the prior studies used relatively small sample sizes, or were conducted at a single site, making it difficult to generalize their results. Second, studies examining the interaction between safety and productivity were generally limited to correlational analysis that lacked robust predictive validation. Third, little development has occurred in the area of developing integrated modeling frameworks that can take safety metrics and convert them into tools for optimizing productivity.

Construction systems are complex, socio-technical environments. Labor productivity is influenced not only by the technical skills of workers and resource allocation, but also by behavioral compliance, fatigue accumulation, and the frequency of disruptions. For example, excessive overtime can increase the amount of fatigue experienced by workers, lower the cognitive ability of workers, and increase the likelihood of errors being made. On the other hand, safety training can improve the level of procedural compliance, coordination, and anticipation of hazards, which can lead to reduced amounts of stoppages and improved effective work hours. Based on these interdependencies, it appears that safety and productivity are dynamically coupled, rather than independently determined [9]. Integrated empirical investigations must be undertaken in regions that are experiencing rapid urbanization, since project volumes are high, and workforce diversity is high. Multi-city data sets provide researchers the opportunity to investigate the structural consistencies of safety-productivity coupling across different geographic areas, while isolating the behavioral factors that influence this coupling. Understanding whether safety-productivity coupling is location dependent or behaviorally dependent has significant implications for managerial standardization and policy design [10].

In addition to understanding the associations between safety and productivity, the industry needs to develop predictive capabilities. Decision makers need to know how increases in PPE compliance, or the hours spent in training, will relate to measurable productivity improvements [11]. Predictive modeling allows decision makers to evaluate the marginal effects, threshold effects, and stability of various operational conditions.

To address the gaps identified above, this study conducts a comprehensive empirical analysis of construction project data from six major urban regions [12]. Using a multi-model predictive framework, this study examines the influence of PPE compliance, safety training hours, near-miss frequency, accident rate, overtime exposure, absenteeism, and workforce characteristics on labor productivity [13]. To go beyond descriptive analysis, this study validates the predictive performance of the models proposed and introduces the concept of a Safety Productivity Coupling Index (SPCI) for managerial applications [14]. The fundamental premise of this study is that occupational safety is not just a regulatory constraint, but a mechanism for maintaining operational performance. By investigating the relationships between safety and productivity using large-scale empirical data, this study makes contributions to both the theoretical and practical aspects of construction management.

Research Questions

- What is the nature and strength of the relationship between occupational safety indicators and labour productivity in construction projects?
- To what extent can labour productivity be reliably predicted using safety-related and workforce variables across multiple project locations?
- Can safety performance variables be integrated into a measurable framework that supports simultaneous safety enhancement and productivity optimization?

Research Objectives

- To empirically quantify the influence of key occupational safety indicators on labour productivity using multi-city construction project data.
- To develop and validate predictive models that estimate labour productivity based on safety compliance, training intensity, and disruption-related variables.
- To construct an integrated safety productivity coupling framework that translates safety improvements into actionable productivity gains for construction management practice.

2 Literature Review

2.1 Occupational Safety as a Determinant of Construction Productivity

Traditional analysis of construction productivity has generally focused upon technical and economic factors, including labor skills, available materials, schedule efficiency, and equipment usage. Within this paradigm, worker safety has historically been treated as an enforceable regulatory obligation, as opposed to a production performance factor. Accordingly, safety programs were typically viewed as indirect costs in the sense that they could limit the rate at which projects can be completed. The fact that safety and productivity are perceived to compete with one another, especially within time-constrained project environments, is consistent with the above view. Nevertheless, several recent empirical investigations demonstrate that adverse working conditions cause real disruption to productive operations. Disruption from accidents, near misses, and unsafe behaviors will disrupt the continuity of workflow, increase rework, increase absenteeism, and create administrative delay. All of these forms of disruption will decrease the number of actual labor hours that are effectively utilized and decrease the efficiency of coordination among workers. Therefore, in addition to contributing to output increases (acceleration), safety performance may also contribute to increased productivity by providing stability to the process of creating value. Recent studies show that organizations that actively engage in managing workplace safety provide greater behavioral discipline, anticipate hazards, and provide more structure to task execution. Examples of such proactive measures include personal protective equipment compliance, regular safety training, and the level of supervision provided to employees. Organizations that have high levels of compliance with the above measures tend to have fewer micro-disruptions in their workflows, improve employee morale, and experience fewer inefficiencies due to stress. The organizations that establish controlled overtime policies will experience a reduction in cognitive impairment caused by fatigue, thereby reducing the incidence of errors and improving the accuracy of tasks. While the mechanisms described above appear to indicate a positive relationship between safety and productivity, the size and consistency of the effect remain unquantifiable in order to utilize large data sets for empirical analysis.

2.2 Predictive Modelling and Integrated Safety–Performance Frameworks

Recent advances in construction analytics have made it possible to use machine learning and statistical techniques to predict productivity based on predictive models. Several studies have used variables such as workforce characteristics (e.g., age, experience), environmental conditions (e.g., weather), and scheduling variables (e.g., time of day, week, month) to estimate performance outcomes. Although recent studies have demonstrated some significant progress in the area of productivity prediction, safety-related variables are frequently considered secondary predictors rather than primary explanatory factors. Many studies are restricted to correlation analysis with no validation of predictive stability. Therefore, without cross validation and multi model comparison, it is very difficult to determine if the relationship between safety productivity is consistent across different contexts.

The integrated modelling approach remains underdeveloped. Very few studies attempt to translate safety metrics into actionable managerial indices that quantify the effect of incremental improvements in compliance or training on productivity. The lack of decision-support frameworks significantly limits practical applicability for project managers who seek to justify their safety investments in economic terms. Very little attention has been paid to the development of multi-city or multi-locations datasets that will enable an evaluation of structural consistency across geographic environments. Therefore, there is a need for comprehensive empirical research that includes both statistical validation, predictive modelling, and managerial interpretation of results within a unified framework.

2.3 Research Gap and Problem Statement

Although it is well recognized by prior research that there are strong interdependencies between occupational safety and labor productivity in construction settings, empirical evidence that quantifies these interdependencies remains sparse and segmented. Most prior research has been based on small sample sizes, single-project cases, and/or perceptual surveys that limit the generalizability of their findings to specific geographically and operationally distinct contexts. Safety indicators (such as PPE use, safety training hours per employee, near miss rates, accident rates, and overtime) have generally been studied in isolation from one another and from each other as a part of a systemic approach to analyzing them. Many prior studies have focused on reporting descriptive statistics or performing simple correlations without subjecting their findings to a rigorous predictive validation process. Due to the scarcity of large-scale empirical data sets collected over multiple locations and the absence of cross-validated modeling techniques, confidence in the magnitude, consistency, and robustness of the relationships between safety and productivity remains uncertain.

The failure to develop methodologies for studying the interdependency between safety and productivity in construction projects results in a more direct practical issue. Construction projects take place in disruption sensitive environments. Disruptions to the flow of work in construction projects can occur due to a variety of factors, including accidents, fatigue, absenteeism, and rework. As previously noted theoretically, unsafe working conditions reduce the stability of operations and therefore, the efficiency of production. However, project managers do not have empirically validated analytical tools that enable them to quantify the impact of incremental improvements in safety performance on measurable increases in labor productivity. Therefore, without an analytical tool that couples safety with productivity, safety management will continue to be viewed by project managers as a compliance driven process rather than as a strategic performance driver. To fill this gap, there needs to be a data-driven empirical framework developed that can model safety and labor productivity at a project level while also allowing for comparisons of productivity and safety across different geographic locations.

3. Methodology

3.1 Research Design

This study will use a quantitative explanatory design to determine the structural relationship between occupational safety indicators (Safety Compliance, Disruption Intensity) and labour productivity within construction projects. In order to achieve this, the research model combines descriptive analysis, correlation analysis, and multi-model predictive regression to allow for both the interpretation of results and statistical robustness. To increase the external validity of the study, while reducing the risk of location specific bias, the research used a multicity dataset. By allowing for the systematic investigation of how the joint combination of safety compliance, disruption intensity, and workforce characteristics influences productivity outcomes, this research design provides a comprehensive understanding of the Impact of these variables on productivity.

The methodology used is sequential. Initially, an exploratory statistical analysis was performed to gain insight into the distributional properties of the data as well as interactions amongst variables. Next, the development and validation of predictive models through out of sample testing was completed to ensure generalisability. The standardized coefficients were used to develop an Integrated safety productivity coupling Index. The stepwise progression of the research design ensures that findings go beyond simple correlation and are predicatively Reliable, therefore providing managerial applicability.

3.2 Data Source and Sampling Framework

The dataset for this study includes data on 1,275 observations from current construction projects throughout 6 metropolitan areas. This dataset has been collected from a combination of site inspection files, safety audit files, and project completion reports. These datasets have been grouped together according to common reporting frequencies to provide comparable data for all sizes of projects. Stratified random sampling is used in this research to account for variation among project size, personnel mix, and complexity of operations. By using stratified random sampling, the diversity of samples increases the ability to generalize results across different contractor categories as well as across different levels of project intensity. In addition, by collecting data from several locations simultaneously, the potential for location specific practice or regulation influencing the outcome of the analysis decreases. Finally, with a total of 1,275 observations available for analysis, the amount of data available to analyze increases statistical power and allows for reliable estimates to be generated through various forms of regression analysis.

3.3 Variable Operationalization

Labour productivity was defined as a normalized productivity index calculated as achieved output relative to planned output within a defined cycle is shown in Table 1. Normalization ensures comparability across heterogeneous construction tasks. Independent variables were categorized into safety compliance, disruption indicators, and workforce operational factors. Safety compliance variables included PPE compliance percentage and safety training hours per worker. Disruption indicators included near-miss frequency, accident rate, work stoppage hours, absenteeism percentage, and rework rate. Workforce-operational variables included overtime exposure, skilled labour ratio, workforce size, and schedule variance. All variables were treated as continuous and standardized where necessary to facilitate coefficient comparison.

Table 1. Variable Description

Category	**Variable**	**Description**	**Measurement Type**
Dependent	Labour Productivity Index	Achieved output / Planned output	Continuous
Safety	PPE Compliance (%)	Percentage of compliant workers	Continuous
Safety	Safety Training Hours	Average hours per worker	Continuous
Disruption	Near-Miss Frequency	Recorded near misses per exposure unit	Continuous
Disruption	Accident Rate	Accidents per exposure unit	Continuous
Disruption	Work Stoppage Hours	Hours lost due to disruptions	Continuous
Workforce	Overtime Hours	Average weekly overtime	Continuous
Workforce	Skilled Labour Ratio (%)	Percentage of skilled workers	Continuous

3.4 Data Preprocessing and Diagnostic Testing

Before modelling, data quality screening was conducted to ensure reliability. Missing values were evaluated using frequency diagnostics. Records with incomplete critical fields were removed to preserve analytical integrity. Outliers were assessed using interquartile range thresholds and standardized z-scores. Extreme values were retained when operationally plausible to preserve real-world variability. Multicollinearity among predictors was evaluated using variance inflation factors. All VIF values remained below accepted thresholds, confirming independence among explanatory variables. Residual normality and homoscedasticity were examined through graphical diagnostics following model estimation. These checks ensured that regression assumptions were satisfied and that coefficient estimates remained unbiased.

3.5 Development of the Safety Productivity Coupling Index

A composite SPCI was developed using standardized regression coefficients from validated models. Indicators that contribute positively to safety (e.g., PPE compliance, safety training) are weighted in a positive direction; whereas indicators related to disruptions (e.g., near misses, overtime exposure) are weighted in a negative direction. The resultant index is a single number that represents the degree to which operational safety and performance are integrated. As a result of this methodology-based process, the output of regression analysis is converted into an actionable management support tool for making decisions regarding investments in safety. The SPCI values will allow projects to be categorized by their degree of coupling (i.e., high or low coupling), enabling safety professionals to prioritize where they need to intervene. The SPCI allows for the practical application of research results into strategies for allocating limited resources.

3.6 Machine learning algorithms

Statistical models that examine the relationship between occupational safety indicators and labor productivity were developed using advanced machine-learning based methods for predicting generalizations with minimum errors and high levels of robustness when comparing data sets. Multiple algorithms have been selected to ensure that all findings are independent of modeling assumptions and that the structural relationships found among the variables will be consistent regardless of how the parameters were estimated.

Quantile Regression is an algorithm that was used to identify the potential distributions in the data set that may produce non-linearities or other forms of asymmetric distributions. Recursive Least Squares (RLS) was used to establish the reliability of the parameter estimates in each step of the recursive process. Generalized Estimating Equations (GEE) were employed to account for potential correlations among observations from the same cluster. Lastly, Weighted Least Squares (WLS) was applied to address issues related to heterogeneity. The use of multiple techniques is a critical aspect of increasing both the methodological rigor and the confidence in the results.

3.6.1 Quantile Regression

Quantile Regression is an extension of linear regression that estimates conditional quantiles of the dependent variable rather than focusing solely on the conditional mean. Instead of minimizing squared residuals, it minimizes weighted absolute deviations, allowing the model to estimate relationships at specific points of the productivity distribution. This approach is particularly useful when the dependent variable exhibits non-normality, skewness, or heterogeneous variance.

In construction productivity analysis, performance variability may arise due to operational disruptions, safety incidents, or workforce imbalances. Quantile Regression provides a more comprehensive understanding of predictor influence across different productivity levels, rather than assuming uniform effects. By estimating relationships at a specified quantile, the model offers robustness against outliers and extreme observations, improving stability in safety performance estimation.

3.6.2 Recursive Least Squares (RLS)

Recursive Least Squares is an adaptive estimation algorithm that updates regression coefficients iteratively as new observations are incorporated. Unlike traditional batch-based ordinary least squares, RLS recalculates parameters sequentially, minimizing cumulative squared error over successive updates. This iterative mechanism allows assessment of parameter stability and sensitivity across observation subsets. The strength of RLS lies in its ability to evaluate whether predictor effects remain consistent as data accumulate. In the context of construction safety modelling, this approach helps verify whether relationships between safety indicators and productivity are structurally stable rather than sample-specific.

RLS is particularly valuable for examining dynamic systems or datasets where relationships may gradually evolve.

3.6.3 Generalized Estimating Equations (GEE)

Generalized Estimating Equations extend generalized linear models to account for correlated or clustered observations. In multi-location construction datasets, observations within the same city or project environment may share similarities due to regulatory, managerial, or climatic factors. Standard regression assumes independence among observations, which may lead to underestimated standard errors if clustering exists. GEE addresses this issue by specifying a working correlation structure to model intra-cluster dependence. By incorporating clustering effects, GEE provides more reliable parameter estimates and robust standard errors. This approach strengthens inference validity when analyzing productivity data collected from multiple geographic regions. The use of GEE ensures that safety productivity relationships are not overstated due to hidden correlation among grouped observations.

3.6.4 Weighted Least Squares (WLS)

Weighted Least Squares is a regression technique designed to correct heteroscedasticity, a condition in which the variance of residuals is not constant across observations. In construction datasets, projects with larger workforce sizes or higher disruption levels may exhibit greater variability in productivity outcomes. Ordinary least squares assume equal error variance, which may produce inefficient estimates under heteroscedastic conditions. WLS addresses this limitation by assigning weights inversely proportional to the variance of each observation. Observations with higher variability receive lower weights, while more stable observations contribute proportionally more to coefficient estimation. This adjustment improves estimation efficiency and enhances model reliability. In safety productivity modelling, WLS ensures that variance instability does not distort the measured influence of safety indicators on labour performance.

3.7 Cross-Validation Strategy

To improve the reliability of the results and to avoid over-reliance upon a single train-test split a K-Fold Cross-Validation methodology was applied. The database was randomly divided into five equal segments. For each segment of this process, a total of four were assigned for model development and one for validation. This cycle of development and validation was performed five different times, with the result being that each segment was utilized once as the validation set. These performance measures were calculated for each iteration and averaged to provide reliable estimates of predictive accuracy. The use of cross-validation provides an opportunity to assess the potential for sample bias and to strengthen the generalization of the results. An evaluation of the cross-validation performance of the model was conducted through an examination of the R^2, RMSE, and MAE. The stability of the model was evaluated through an examination of the dispersion (standard deviation) of these measures across the multiple iterations of the cross-validation. A low level of dispersion would indicate that the model does not produce significantly different results based upon the particular division of the data. The application of the cross-validation methodology, both the credibility of the methodologies employed to derive the results and the structural basis of the relationship between safety and productivity within the data, can be confirmed.

Table 2. Cross-Validation Configuration

Parameter	Specification
Validation Method	k-fold cross-validation
Number of Folds (k)	5
Training Proportion per Fold	80%
Validation Proportion per Fold	20%
Evaluation Metrics	R^2, RMSE, MAE
Stability Assessment	Mean ± Standard Deviation across folds

3.8 Hyperparameter Configuration

While both models chosen are regression-based, each has additional required configuration settings that determine how the model estimates, and these are presented in Table 3. The Quantile Regression model used a fixed quantile (q=0.6) to reflect central-to-higher percentiles of productivity performance. To ensure numerical convergence when sequentially estimating coefficients, recursive least squares models require the user to specify an initial vector of coefficients and update rules. A working correlation structure, which accounts for the clustering that occurs across locations, is also specified for generalized estimating equations. The weights assigned in weighted least squares models were based on the inverse of the estimated residual variance. Model parameter configurations were determined by minimizing cross-validation errors in predictions. Configurations that provided consistent R^2 and low root mean square errors (RMSE) in cross-validation were kept. The control of model parameter specifications helps to prevent model overfitting and ensures that the amount of complexity in the model does not exceed the number of data points.

Table 3. Hyperparameter Settings for Implemented Models

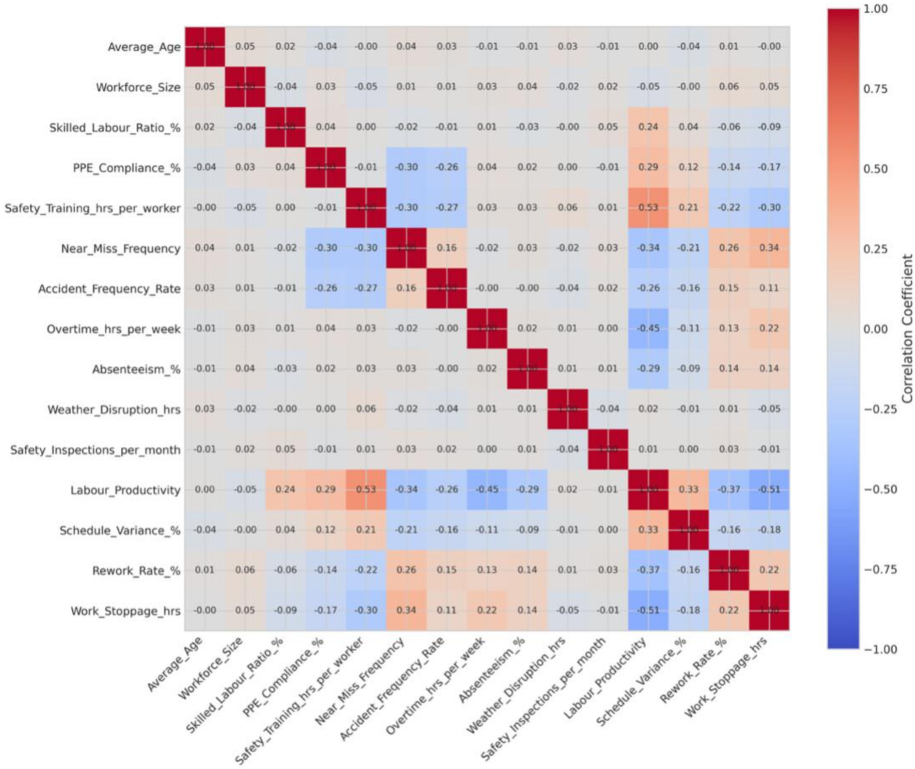
Model	Key Hyperparameter	Selected Configuration
Quantile Regression	Quantile Level (q)	0.6
Recursive Least Squares	Initialization Scheme	Stable recursive update
Recursive Least Squares	Update Mechanism	Sequential least squares minimization
GEE	Working Correlation Structure	Location-based clustering
Weighted Least Squares	Weighting Scheme	Inverse residual variance

3.9 Robustness and sensitivity analysis

To add to the reliability of findings, robustness and sensitivity analyses were conducted. Model consistency was first assessed by compiling coefficient direction and magnitude across Quantile Regression, recursive least squares, Generalized Estimating Equations, and Weighted Least Squares. For all models, the signs of key safety indicators, including PPE compliance, safety training, near-miss frequency, and overtime exposure, remained consistent. The narrow range of test R^2 values reported in the performance summary confirms the structural stability of the safety-productivity relationship. A sensitivity analysis was also performed by simulating marginal changes in key safety indicators to observe their effect on predicted productivity levels. Positive changes in productivity estimates occurred with incremental increases in PPE compliance and training hours. However, positive changes in productivity estimates did not occur with either increments in overtime or in near miss frequency. These simulations provide evidence of the practical interpretability of regression coefficients and support the development of the safety–production coupling index. The robustness of findings across models and stability under simulated variation demonstrate that the results are not dependent on the model but rather reflect inherent relationships within the dataset

3.10 Heatmap

The data in Figure 1 illustrates a Pearson correlation heatmap to show how various workforce characteristics and safety indicators are related to labor disruption variables and labor productivity. A warmer tone in the heat map illustrates positive correlation, while a cooler tone represents a negative correlation. Labor productivity has the largest positive correlation with safety training hours per employee (r = 0.53), PPE compliance (r =0.29), and the skilled labor ratio (r= 0.24). This indicates that structured training programs, high levels of compliance, and workforce composition skills positively impact the operational output. Conversely, there were large negative correlations for productivity for overtime hours (r = −0.45), work stoppage hours (r =−0.51), near miss frequency (r = −0.34), absenteeism (r =−0.29), and accident frequency rate (r =−0.26). These coefficient values also suggest that factors which cause disruptions and lead to fatigue have a greater detrimental effect on productivity than demographics such as average age or workforce size.



	Average_Age	Workforce_Size	Skilled_Labour_Ratio_%	PPE_Compliance_%	Safety_Training_hrs_per_worker	Near_Miss_Frequency	Accident_Frequency_Rate	Overtime_hrs_per_week	Absenteeism_%	Weather_Disruption_hrs	Safety_Inspections_per_month	Labour_Productivity	Schedule_Variance_%	Rework_Rate_%	Work_Stoppage_hrs
Average_Age	1.00	0.05	0.02	-0.04	-0.00	0.04	0.03	-0.01	-0.01	0.03	-0.01	0.00	-0.04	0.01	-0.00
Workforce_Size	0.05	1.00	-0.04	0.03	-0.05	0.01	0.01	0.03	0.04	-0.02	0.02	-0.05	-0.00	0.06	0.05
Skilled_Labour_Ratio_%	0.02	-0.04	1.00	0.04	0.00	-0.02	-0.01	0.01	-0.03	-0.00	0.05	0.24	0.04	-0.06	-0.09
PPE_Compliance_%	-0.04	0.03	0.04	1.00	-0.01	-0.30	-0.26	0.04	0.02	0.00	-0.01	0.29	0.12	-0.14	-0.17
Safety_Training_hrs_per_worker	-0.00	-0.05	0.00	-0.01	1.00	-0.30	-0.27	0.03	0.03	0.06	0.01	0.53	0.21	-0.22	-0.30
Near_Miss_Frequency	0.04	0.01	-0.02	-0.30	-0.30	1.00	0.16	-0.02	0.03	-0.02	0.03	-0.34	-0.21	0.26	0.34
Accident_Frequency_Rate	0.03	0.01	-0.01	-0.26	-0.27	0.16	1.00	-0.00	-0.00	-0.04	0.02	-0.26	-0.16	0.15	0.11
Overtime_hrs_per_week	-0.01	0.03	0.01	0.04	0.03	-0.02	-0.00	1.00	0.02	0.01	0.00	-0.45	-0.11	0.13	0.22
Absenteeism_%	-0.01	0.04	-0.03	0.02	0.03	0.03	-0.00	0.02	1.00	0.01	0.01	-0.29	-0.09	0.14	0.14
Weather_Disruption_hrs	0.03	-0.02	-0.00	0.00	0.06	-0.02	-0.04	0.01	0.01	1.00	-0.04	0.02	-0.01	0.01	-0.05
Safety_Inspections_per_month	-0.01	0.02	0.05	-0.01	0.01	0.03	0.02	0.00	0.01	-0.04	1.00	0.01	0.00	0.03	-0.01
Labour_Productivity	0.00	-0.05	0.24	0.29	0.53	-0.34	-0.26	-0.45	-0.29	0.02	0.01	1.00	0.33	-0.37	-0.51
Schedule_Variance_%	-0.04	-0.00	0.04	0.12	0.21	-0.21	-0.16	-0.11	-0.09	-0.01	0.00	0.33	1.00	-0.16	-0.18
Rework_Rate_%	0.01	0.06	-0.06	-0.14	-0.22	0.26	0.15	0.13	0.14	0.01	0.03	-0.37	-0.16	1.00	0.22
Work_Stoppage_hrs	-0.00	0.05	-0.09	-0.17	-0.30	0.34	0.11	0.22	0.14	-0.05	-0.01	-0.51	-0.18	0.22	1.00

Fig 1. Pearson correlation heatmap illustrating relationships among safety, workforce, disruption, and labour productivity variables.

Additional inter-variable relationships illustrate a structural relationship between elements in the safety-productivity system. There was a strong negative correlation between safety training/PPE compliance and accident frequency, indicating that the use of preventive measures reduces operational incident rates. The near miss frequency was strongly positively correlated with the rework rate and work stoppage hours, indicating that it serves as an early indication of potential operational disruptions. Exposure to overtime had a moderate relationship with absenteeism and stoppages, confirming the presence of a fatigue disruption pathway. The schedule variance had a positive correlation with productivity (r = 0.33), indicating that when schedules are adhered to, employees can be more productive. The data from the heat map confirm that the compliance variables representing safety have a positive relationship with productivity and the indicators representing disruptions have a negative relationship with productivity, thus supporting the hypothesis that there exists a coupled relationship between safety and productivity.

4. Results and Discussion

4.1 Model Performance

Table 4 demonstrates the comparable training and testing performance for the four regression-based predictive models employed to predict labor productivity based on a combination of safety and workforce variables. With respect to their training performance, RLS, GEE, and WLS show nearly equivalent performance; each has an R^2 value of .777 and an RMSE value of .0366. While the R^2 value for Quantile Regression (q = .6) is lower at .7579 and the RMSE is higher at .0381 than that of the other three models, it still exhibits robust performance. Each of the four models recorded a WMAPE (Weighted Mean Absolute Percentage Error) value less than 3%, which reflects very low relative prediction errors. Furthermore, the VAF (Variance Accounted For) values for the top performing models exceed 77%, and the RSR (Root Sum of Squared Residuals) values for all four models are less than .50. This evidence indicates that all of the models have strong in-sample explanatory capacity.

The testing performance of the models is similar to their training performance, with only minor declines in R^2 values, which indicates strong generalizability as well as minimal evidence of over-fitting. Of the four models, WLS demonstrated the strongest predictive accuracy during testing (R^2 = .7392; RMSE = .0362), which was very close to the predictive accuracy of both RLS and GEE. The latter models were also extremely close in terms of their predictive accuracies. In contrast, the predictive accuracy of Quantile Regression (R^2 = .7213) was somewhat lower than that of the three former models, although it maintained competitive error metric values. The small difference in performance between training and testing performance for each of the models' performance indicators further supports the presence of structural stability in the relationship between safety and productivity. These findings provide support for the robustness and reliability of the integrated predictive framework.

Table 4. Comparative training and testing performance based models

Model	Dataset	R^2	RMSE	MAE	WMAPE (%)	NSE	VAF (%)	RSR	LMI
Quantile Regression	Train	0.7579	0.0381	0.0298	2.95	0.7579	77.35	0.4921	0.5206
	Test	0.7213	0.0374	0.0303	2.99	0.7213	73.42	0.5279	0.4734
Recursive Least Squares	Train	0.7773	0.0366	0.0291	2.88	0.7773	77.73	0.4719	0.5327
	Test	0.7382	0.0362	0.0291	2.88	0.7382	73.88	0.5117	0.4928
Generalized Estimating Equations	Train	0.7771	0.0366	0.0291	2.88	0.7771	77.73	0.4721	0.5327
	Test	0.7375	0.0363	0.0292	2.89	0.7375	73.89	0.5123	0.4926
Weighted Least Squares	Train	0.7771	0.0366	0.0291	2.88	0.7771	77.71	0.4721	0.5325
	Test	0.7392	0.0362	0.0290	2.87	0.7392	73.97	0.5107	0.4948

4.2 Regional comparison of labour productivity

Figure 2 displays the average labour productivity indexes for the six largest urban areas in India, Kochi, Bengaluru, Chennai, Hyderabad, Coimbatore, and Mysore. Labour productivity is measured to be nearly consistent across all urban areas having an average productivity index ranging from 1.000 to 1.012. The urban area of Kochi has the highest productivity average at 1.012; the urban areas of Bengaluru and Chennai have slightly lower productivities at 1.011 each; and the urban area of Mysore has the lowest at 1.000. Although there may be some variation in the size of projects undertaken in different cities, the size and composition of workforces, and local regulatory requirements, the narrow spread of productivity values indicate that overall productivities are generally very similar among cities.

Thus, it appears that geography alone is not responsible for significant performance variability between locations within this data set. Rather, it seems that operational and safety-related influences are likely to have had greater impacts on productivity than geographical factors. Therefore, the methodological choice to treat safety compliance and disruptions as the primary explanatory drivers of productivity, and not as location-specific influences is supported.

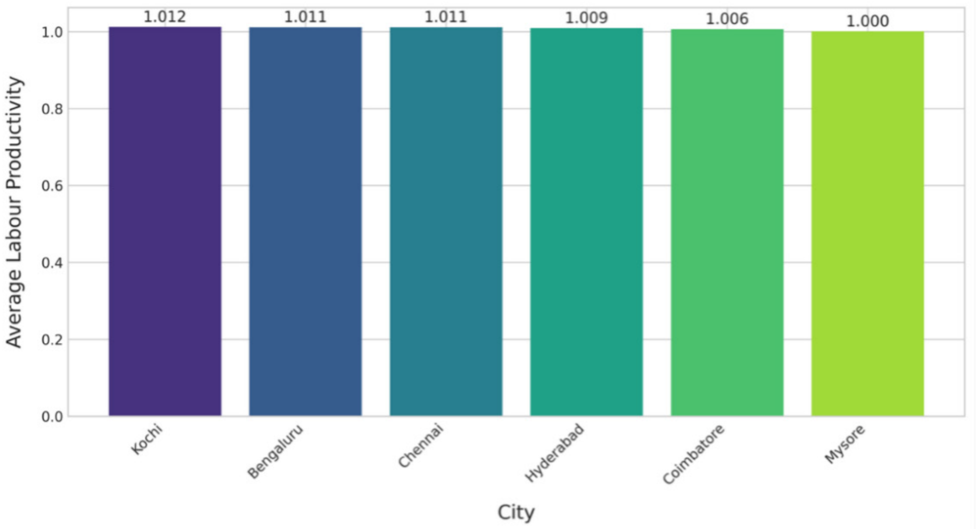


Fig 2. Average labour productivity index across six urban construction regions.

4.3 Gender-Based distribution of Labour Productivity

The data in Figure 3 shows how the labor productivity of male and female workers is distributed. The medians for each group were very close to one another, as were their respective IQRs, which indicate little difference in dispersion (variability) in productivity among the two groups. The mean productivities shown inside the boxes were also very similar between the male and female groups.

In addition to the fact that the majority of the data points fall near the median line, the small number of outliers in both groups at the low end of productivity indicate that there have been the rare instances where productivity has dipped, most likely due to some form of project-specific disruption or event, as opposed to being related to any structural issues based upon gender. The overall pattern of the data indicates that labor productivity does not differ much between males and females in this data set. Therefore, it appears that operational, safety, and disruption factors have had a larger impact on the productivity output of these individuals than did their demographic information.

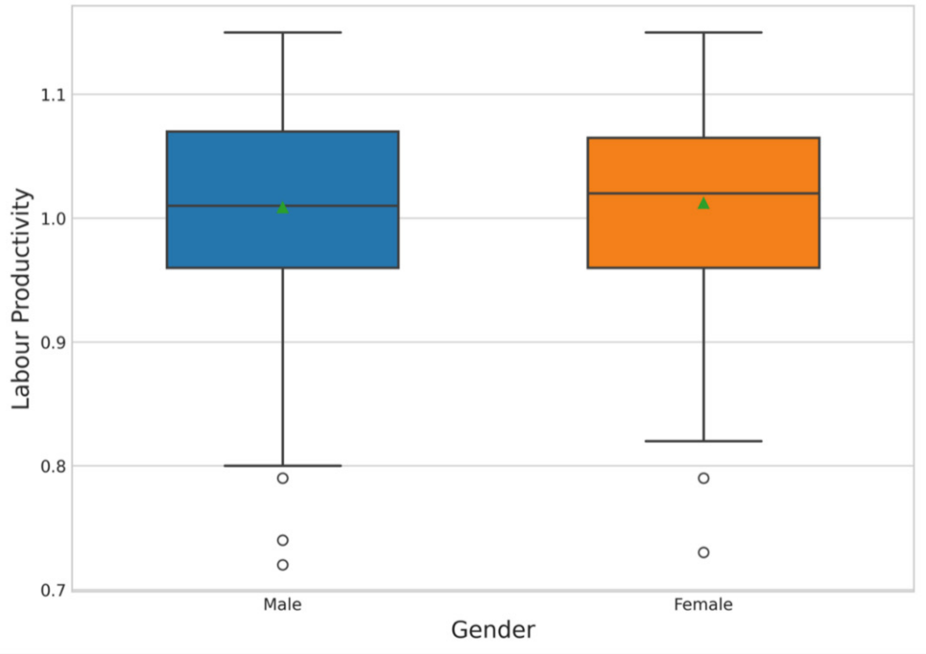


Fig 3. Boxplot comparison of labour productivity distribution by gender.

4.4 Relationship between accident frequency and labour productivity

The illustration in Figure 4 indicates the general relationship between accident frequency and labour productivity. There is an inverse association between accident frequency (with accident frequency increasing) and labour productivity (with productivity decreasing). While this is not a perfect linear relationship, the majority of the higher productivity values are at the lower accident frequencies; the accident frequency has been shown to increase with the increased dispersion and decreased level of productivity. These trends support the finding from the regression that accident frequency is a negative influence on operational stability or performance.

In addition to reducing mean labour productivity, the presence of accidents will result in a greater degree of variation in labour output. Due to these operational interruptions, administrative delays, and loss of manpower, projects with a high frequency of accidents will have greater variability in their labour output. The lack of productivity clusters at the higher accident frequencies further supports the notion that preventing accidents stabilizes production rather than being simply a regulatory obligation.

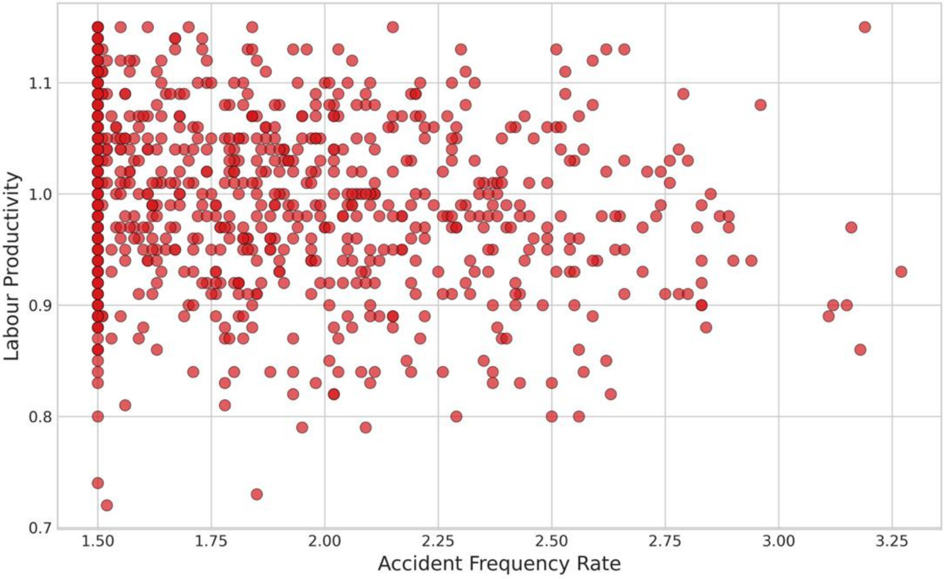


Fig. 4. Scatter plot showing the relationship between accident frequency rate and labour productivity.

4.5 PPE Compliance, Skilled Labour Ratio, and Labour Productivity

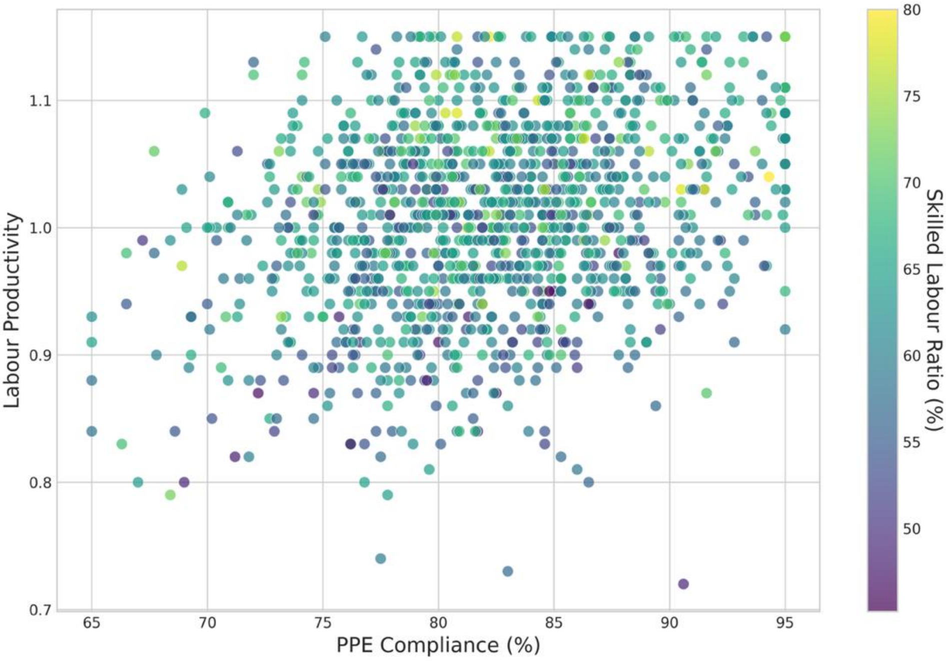


Fig 5. Scatter plot showing the relationship between PPE compliance and labour productivity

Figure 5 shows how compliance with PPE correlates with labour productivity based on colour intensity for the skilled labour ratio. There seems to be a positive correlation, i.e., higher percentages of compliance with PPE corresponded with higher levels of productivity. Values of productivity seemed to have clustered in the top half of the scale when PPE compliance was greater than 80%. Therefore, it appears that well-structured adherence to safety principles contributes to stable operations. While the overall trend upwards was somewhat scattered, the findings from the regression analysis also suggest that compliance with PPE is a highly positive predictor of labour productivity.

In addition to the main effects observed, an interaction effect can also be identified using the colour gradient. In particular, observations (i.e., projects) with higher skilled labour ratios (warming colours) were found to be located at higher productivity levels. Thus, it appears that workforce competence enhances the impact of PPE compliance. Those projects having both a high level of compliance and a large number of skilled workers had fewer low-productivity outliers and tended to perform more consistently. The pattern therefore supports the conclusion that a combination of workplace safety discipline and workforce capability are joint stabilizer within systems of productivity in the construction industry.

4.6 Relationship Between Safety Training Hours and Labour Productivity

Fig 6. Scatter plot showing the relationship between safety training hours per worker and labour productivity

Figure 6 depicts a scatter plot that demonstrates the relationship between labor productivity and the number of safety training hours each worker receives. It can be seen from this plot that an increase in hours spent on training corresponds to an increase in productivity. Lower levels of training (especially those below 1-2 hours) exhibit a greater amount of variance; they are more densely populated by data points whose productivity is less than the average or "baseline" productivity. Training that exceeds 3 hours/worker results in data points that cluster much closer to and typically above the "average" or "baseline", indicating better operational consistency. This upward trend indicates that safety training serves as both a method for ensuring compliance as well as improving efficiency. Improved awareness of hazards, task coordination, and compliance with procedures likely result from increased training hours. Reduced workflow disruptions and error-based inefficiencies result when fewer errors occur due to improved procedural compliance. While there will always be variability present in the data observed, it clearly demonstrates the regression finding that the amount of safety training hours workers receive is positively correlated to their labor productivity within the multi-city construction dataset.

5 Conclusion

The study evaluated the correlation between the performance of occupational safety and labor productivity in construction projects, using a dataset containing 1,275 observations from 6 different cities. Statistical data mining was used along with predictive modeling to assess whether various factors (safety compliance, disruption indicators, and characteristics of workers) affect productivity. Findings indicate that occupational safety is an operational performance variable and is not solely a regulatory variable. All of the predictive models utilized (QR, RLS, GEE, and WLS) performed consistently well with R^2 values ranging from approximately 0.72 to 0.74, indicating that the majority of variability in productivity can be explained by variables related to safety.

The empirical data indicated a strong positive correlation between safety compliance and labor productivity. Improved compliance rates for PPE and additional safety training hours were positively correlated with both productivity and the consistency of output on construction projects. The disruption-related variables (accident frequency, near-miss incidents, work stoppage hours, overtime exposure, and absenteeism) negatively correlated with productivity. These variables create workflow disruptions, fatigue, and delay coordination that result in less effective labor output. The ability of the workforce was identified as a significant positive reinforcing factor; i.e., a higher ratio of skilled labor resulted in a more consistent level of productivity.

The SPCI, which converts regression results into a useful decision support tool. The SPCI combines variables representing positive safety compliance with disruption indicators to provide a measure of the total impact of safety on productivity for each project. This provides project managers with a means to determine when operational conditions are contributing to reduced productivity stability and allows them to identify those safety interventions that will improve their production while improving safety. The findings of this research also refute the commonly held belief that safety improvements lead to decreased productivity and show that structured safety programs can maintain continuity of the workflow and minimize losses due to disruption.

References

1. J. Fiegler-Rudol, K. Lau, A. Mroczek, and J. Kasperczyk, "Exploring Human–AI Dynamics in Enhancing Workplace Health and Safety: A Narrative Review," (2025). https://doi.org/10.3390/ijerph22020199
2. Qu, J., Liu, L., Zeng, J., Maraseni, T. N., & Zhang, Z.. City-Level Determinants of Household CO2 Emissions per Person: An Empirical Study Based on a Large Survey in China. *Land*, (2025), *11*(6). https://doi.org/10.3390/land11060925
3. J. M. Nwaogu, A. P. C. Chan, J. A. Naslund, and S. Anwer, "The Interplay Between Sleep and Safety Outcomes in the Workplace: A Scoping Review and Bibliographic Analysis of the Literature," (2025). doi: 10.3390/ijerph22040533.

4. Y. Gu, C. Wang, Y. Liu, and R. Zhou, "An ontology-based multi-hazard coupling accidents simulation and deduction system for underground utility tunnel - A case study of earthquake-induced disaster chain," Reliab. Eng. Syst. Saf.253,110559, (2025). doi: https://doi.org/10.1016/j.ress.2024.110559.
5. R. Lohitashwa, N. Kadli, R. Kisan, S. A, and D. Deshpande, "Effect of stress on sleep quality in young adult medical students: a cross sectional study," Int. J. Res. Med. Sci.**3**, 3519–3523, (2017). doi: 10.18203/2320-6012.ijrms20151391.
6. H. Sarvari, D. J. Edwards, I. Rillie, and C. Roberts, "Barriers, Bottlenecks, and Challenges in Implementing Safety I- and Safety II-Enabled Safe Systems of Working in Construction Projects: A Scoping Review," (2025). doi: 10.3390/buildings15030347.
7. S. Sathvik, A. Alsharef, A. K. Singh, M. A. Shah, and G. ShivaKumar, "Enhancing construction safety: predicting worker sleep deprivation using machine learning algorithms," Sci. Rep.14,15716, (2024). doi: 10.1038/s41598-024-65568-2.
8. S. Vidović et al., "Sleep Quality and Mental Health Among Medical Students: A Cross-Sectional Study," (2025). doi: 10.3390/jcm14072274.
9. S. S. Chandra, K. Loganathan, B. O. Awuzie, and F. Wang, "A Longitudinal Study Examining the Association between Cognitive Behavior and Rational Abilities and the Effect of Sleep Quality on Construction Laborers," (2023). doi: 10.3390/su15076257.
10. Y Huang, X Zhu, R Wang,Y Xie, & S Fong. A Dynamic Global–Local Spatiotemporal Graph Framework for Multi-City PM2.5 Long-Term Forecasting. *Remote Sensing*, *17*(16) (2025). https://doi.org/10.3390/rs17162750
11. S Jia, W Li, & X Guo. Multisource data-driven resilience assessment and optimization of metro station public spaces: a case study of Suzhou metro. *Journal of Asian Architecture and Building Engineering*. (2025) https://doi.org/10.1080/13467581.2025.2608446
12. M Turner, H Lingard, & P Pirzadeh. How Does the Five-Day Work Week Impact Labour Productivity? Exploring the Perceptions of Australian Construction Workers. *Buildings*, *15*(5) (2025). https://doi.org/10.3390/buildings15050787
13. H. Ikeda, T. Kubo, S. Izawa, N. Nakamura-Taira, T. Yoshikawa, and R. Akamatsu, "The Joint Association of Daily Rest Periods and Sleep Duration with Worker Health and Productivity: A Cross-Sectional Web Survey of Japanese Daytime Workers," (2022). doi: 10.3390/ijerph191711143.
14. A. Baghdadi, "An Evaluation Study of the Impact of Occupational Health and Safety on Productivity in Saudi Arabia Construction Industry," Am. J. Civ. Eng. Archit.12,36–**43**, (2024).doi: 10.12691/ajcea-12-2-3.