# Introduction & Literature Review

The success or failure of technological solutions within a professional organization or public institution is dependent on many factors going into the project itself. Variables as diverse as process, governance, financial investment, and executive sponsorship have direct influence. An often-overlooked factor, however, is the readiness of human resources beyond the simple communication of upcoming changes.

Employee training, though often included as part of the scope of overall management practices, has not been examined in depth in terms of its direct influence on the success of an organization as measured by that organization’s performance or the measured success of individual projects. As an example, Jakob Shneebacher provides a general examination of management practices throughout the United Kingdom (Schneebacher, 2021). In his analysis, Schneebacher cites a strong and significant relationship between management practices and measurements of business success. Employee training, however, is included as a part of a group of measures that include promotion and employee underperformance and is not called out separately.

Similarly, George Ochola (Ochola, 2018) examines employee motivation as a larger subject that includes employee training. Ochola does emphasize organizational performance in relation to motivation, stating that "the capability of drawing, holding and advancing employees that are talented are the main characteristics of a business that is successful" (Ochola, 2018). Training and skills development specifically are identified as key factors that ensure an employee feels empowered and is more likely to embrace organizational goals. Ochola does not engage in experimental analysis, though, and concentrates on a review of existing literature.

Additional analysis by Jalal Hanaysha (Hanaysha, 2016) has demonstrated strong relationships between employee training and employee commitment. Commitment is roughly defined as whether an employee identifies themselves with the organization and aligns to the organization's goals and values. More directly, it is the willingness of the individual to remain associated with the organization (i.e. remain employed) and their commitment to their work. Employee commitment goes beyond merely doing the job and manifests as a desire to be productive within that job. Hanaysha examines employee empowerment and teamwork as well as employee training and concluded that all three have a significant positive effect on employee commitment with training demonstrating the highest t-value of the three.

Similar to Hanaysha’s examination of employee commitment is Debra Truitt’s analysis of employee training (Truitt, 2011). Truitt does a thorough analysis of employee training and coaching, especially the impact of updated training programs. She concludes there is a direct relationship between an employee’s experience of training and their attitudes related to their own proficiency in their jobs as well as their attitude about their jobs. Although improved organizational performance and project performance might be inferred from this improved attitude, Truitt does not make that association.

R.A. Khan et. al. (Khan, Khan, & Khan, 2011) provide the most direct association, showing that not only does training improve employee performance, but that improved performance leads to improved organizational performance. As with Ochola and Schneebacher, Khan et. al. state that employee performance is dependent on many factors including job satisfaction and management practices. Their analysis, though, concentrates on employee training, examining three factors of training: on-the-job (vs. through books or online), training design, and the delivery style of the trainer. The authors conclude all three variables have a strong positive influence on organizational performance.

Of note across much of the literature examined is the absence of experimental links between training employees on a skill set and the direct application of that skill set toward the success of an organization's projects or overall performance. Ochola's literature review (Ochola, 2018), though citing several other analyses, is noticeably lacking any experimentation. Hanaysha (Hanaysha, 2016) indicates an association between employee commitment and organizational performance but stops short at examining that commitment. Further, Hanaysha focused on the public universities in northern Malaysia, both limiting the geographical scope of the study as well as possibly biasing the outcomes based on individuals already committed to education.

Although the analysis conducted by Khan et. al. does include experimentation, they admit in their introduction that the results are "strongly based on the literature review" (Khan, Khan, & Khan, 2011). The actual experiment consists of analyzing 79 surveys with 15 questions, indicating the employees themselves have assessed the value of the training they have received with limited objective evidence of improved performance.

The analysis in this document attempts to provide a more empirical demonstration of the direct relationship between training individuals on new technology and the success of those technological solutions. More specifically, through the analysis of survey data collected by the World Bank Group, I will examine the hypothesis that training has a positive impact on organizational success and on the outcome of specific projects.

## Hypotheses

H0A [NULL]: The introduction of internal employee training programs on technology and data have no impact on the technological maturity of an organization.

H1A: Introducing internal employee training programs on technology and data improve the overall technological maturity of an organization.

H0B [NULL]: The introduction of internal employee training programs on technology and data have no impact on the success rate of specific technology projects implemented within an organization.

H1B: Introducing internal employee training programs on technology and data improve the success rate of specific technology projects implemented within those organizations.

# Data & Methodology

For this project, I will examine the success rate of implementing technological solutions within world government structures in relation to the inclusion of employee training on technology. Success will be measured based on two indicators:

* The GovTech Maturity Index (GTMI) score assigned by the World Bank Group (WBG), indicating the maturity of a country's digital government transformation.
* Outcome ratings (Satisfactory/Unsatisfactory) of WBG funded projects by country.

All data has been obtained from the WBG either through their annual GovTech Maturity Index (GTMI) Update or the Digital Governance Projects Database. The GTMI is a report on the state of Technology services and solutions as implemented by central governments throughout the world. The GTMI assessment includes 198 world economies obtained through surveys and remote data collected from non-participating countries.

The GTMI was launched in 2020, and all data is published by the WBG in the GovTech Dataset ([GovTech Dataset | Data Catalog (worldbank.org)](https://datacatalog.worldbank.org/search/dataset/0037889/GovTech-Dataset)). However, due to restrictions in accessing this dataset through the API, I chose to make use of the published data in the Excel files downloaded from the site. GovTech data published in October 2022 includes results for the dataset published in December 2020. I did not include the 2020 data due to specific indicators/metrics that were not implemented that year. More specifically, the metrics used for this analysis were not in the 2020 dataset.

Additional data used for this analysis came from the Digital Governance Projects Database ([Digital Governance Projects Database | Data Catalog (worldbank.org)](https://datacatalog.worldbank.org/search/dataset/0038056/digital-governance-projects-database)). This provides details of 1,449 projects funded by the WBG in 147 countries, including cost, duration, and outcome ratings of completed activities. For this analysis, I focused on the success rating of projects within countries also evaluated in the GovTech Dataset. Again, data from the October 2022 data set was used through a download of the available Excel file at the site. Additionally, country lookup data was established through code to account for mismatched country names between the two data sets.

## Data Tables

The following tables show the variables extracted from each file. Although each file includes an extensive set of variables, only the variables listed in these tables will be considered for the purpose of this project. Note that several variables were renamed from the original source for ease of reference within this project. In the following tables, the original name from the data source is provided (Original Name), along with the name to be used in this project (Project Name).

### Data Source: WBG\_GovTech Dataset\_Oct2022.xlsx (GovTech)

<https://datacatalog.worldbank.org/search/dataset/0037889/GovTech-Dataset>

|  |  |  |  |
| --- | --- | --- | --- |
| Original Name | Project Name | Description | Values |
| Code | Code | Standard WBG Abbreviation for a country. | Text (3 characters) |
| Economy | Country | The name of the country | Text |
| Population | Population | Population of the country in thousands | Numeric |
| Level | IncomeLevel | Income level (abbreviation) | H (High) / UM (Upper middle) / LM (Lower middle) / L (Low) |
| Reg | Region | Mapping to the World Bank regions (abbreviation) | AFR / EAP / ECA / LCR / MNA / SAR |
| Grp | Group | GovTech Maturity Index group: | A: Very high (>= 0.75) B: High (>= 0.50 and < 0.75) C: Medium (>= 0.25 and < 0.50) D: Low (< 0.25) |
| GTMI | GTMI | GTMI score | 0 to 1 |
| CGSI | CGSI | Core Government Systems Index score | 0 to 1 |
| PSDI | PSDI | Public Sector Delivery Index score | 0 to 1 |
| DCEI | DCEI | Digital Citizen Engagement Index score | 0 to 1 |
| GTEI | GTEI | GovTech Enablers Index score | 0 to 1 |
| I-45 | DS\_Strategy\_Program | Is there a government strategy/  program to improve digital skills in the public sector? | 0 = No, 1 = Yes (Only strategy or program), 2 = Yes (Both strategy and program) |
| I-45.4 | FocusArea | Focus areas of the DS strategy | 0 = Unknown, 1 = Basic digital skills, 2 = Basic digital skills + Data literacy, 3 = Advanced digital skills + Data literacy |
| I-45.5 | DSProgram | Is there a DS program? | 0 = No, 1 = Yes |
| I-45.5.1 | DSProgramType | If Yes > Type of primary DS  program(s) | 1 = Academic program, 2 = Public sector program, 3 = CSO/Private program |
| I-45.5.3 | DSProgramMandatory | If Yes > DS program mandatory for  new public employees? | 0 = Unknown, 1 = Not mandatory, 2 = Mandatory |
| I-45.6 | DSProgramExternal | Are there digital skills programs offered by governments for  citizens/schools? | 0 = No, 1 = Yes (fee-based programs), 2 = Yes (freely available programs) |
| I-45.7 | DSProgramPublished | Publishing of the results/progress in DS programs? | 0 = No, 1 = Yes (internal, not published), 2 = Yes (public, published) |

### Data Source: WBG\_DG-GovTech\_Projects\_Oct2022.xlsx (Projects)

<https://datacatalog.worldbank.org/search/dataset/0038056/digital-governance-projects-database>

|  |  |  |  |
| --- | --- | --- | --- |
| Original Name | Project Name | Description | Values |
| Project ID | ProjectID | Project Identifier | Text |
| Region | Region | Mapping to the World Bank regions (abbreviation) | AFR / EAP / ECA / LCR / MNA / SAR |
| Country | Country | The name of the country where the project occurred | Text |
| ICROut | ICROutcome | Implementation Completion Report (ICR) Project Outcome rating – based on assessment completed within 6 months after closure of project by WBG project team | Standard ICR ratings : HS / S / MS / MU / U / HU |
| IEGOut | IEGOutcome | Independent Evaluation Group (IEG) Project Outcome rating = based on a review within 6 months after delivery by independent review board for validation. | Standard ICR ratings : HS / S / MS / MU / U / HU |

### Data Source: API\_NY.GDP.MKTP.CD\_DS2\_en\_csv\_v2\_4770391.csv (GDP)

<https://data.worldbank.org/indicator/NY.GDP.MKTP.CD>

|  |  |  |  |
| --- | --- | --- | --- |
| Original Name | Project Name | Description | Values |
| Country Name | Country | The name of the country | Text |
| Country Code | Code | Standard WBG Abbreviation for a country. | Text (3 characters) |
| 2021 | GDP2021 | The country’s Gross Domestic Product (GDP) for 2021 in US dollars. | Numeric |

### Summary Statistics

The following tables include summary statistics for some of the key variables I will be using in this project. I have separated categorical and non-numeric variables from numeric.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Count | Mean | Median | Std | Min | Max |
| GTMI | 198 | 0.552226 | 0.572181 | 0.261910 | 0.019134 | 0.991409 |
| CGSI | 198 | 0.574616 | 0.581023 | 0.237276 | 0.002509 | 0.990160 |
| PSDI | 198 | 0.649481 | 0.738621 | 0.284674 | 0.000000 | 1.000000 |
| DCEI | 198 | 0.448667 | 0.394079 | 0.314998 | 0.002499 | 0.997501 |
| GTEI | 198 | 0.536139 | 0.568245 | 0.276343 | 0.036503 | 0.983757 |
| GDP | 245 | $3,275,719M | $67,404M | $10,795,732M | $63M | $96,513,077 |

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | Variable | Count | Comment |
| GovTech | Code | 198 | 198 countries were included in the original GovTech assessment. |
| GovTech | DS\_Strategy\_Program | 0: 65  1: 67  2: 66 | 0 = No Strategy or Program  1 = Yes (Only strategy or program)  2 = Yes (Both strategy and program) |
| GovTech | FocusArea | 0: 69  1: 31  2: 32  3: 66 | 0 = Unknown focus for the strategy/program  1 = Basic digital skills  2 = Basic digital skills + Data literacy  3 = Advanced digital skills + Data literacy |
| GovTech | DSProgramType | 0: 100  1: 7  2: 76  3: 15 | 0 = None or unknown  1 = Academic program  2 = Public sector program  3 = CSO/Private program |
| Projects | ProjectID | 1024 | 1024 projects subsidized by WBG with at least one project outcome rating |
| Projects | ICROutcome | |  |  | | --- | --- | | HS: | 33 | | S: | 485 | | MS: | 316 | | MU: | 104 | | U: | 64 | | HU: | 5 | | |  |  | | --- | --- | | HS: | Highly Satisfactory | | S: | Satisfactory | | MS: | Moderately Satisfactory | | MU: | Moderately Unsatisfactory | | U: | Unsatisfactory | | HU: | Highly Unsatisfactory | |  |  | |
| Projects | IEGOutcome | |  |  | | --- | --- | | HS: | 25 | | S: | 295 | | MS: | 372 | | MU: | 158 | | U: | 80 | | HU: | 12 | | |  |  | | --- | --- | | HS: | Highly Satisfactory | | S: | Satisfactory | | MS: | Moderately Satisfactory | | MU: | Moderately Unsatisfactory | | U: | Unsatisfactory | | HU: | Highly Unsatisfactory | |  |  | |

## Analysis Methodology

This analysis will use both Linear (OLS) Regression and Logistic (Logit) Regression to analyze the influence of the inclusion of a Digital Skills strategy or program within each government, including the level of skills training provided, and the type. Additional independent variables are included to reduce omitted variable bias (OVB) including population and GDP.

Specifically, my dependent variables will be: GTMI, ICROutcome, and IEGOutcome. My primary independent variables will include:

|  |
| --- |
| GTEI |
| DS\_Strategy\_Program |
| FocusArea |
| DSProgram |
| DSProgramType |
| DSProgramMandatory |
| DSProgramExternal |
| DSProgramPublished |

I introduce three independent variables to reduce omitted variable bias (OVB) including population, income levels, and GDP. Although these variables can not eliminate OVB, my goal is to reduce bias to a level that can be used to assess the significance of the independent variables of interest.

Because the **ICROutcome** and **IEGOutcome** variables are categorical rather than continuous, these variables are converted into simple bivariate variables indicating Successful (1) or Unsuccessful (0) for a project. With this conversion, we can leverage Logit regression to understand the significance of Skills Training variables on project success.

## Controlling for Omitted Variable Bias

With the large number of countries considered in this analysis, there is substantial risk of Omitted Variable Bias. I have attempted to control for OVB through the introduction of three variables not directly related to the variables of interest. Each of these, however, can vary across the countries represented. They include Income Level, a categorical variable, Population, and GDP.

In this section, I am examining how each of these variables is distributed across the GTEI scores to assess whether it can be introduced as a control variable. I start with Income Level.

A picture containing screenshot, text, diagram, line

Description automatically generated

Unsurprisingly, the GTEI data appears strongly correlated with the GTMI score, since GTEI is one of 4 composite scores that are used to generate GTMI. In this plot, however, the interesting aspect is that Income Level is generally scattered across the values. This indicates that the average Income Level may NOT be an influencing factor for Tech Maturity.

Population and GDP are the other two variables we are interested in examining for OVB. First, let's look at the distribution of the two data sets.

A picture containing screenshot, text, diagram, line

Description automatically generated

Both distributions are highly skewed. The outliers may affect the outcome of the regression analysis. To account for this, I convert to logarithmic values and reexamine the distribution.

A picture containing diagram, plot, line, screenshot

Description automatically generated

The distributions now appear closer to normal. Additionally, I examine how these values are distributed across the GTMI and GTEI scores in a scatterplot similar to the one used for Income Level.

A picture containing screenshot, text

Description automatically generated

A picture containing screenshot, text, diagram

Description automatically generated

The scatter plots show a slightly increased concentration of higher population and higher GDP in the upper right. However, the concentrations do not appear significant through visual observation. In general, the values are spread throughout the GTMI and GTEI results. I concluded that all three variables could be added as control variables to the regression analyses.

Note that as a categorical variable, Income Level was converted to dummy variables, and the Low (L) dummy was dropped to avoid multicollinearity.

# Analysis

This assessment of the influence of TechSkills training includes two separate analyses:

* Analysis of Government Tech Maturity through the WBG GTMI score.
* Analysis of Government Tech Project Success through satisfaction outcome scores.

Though both analyses leverage the same Tech Enablement variables to examine influence, the following sections explore these analyses separately.

## Analysis of GTMI Score

The goal in this first analysis is to assess to what extent the Tech Skills enablement variables influence the overall Tech maturity of a government organization. The initial hypothesis, as stated earlier, is:

H0A [NULL]: The introduction of internal employee training programs on technology and data have no impact on the technological maturity of an organization.

H1A: Introducing internal employee training programs on technology and data improve the overall technological maturity of an organization.

The Tech Skills variables are components of an overall GTEI (Technical Enablement) variable used to generate the GTMI (Tech Maturity) variable. To understand the relationship of these individual scores to the GTEI and GTMI scores, we can use a correlation matrix. Note that due to the number of variables considered, this correlation analysis is split into two tables.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **GTMI** | **GTEI** | **DS\_Strategy\_Program** | **FocusArea** | **DSProgram** | **DSProgramType** |
| **GTMI** | 1.0000 | 0.9552 | 0.6869 | 0.6427 | 0.6890 | 0.7167 |
| **GTEI** | 0.9552 | 1.0000 | 0.7537 | 0.6772 | 0.7355 | 0.7426 |
| **DS\_Strategy\_Program** | 0.6869 | 0.7537 | 1.0000 | 0.7358 | 0.7637 | 0.7281 |
| **FocusArea** | 0.6427 | 0.6772 | 0.7358 | 1.0000 | 0.6879 | 0.6563 |
| **DSProgram** | 0.6890 | 0.7355 | 0.7637 | 0.6879 | 1.0000 | 0.9163 |
| **DSProgramType** | 0.7167 | 0.7426 | 0.7281 | 0.6563 | 0.9163 | 1.0000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **GTMI** | **GTEI** | **DSProgramMandatory** | **DSProgramExternal** | **DSProgramPublished** |
| **GTMI** | 1.0000 | 0.9552 | 0.5978 | 0.6562 | 0.4236 |
| **GTEI** | 0.9552 | 1.0000 | 0.6422 | 0.6986 | 0.4732 |
| **DSProgramMandatory** | 0.5978 | 0.6422 | 1.0000 | 0.6802 | 0.4577 |
| **DSProgramExternal** | 0.6562 | 0.6986 | 0.6802 | 1.0000 | 0.5172 |
| **DSProgramPublished** | 0.4236 | 0.4732 | 0.4577 | 0.5172 | 1.0000 |

Based on the above correlation matrices, it appears the variables with the least correlation with GTMI and GTEI are DSProgramMandatory (indicating whether new employees are required to take the Tech Skills training), and DSProgramPublished (indicating whether the results of the training program are published). We can keep this in mind as we turn to Regression analysis.

Turning to a regression analysis, I started with a simple regression of GTEI on GTMI.

GTMI = β0 + (GTEI x β1)

|  |  |
| --- | --- |
| Intercept (β0) | 0.0668 (0.012) |
| GTEI (β1) | 0.9053 (0.020) |
| R-Value | 0.9552 |
| R-Squared | 0.9124 |
| P-Value | 0.0000 |

In this simple regression model, the GTEI enablement score has a coefficient of 0.9053 (0.02), indicating that a value of 1 for the GTEI score will increase the GTMI score by 0.9053. However, we have to keep in mind that GTEI and GTMI scores have a range of 0 to 1. Therefore, we could interpret this as an increased value of 0.1 in the GTEI score (enablement) improves the overall Tech Maturity of the GTMI score by 0.09, or almost 0.1, indicating a very near 1 to 1 correlation.

We can illustrate this with our earlier scatterplot, this time including the regression line.

A picture containing text, screenshot, line, diagram

Description automatically generated

My next question was whether the coefficient on GTEI changes significantly when we include our control variables: Income level, Population, and GDP. This requires a more complex linear regression. Note that the Income Level, as stated before, has been converted into three dummy variables: IL\_LM, IL\_UM, and IL\_H.

GTMI = β0 + (GTEI x β1) + (logPopulation x β2) + (logGDP x β3) +   
(IL\_LM x β4) + (IL\_UM x β5) + (IL\_H x β6)

The results of this regression appear in the table below. An asterisk (\*) indicates significance at the 95% level. Standard errors are expressed in parentheses.

|  |  |
| --- | --- |
|  | Dependent variable: GTMI |
| Intercept | 0.0033 (0.153) |
| GTEI | 0.8576\* (0.029) |
| logPopulation | 0.0068 (0.012) |
| logGDP | -0.0002 (0.011) |
| IL\_LM | 0.0201 (0.023) |
| IL\_UM | 0.0648\* (0.033) |
| IL\_H | 0.0504 (0.048) |
|  |  |
| R-Squared | 0.924 |
| Adj. R-Squared | 0.922 |
| Observations | 184 |

In the output from the above, we see a slightly smaller coefficient for the GTEI score, 0.8576, compared to 0.9053 on the simple regression. This score remains significant with a t-score of 29.8, and an extremely small p-value (5.642e-71). Again, this is not a surprise, as the GTEI score is one of 4 used to generate the GTMI score.

The GTEI score itself is comprised of 56 separate numerical scores, some weighted. Of these, 23 scores fall to the category of *Human Capital*, including the 7 scores we are most interested in, relating to Technical Skills enablement.

With a  baseline regression provided above, we can now examine what influence these individual scores have on a government organization's overall Tech Maturity (GTMI). Note that these Technical Skills variables all have limited values, some binary (0 or 1), and some with only three values (0,1,2 or 1,2,3). For this reason, they can be considered categorical to some extent. As each score is ordered, i.e. the higher the number the higher the level of skills enablement, they can still be used for the regression analysis.

In this analysis, I am performing two regressions:

1. A regression on GTMI for all Tech Skills enablement variables.

2. A regression on GTEI, the Tech Enablement specific aspect of GTMI.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Simple Regression | Simple Regression w/ Control | Complete Regression | Complete Regression |
| Dependent variable | GTMI | GTMI | GTMI | GTEI |
| Intercept | 0.0668 (0.012) | 0.0033 (0.153) | -0.3613 (0.285) | -0.3848 (0.276) |
| GTEI | 0.9053\* (0.020) | 0.8576\* (0.029) |  |  |
| DS\_Strategy\_Program |  |  | 0.0359 (0.022) | 0.0712\* (0.022) |
| FocusArea |  |  | 0.0267\* (0.012) | 0.0203 (0.012) |
| DSProgram |  |  | -0.0443 (0.058) | -0.0351 (0.056) |
| DSProgramType |  |  | 0.0634\* (0.024) | 0.0544\* (0.023) |
| DSProgramMandatory |  |  | 0.0233 (0.023) | 0.0262 (0.023) |
| DSProgramExternal |  |  | 0.0300 (0.018) | 0.0325 (0.017) |
| DSProgramPublished |  |  | 0.0215 (0.015) | 0.0367\* (0.014) |
| logPopulation |  | 0.0068 (0.012) | 0.0195 (0.022) | 0.0162 (0.021) |
| logGDP |  | -0.0002 (0.011) | 0.0185 (0.021) | 0.0194 (0.020) |
| IL\_LM |  | 0.0201 (0.023) | 0.0658 (0.043) | 0.0501 (0.042) |
| IL\_UM |  | 0.0648\* (0.033) | 0.1732\* (0.061) | 0.1281\* (0.059) |
| IL\_H |  | 0.0504 (0.048) | 0.2180\* (0.090) | 0.2059\* (0.087) |
|  |  |  |  |  |
| R-Squared | 0.9124 | 0.924 | 0.756 | 0.797 |
| Adj. R-Squared |  | 0.922 | 0.739 | 0.783 |
| Observations | 198 | 184 | 184 | 184 |

The regression outcomes of the complete regression models with all independent variables indicate significant results (at the 95% Confidence Level) for FocusArea, with respect to GTMI, and DSProgramType, for both GTMI and GTEI.

* *FocusArea* is an indication of the type of Tech Skills provided: 0=Unknown, 1=Basic digital skills, 2=Basic digital skills + Data literacy, 3=Advanced digital skills + data literacy.
* *DSProgramType* is an indication of the type of learning program provided: 1=Academic program, 2=Public sector program, 3=CSO/Private program.

However, the coefficient on both variables is small: 0.0267 (GTMI) and 0.0203 (GTEI) for FocusArea, and 0.0634 (GTMI) and 0.0544 (GTEI) for DSProgramType. This indicates that these variables have a definite influence on the GTMI and GTEI scores, but that it is only a small influence.

In the GTEI regression, *DS\_Strategy\_Program* and *DSProgramPublished* also appear significant at the 95% Confidence Level. As with *FocusArea* and *DSProgramType*, though, these variables also have small coefficients. Of the four, *DS\_Strategy\_Program* has the highest coefficient at 0.0712. Possible values for DS\_Strategy\_Program are: 0=No Strategy or Program for digital skills exists, 1=Yes, but only one or the other (Strategy or Program) exists, and 2=Both a Strategy and a Program for digital skills enablement exists.

## Analysis of Project Success Rates

Project Success is recorded by one of two project evaluation variables: ICR Outcome and IEG Outcome. An Implementation Completion Report (ICR) is an evaluation done at the completion of the project and represents an assessment of the project itself. An Independent Evaluation Group (IEG) evaluation is a review of the ICR and associated materials completed within six months of the project end to provide a secondary independent evaluation of the success of the project. Refer to the Data & Methodology section for possible scores.

H0B [NULL]: The introduction of internal employee training programs on technology and data have no impact on the success rate of specific technology projects implemented within an organization.

H1B: Introducing internal employee training programs on technology and data improve the success rate of specific technology projects implemented within those organizations.

Starting with an assessment of the distribution of the scores, we can quickly see that ICR Scores tend toward more satisfactory results than the follow-up IEG scores.

A picture containing text, screenshot, diagram, plot

Description automatically generated A picture containing text, screenshot, diagram, plot

Description automatically generated

However, it should also be noted that the data includes far fewer IEG scores (692) than ICR scores (834). For the purpose of our analysis, we have converted any satisfactory score to a value of 1, and any unsatisfactory score to a value of 0. This will allow us to leverage logistic regression through a Logit model.

For this analysis, I used both a Generalized Linear Model (GLM) from the *statsmodels* library as well as a Logit model from the *SciKit-Learn* library. The first two columns in the table below show the outcome of the GLM mode on each of the binary variables. The second two columns are the result of the Logit model. In both cases, all independent variables of interest, including control variables, are used. Exceptions are *DSProgramPublished* and *logGDP* do not appear in the Logit models due to the presence of null values.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | GLM | GLM | Logit | Logit |
| Dependent variable | ICROutcomeB | IEGOutcomeB | ICROutcomeB | IEGOutcomeB |
| Intercept | 0.2287 (0.467) | 0.1975 (0.559) | 1.0144 | 0.1350 |
|  |  |  |  |  |
| DS\_Strategy\_Program | 0.0058 (0.030) | 0.0528 (0.035) | 0.0940 (0.1963) | 0.2531 (0.196) |
| FocusArea | -0.0089 (0.017) | -0.0251 (0.020) | -0.0282 (0.110) | -0.1077 (0.110) |
| DSProgram | 0.2417 (0.099) | 0.1350 (0.118) | 1.6431 (0.700) | 0.6169 (0.700) |
| DSProgramType | -0.0687 (0.041) | -0.0134 (0.049) | -0.4490 (0.273) | -0.0480 (0.273) |
| DSProgramMandatory | -0.0343 (0.031) | -0.0252 (0.037) | -0.2367 (0.215) | -0.0959 (0.215) |
| DSProgramExternal | -0.0325 (0.026) | -0.0488 (0.031) | -0.2098 (0.183) | -0.2169 (0.183) |
| DSProgramPublished | 0.0113 (0.021) | 0.0182 (0.025) |  |  |
| logPopulation | -0.0400 (0.037) | -0.0271 (0.044) | 0.0000 (0.020) | -0.0062 (0.020) |
| logGDP | 0.0398 (0.035) | 0.0255 (0.042) |  |  |
| IL\_LM | -0.0030 (0.059) | 0.0590 (0.070) | 0.3167 (0.225) | 0.3902 (0.225) |
| IL\_UM | 0.0017 (0.096) | 0.1568 (0.114) | 0.6891 (0.236) | 1.0031 (0.236) |
| IL\_H | -0.0300 (0.130) | 0.1639 (0.155) | 0.6504 (0.416) | 1.1089 (0.416) |
|  |  |  |  |  |
| Pseudo R-Squared | 0.01617 | 0.03854 | n/a | n/a |
| Observations | 964 | 964 |  |  |

With the logistic models, no coefficients show significance at the 95% level, indicating no influence on project success. Of note in these models is the large number of negative coefficients, indicating any influence would reduce the probability of success.

More importantly for these Logistic models is the prediction of success based on these models. Starting with the GLM models, and using the original data to develop the model, I compare the actual percentage of successful projects to the predicted percentage of successful projects.

|  |  |  |
| --- | --- | --- |
|  | ICR Outcome | IEG Outcome |
| Actual Success Rate | 0.8138 | 0.6787 |
| Predicted Success Rate | 1.0000 | 0.9803 |
| Difference | **0.1862** | **0.3016** |

With differences of 30% and nearly 20%, the GLM models do not appear to do a good job of predicting project success rate. Although, this may require more refinement of the model itself, these results imply no significant relationship between the Tech Enablement variables, and the Project Success rate.

The Logit models from **Sci-Kit Learn** provide more complete classification reports. When we focus on the accuracy, though, these models show identical results to the GLM models from **statsmodels**.

**Partial Classification Report: ICR Outcome from Logit Model**

|  |  |  |  |
| --- | --- | --- | --- |
| Value | Precision | Recall | # actual values |
| 0 | 0.00 | 0.00 | 186 |
| 1 | 0.81 | 1.00 | 813 |

**Accuracy of ICR Outcome Logit Model: 0.81**

**Partial Classification Report: IEG Outcome from Logit Model**

|  |  |  |  |
| --- | --- | --- | --- |
| Value | Precision | Recall | # actual values |
| 0 | 0.38 | 0.01 | 321 |
| 1 | 0.68 | 0.99 | 678 |

**Accuracy of IEG Outcome Logit Model: 0.68**

* Precision = accuracy of a predicted positive outcome – when we predict a value of 1, what percentage were actually 1?
* Recall = ability of model to predict a positive outcome – if the value is 1, at what rate did we predict it is 1?

# Discussion and Conclusion

Contrary to my expectations, there is no significant evidence of the enablement of Tech Skills strategies or programs improving the outcome of either the general Tech Maturity of an organization, or the likelihood of the success of a Technology project within an organization. The influence of the Tech Enablement variables on the GTMI score for countries in the WBG report showed significance for only 2 variables, and the coefficient for those was small enough that it would not shift the GTMI score at a level that substantially mattered for the WBG report. Similarly, the GTEI enablement score was not influenced substantially by the Tech Skills variables, especially surprising since they are part of the composite score that makes up the GTEI.

For Project Success, the Logistic models did not perform well for predicting success rates for either of the success metrics. Since both the GLM models and the Logit models returned the same results, either the model itself is incorrect, or the Tech Skills variables are not good predictors of project success. As a follow up, I would find it useful to work with the model itself more, adjusting some of the independent variables used, and perhaps removing those with the least significance and/or lowest coefficients.

Some of our highest coefficients were on the Income Level control variables, with both Upper-Middle and High income dummy variables showing significant influence on both the GTMI and GTEI scores. Although not the focus of this analysis, a follow up investigation of Income Level on Tech Maturity may reveal more interesting results.

Additionally, I would like to evaluate regression outcomes across the groups identified by the WBG report. Based on the GTMI score, WBG identifies a grouping of the country in terms of its GovTech maturity. Theoretically, countries scoring higher in the GTMI would be more likely to 1) have implemented a Digital Skills program and 2) have a higher rate of success on projects.

Based on the analysis in section 3.1, a value of 2 for the DS\_Strategy\_Program variable, indicating the presence of both a Digital Skills Strategy and an internal Program, the GTEI score is increased by a score of 0.1424. The value of this could be further analyzed based on the various groupings of the governments according to the Group variable provided in the WBG GovTech data. In other words, does that value of 0.1424 move the government up one level in maturity?

For example, according to the GovTech Maturity Index document (World Bank Group, 2022), average GTEI values for the 4 Maturity groups are: A=0.845, B=0.648, C=0.362, and D=0.144. The smallest difference between those four levels is 0.197 between Group A and Group B, less than the 0.1424 coefficient that a \_DS\_Strategy\_Program\_ value of 2 would provide. Thus, initial indication is that Tech Skills would not move the organization up a level. However, if we reassessed the model focusing on each group in turn, we may see more significant results in one or more of the maturity groups.

In conclusion, based on this investigation, there is no evidence to indicate the implementation of Tech Skills programs will improve technical capabilities or Project Success within a Government organization.

# References

(2022). *An Overview of World Bank Group Digital Governance and GovTech Projects.* Washington, DC: World Bank. Retrieved from https://datacatalog.worldbank.org/search/dataset/0038056/digital-governance-projects-database

Hanaysha, J. (2016, August). Examining the Effects of Employee Empowerment, Teamwork, and Employee Training on Organizational Commitment. *Procedia - Social and Behavioral Sciences, 229*(19), 298-306. doi:10.1016/j.sbspro.2016.07.140

Khan, R. A., Khan, F. A., & Khan, M. A. (2011, July). Impact of Training and Development on Organizational. *Global Journal of Mangement and Business Research, 11*(7). Retrieved from https://globaljournals.org/GJMBR\_Volume11/8-Impact-of-Training-and-Development-on-Organizational-Performance.pdf

Ochola, G. O. (2018, November). Employee Motivation, An Organizational Performance Improvement Strategy (A Review on Influence of Employee Motivation on Organizational Performance). *JOJ Sciences, 1*(5), 120-125. doi:10.19080/JOJS.2018.01.555575

Schneebacher, J. (2021, May 17). Management practices in Great Britain: 2016 to 2020. *Plus Company Updates*. Retrieved from https://www.ons.gov.uk/economy/economicoutputandproductivity/productivitymeasures/articles/managementpracticesingreatbritain/2016to2020

Truitt, D. L. (2011, December 27). The effect of training and development on employee attitude as it relates to training and work proficiency. *SAGE Open, 1*(3). doi:10.1177/2158244011433338

(2022). *WBG GovTech Maturity Index 2022 Update: Trends in Public Sector Digital Transformation.* Washington, DC: World Bank. Retrieved from http://hdl.handle.net/10986/38499