



# PEFA

## RESEARCH PAPER SERIES

Improving public financial management. Supporting sustainable development.

### *Analyzing the Link between PFM System Strength and Public Investment Performance*

Authors: Tim Robinson, Joseph McGrann, Adam Kadir, Inés Jiménez-Ontiveros, and Uzma Ashraf Barton

(May 2021)

For more information visit  
[www.pefa.org](http://www.pefa.org)

Follow PEFA on social media



#### PEFA Partners



## Abstract

Investment in key infrastructure can promote economic growth and improved service delivery. However, weaknesses in public financial management (PFM) and public investment management (PIM) systems can constrain countries from realizing the full gains of public investments. Existing research and empirical evidence point to positive linkages between the quality of PFM systems, aggregate fiscal discipline, and budget credibility. Assessing the effects of PFM reforms on public investments and service delivery, however, remains underexplored. This study seeks to address this gap by examining the link between the strength of country PFM systems and public investment outcomes. Using an event study framework, we analyze the relationship between improvements in PFM systems (as measured by changing scores on select PEFA indicators) and a range of public investment outcomes (as measured by various international benchmarks for infrastructure quality and infrastructure-related service delivery). Our results suggest that PFM strengthening efforts oriented toward achieving quantifiable intermediate outcomes (e.g., improving adherence to approved budgets) are more likely to lead to observable improvements in public investment and service delivery results. These intermediate outcomes provide a clearer base of assessment to link PFM improvements to public investment outcomes.

---

*This paper is a product of the 2019 PEFA Research Competition that focused on the role of Public Financial Management (PFM) in public service delivery. The PEFA Research Paper Series provides open access to PEFA sponsored research to quickly disseminate knowledge that contributes to ongoing PFM discussions around the world. The broader objectives of the PEFA Research Competition are to contribute to addressing gaps in knowledge on fiscal management, how to improve PFM systems, and the practical implementation of PFM reform. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in the papers are entirely those of the authors. They do not necessarily represent the views of the PEFA Program or those of the PEFA Partners.*

## Analyzing the Link between PFM System Strength and Public Investment Performance

SUBMITTED BY  
Nathan Associates Inc.

AUTHORED BY  
Tim Robinson, Joseph McGrann, Adam Kadir, Inés Jiménez-Ontiveros, Uzma Ashraf Barton



**JEL:** H110, H610, H830

---

## Contents

<b>Acknowledgments</b>	<b>6</b>
<b>Executive Summary</b>	<b>7</b>
<b>Abbreviations</b>	<b>9</b>
<b>1. Introduction, Aim, and Research Question</b>	<b>10</b>
<b>2. Literature Review: Quality of PFM Systems and Outcomes</b>	<b>11</b>
PFM and High-Level Fiscal Conditions	12
PFM and Budget Credibility	13
PFM, Budget Allocations, and Allocative Efficiency	13
PFM and Higher-Level Outcomes	14
<b>3. Conceptual Framework</b>	<b>17</b>
Measuring the Quality and Strength of PFM Systems	17
Selection of the PEFA Framework 2011 Indicators	20
PIM and the Capital Budgeting Process	22
A Model for Understanding Effects	23
<b>4. Methodology</b>	<b>28</b>
Measures of Performance	28
Event Study	30
Transitions	30
Defining Events	30
Time and Measurement	32
Testing Significance in Results Sheets	32
Looking at the Adjustment in Growth	33
Mapping Results to the Capital Budget Process	34
Fitting a Curve to the Data	34
Limitations and Trade-offs	37
Short Time Series	37

Nondiscrete Events	38
Grouping	38
Lag	39
<b>5. Results</b>	<b>41</b>
Descriptive Statistics	41
Main Results	43
Select Tables for Measures with Statistically Significant Results	47
Curve-Fitting Results	56
Shocks and Spending	60
<b>6. Conclusions and Recommendations</b>	<b>62</b>
Conclusions	62
Recommendations for programming	64
Areas for Future Research	64
<b>Appendix A. Considered Measures</b>	<b>67</b>
<b>Appendix B. Results Sheets</b>	<b>71</b>
Aggregate Expenditure Outturn Compared to Original Approved Budget	71
Transport Composite Index	71
People Using Safely Managed Sanitation Services	72
People Using At Least Basic Sanitation Services	73
Estimated Gross Fixed Capital Formation	74
Mobile Cellular Subscriptions	75
Net Acquisition of Non-Financial Assets	76
Composition of Expenditure Outturn Compared to Original Approved Budget	77
Water and Sanitation Composite Index	77
People Using Safely Managed Drinking Water Services	78
People Using At Least Basic Sanitation Services	79
People Using At Least Basic Drinking Water Services	80
Access to Electricity	81
Mobile Cellular Subscriptions	82
Competition, Value for Money, and Controls in Procurement	83

Water and Sanitation Composite Index	83
People Using Safely Managed Drinking Water Services	84
People Using At Least Basic Sanitation Services	85
Mobile Cellular Subscriptions	86
Effectiveness of Internal Controls for Non-salary Expenditure	87
Water and Sanitation Composite Index	87
People Using Safely Managed Drinking Water Services	88
People Using At Least Basic Sanitation Services	89
People Using At least Basic Drinking Water Services	90
Estimated Public Gross Fixed Capital Formation	91
Mobile Cellular Subscriptions	92
Net Acquisition of Non-Financial Assets	93
Multiyear Perspective in Fiscal Planning, Expenditure Policy, and Budgeting	94
Water and Sanitation Composite Index	94
People Using Safely Managed Sanitation Services	95
People Using At Least Basic Sanitation Services	96
Mobile Cellular Subscriptions	97
<b>References</b>	<b>98</b>

## Acknowledgments

---

We would like to thank the World Bank PEFA Secretariat for their support, particularly Urška Zrinski, who guided us in the early days of the study; Karen Ortega, who kept the wheels in motion; and Richard Sutherland, Tia Raappana, James Sampi, and Emeline Bredy, who provided valuable insights and feedback that helped us to refine and improve the final work product.

We also thank the African Development Bank Statistics Department for providing an overview of the African Infrastructure Development Index to inform our analysis.

We would also like to thank Eric Uggen of Nathan Associates for his review, guidance, and support throughout.

Lastly, a special thank you to Janine Mans, who played a key role in the initial design of this study and without whom it would never have gotten off the ground.

All errors remain our own.

## Executive Summary

---

Investment in key infrastructure can promote economic growth, improve the performance of public investment, and achieve higher-level outcomes like infrastructure-related service delivery. Weaknesses in public financial management (PFM) and public investment management (PIM) systems can prevent countries from realizing the full potential of their public investments.

Research and empirical evidence indicate the existence of positive links between the quality of PFM systems, aggregate fiscal discipline, and budget credibility. However, the effects of PFM reforms on public investment and higher-level outcomes related to the provision of public infrastructure and other services have yet to be assessed. This study seeks to address this gap by examining the link between the strength of a country's PFM systems and public investment performance. Public investment performance can be measured through the impact of PIM on outcomes such as public infrastructure quality and growth (IMF 2015, 27). This study sets out to answer the question of whether countries whose PFM performance is improving see greater subsequent improvements in public investment outcomes than countries whose PFM systems are not improving.

Using an event study framework, this study analyzes the relationship between improvements in PFM systems (as measured by changing scores on select PEFA indicators) and a range of public investment performance outcomes (as measured by various international benchmarks for infrastructure quality). Five PEFA indicators from the 2011 framework are used to measure this relationship: PI-12 (Multiyear perspective in fiscal planning, expenditure policy, and budgeting), PI-1 (Aggregate expenditure outturn compared to original approved budget), PI-2 (Composition of expenditure outturn compared to original approved budget), PI-19 (Competition, value for money, and controls in procurement), and PI-20 (Effectiveness of internal controls for nonsalary expenditure). This study links these five indicators to the capital budgeting process and evaluates the relationship between each indicator and a set of public investment performance outcomes. For the purposes of this research, the event study framework divides countries into "improving" and "nonimproving" for each PEFA indicator, defined as having a score that improves by 0.5 or more. In all, 15 measures of public investment and infrastructure-related outcomes are tested, including outcomes related to water and sanitation, electricity and communication, transport, as well as the African Infrastructure Development Index (AIDI) composite indicator.

The results suggest that PFM-strengthening efforts oriented toward achieving quantifiable intermediate outcomes (for example, improving adherence to approved budgets) are more likely to lead to observable improvements in public investment and higher-level outcomes and that those intermediate outcomes provide a clearer base of assessment for linking improvements in PFM to public investment outcomes. A stronger link is found between quantifiable improvements in PFM performance (adherence to approved budgets) and public investment and infrastructure-related outcomes than between qualitative improvements in public financial management (establishment of commitment controls) and those same outcomes.

These conclusions have two main implications. First, individual reforms should be linked to broader intermediate outcomes measured quantitatively. Second, these intermediate outcomes should be the focus of future analysis linking reforms to higher-level outcomes, as individual



components of reforms appear to have limited measurable effect but can significantly influence the quality of public investment performance outcomes in the medium to long term.

This analysis yields several recommendations. First, the areas of PFM performance should be linked to quantitative, intermediate measures that they can reasonably be shown to influence. Second, PFM-strengthening programs should establish quantifiable immediate outcomes for reforms as well as identify complementarity between reforms. Third, future PFM-strengthening initiatives should be based on evidence of their impact on specific PFM performance outcomes.

Recommendations for further research include expanding the scope of indicators, examining the timing and complementarity of PFM reforms and the relationship between quantitative versus qualitative indicators, conducting deeper country-level analysis of the data, and studying the policy implications of this research.

## Abbreviations

---

AAP	Assessment and Action Plan
AIDI	African Infrastructure Development Index
CPIA	Country Policy and Institutional Assessment
DeMPA	Debt Management Performance Assessment
DSA	Debt sustainability analysis
FAO	Food and Agriculture Organization
FE	Forward estimate
FTE	Fiscal Transparency Evaluation
GDP	Gross domestic product
GFCF	Gross fixed capital formation
GFS	Government Finance Statistics
HIPC	Heavily Indebted Poor Countries
IMF	International Monetary Fund
JMP	Joint Monitoring Program
KDE	Kernel density estimation
LPI	Logistics Performance Index
MTF	Medium-term framework
MTFF	Medium-term fiscal framework
OBI	Open Budget Index
OECD	Organization for Economic Co-operation and Development
PER	Public Expenditure Review
PFM	Public financial management
PI	Performance indicator
PIM	Public investment management
PIMA	Public Investment Management Assessment
ROSC	Report on the Observance of Standards and Code
TADAT	Tax Administration Diagnostic Assessment Tool
TEU	Twenty-foot-equivalent unit
UNCTAD	United Nations Conference on Trade and Development
UNICEF	United Nations Children's Fund
WEF	World Economic Forum
WHO	World Health Organization

# 1. Introduction, Aim, and Research Question

Public investment is government spending on the creation or improvement of physical assets, including both economic infrastructure (airports, roads, railways, water and sewerage systems, electricity utilities, telecommunications) and social infrastructure (schools, hospitals, prisons) (IMF 2015; Miller and Hart 2017). Such investment can promote economic growth and improved delivery of vital social services. However, investments that are reliant on public finances may be constrained by weaknesses in public financial management (PFM) and public investment management (PIM) systems and processes. An International Monetary Fund (IMF) study finds an average loss of around 30 percent of the returns on its investments due to PIM inefficiencies (IMF 2015).

Existing knowledge points to the theoretical links between improved PFM system performance and higher and better managed public investment (see section 2). A sound PIM system should adhere to the established goals of public financial management: (1) improved fiscal discipline and sustainability and long-term consistency with total public investment spending; (2) allocative efficiency, in which selected projects align with sectoral priorities and resources are shifted to productive sectors; and (3) technical efficiency, in which projects deliver outcomes and outputs in a more cost-effective manner (Cangiano Curristine, and Lazare 2013). Empirical studies establish that strengthening PFM can lead to these outcomes. Thus, increases in the level of public investment performance may be associated with a PFM system that is better able to prioritize public investment within its budgetary process. Improvements in public investment performance may also indicate an improvement in a country's ability to implement its budget as planned.

Yet while various methodologies and indicators exist for evaluating PFM systems and PIM outcomes, limited research has been conducted on the links between PFM performance, level of public investment, and higher-level outcomes, like the quality of public infrastructure and related services. This dearth of research is a result of both the long-term nature of these reforms and the limited availability of data for identifying future effects. It may also speak to the difficulty of establishing causality from changes in PFM systems and processes to changes in infrastructure and service delivery outcomes, as these changes may be several steps removed from each other.

This study contributes to the research by examining the links between PFM system strength and public investment performance. Public investment performance can be measured through the impact of PIM on public infrastructure quality and growth outcomes (IMF 2015, 27). This study identifies the links between improved budgeting, procurement, and public investment processes, on the one hand, and higher levels of infrastructure investment and performance, on the other. It also measures the extent to which improvements in scores on PEFA indicators and dimensions are associated with (1) increases in the level of public investment and (2) improvements in public

## Box 1.1. Key Terms

**Measure:** the measure of public investment performance

**Indicator:** the PEFA indicator, reflecting a change in public financial management (PFM) quality on a particular metric

**Improving country:** countries seeing an improvement in their PEFA score for the examined indicator during the time period under examination

**Non-improving country:** countries not seeing an improvement in their PEFA score for the examined indicator during the time period under examination

**Test period:** the period during which the PFM improvements occur as measured by two PEFA assessments.

investment performance, measured through a set of infrastructure-related outcomes. Box 1.1 defines key terms used in this study.

In this study, existing measurement frameworks, such as the PEFA framework, are used to establish the scale of PFM improvements as well as the subsequent outcomes related to public investment and infrastructure. Initial steps are also taken to test the significance of these results. The analysis covers a range of sectors, including transport, water and sanitation, and electricity for 68 countries. While it does not resolve all outstanding measurement and causality issues, it does establish a framework for examining these relationships and contributes to an understanding of the impacts of strengthening PFM systems. It sets out to answer the following key question: *Do countries with PFM systems that are improving see greater subsequent improvements in public investment performance outcomes than countries with PFM systems that are not improving?*

To address this question, an event study framework is used that does not impose any assumptions on the path of improvement or deterioration. Significance tests are conducted for these event studies, comparing a set of countries with improving PFM scores to a set without. This effort yields two dimensions of comparison: countries whose PFM scores are improving are compared against themselves over time and against countries whose scores are not improving. Where possible, an assumed structure is fit to an S-shaped curve, and improving and non-improving country groups are tested against each other.

While exploring the link between improving PFM systems and greater subsequent improvements in public investments and higher-level outcomes, this study builds the evidence base for prioritizing certain PFM-strengthening measures to enable better public investment management practices and outcomes. The research is intended for a technical audience, including public finance practitioners, researchers, and scholars, to deepen their understanding of the role of stronger PFM practices in improving public investment outcomes. In other words, this study helps to explain the relationship between PEFA indicators and their impact on PIM outcomes.

The rest of the study proceeds as follows: following a literature review of the existing research on the outcomes of strengthened PFM systems and an explanation of the conceptual framework as well as the data and methodological approach used, the study presents the main results and the implications of the findings, pointing to areas for further research.

## **2. Literature Review: Quality of PFM Systems and Outcomes**

---

This study seeks to assess measurable links establishing the relationship between the strength of a country's public financial management (PFM) systems and its public investment performance. This section conducts a literature review of the existing empirical evidence on the outcomes of strengthened PFM systems. The term "strengthened PFM system," as used here, refers to improved PFM quality and performance, which are measured using a variety of approaches (see *Measuring the Quality and Strength of PFM* in section 3). Some studies focus on the impact of specific PFM reforms, such as the adoption of medium-term frameworks (MTFs), performance-

based budgeting, fiscal rules, and risk management, while others assess the quality of PFM systems more broadly.

Empirical studies point to the impact of strengthened PFM systems on the following outcomes: high-level fiscal conditions (the deficit), budget credibility (adherence to the budget), budget composition (level of spending and allocations), and higher-level outcomes (outcomes tied to the provision of public infrastructure and service delivery). Budget composition and higher-level outcomes remain largely unexplored—despite their potential links to public investment management (PIM) and public investment outcomes.

## **PFM and High-Level Fiscal Conditions**

Multiple studies indicate that a strong PFM system contributes to aggregate fiscal discipline by restraining expenditures and providing strong accountability mechanisms to monitor and enforce budget decisions (Mustapha 2019a; Prakash and Cabezón 2008). Empirical studies examine the impact of PFM reforms and improved PFM systems on aggregate fiscal outcomes. Some studies focus on the impact of specific PFM reforms as opposed to the quality of PFM systems broadly. Vlaicu et al. (2014) use a panel of 181 countries to examine the impact of having a medium-term framework for fiscal discipline, measured by the primary balance. They find that fiscal discipline improves with the adoption of an MTF and that improvements increase with each successive MTF phase. In a series of event studies, World Bank (2013) observes that fiscal discipline increases with the adoption of an MTF and with the progression of MTF phases and that medium-term budgetary frameworks and medium-term performance frameworks are associated with larger improvements than medium-term fiscal frameworks.

Other studies examine cross-country data and assess the quality of PFM more broadly. Using both the Heavily Indebted Poor Countries—Assessment and Action Plan (HIPC AAP) and Country Policy and Institutional Assessment (CPIA)-13 to measure PFM quality, Prakash and Cabezón (2008) conduct ordinary least squares regressions and find that improving PFM leads to improved fiscal outcomes, as measured by the overall fiscal balance and external debt levels for 39 Sub-Saharan HIPCs. These findings echo previous research on PFM reform in Europe and Latin America. Creating their own index of PFM quality from numerous sources, Dabla-Norris et al. (2010) also find that strong budget institutions improve public external debt and fiscal balance outcomes in 65 low- and middle-income countries. This index of PFM quality draws on PEFA assessments, the Organisation for Economic Co-operation and Development (OECD) International Budget Practices and Procedures database, the International Budget Partnership’s Open Budget Index (OBI), and the International Monetary Fund (IMF) Reports on the Observance of Standards and Codes (ROSCs), all of which focus on the expenditure side of a budget (Dabla-Norris et al. 2010).<sup>1</sup> Meanwhile, using PEFA-based measures of the quality of the PFM system, Fritz, Sweet, and Verhoeven (2014) find that improved PFM quality does not improve deficit levels. Using the same measure for PFM quality and adding more controls,

---

<sup>1</sup> Specifically, Dabla-Norris et al. (2010) use the PEFA framework for criteria related to budget performance and practice, the International Budget Partnership OBI and the IMF ROSCs for criteria related to transparency and the “comprehensiveness of fiscal information,” and the OECD database for information on legal regulations and procedural rules (Dabla-Norris et al. 2010, 9). See Dabla-Norris et al. (2010, apps. 1 and 2) for a detailed breakdown of their index and the relevant data sources used.

Mustapha (2019a) also finds no relationship between PFM quality and deficit and debt ratios. In short, despite the mixed evidence, stronger PFM systems are associated with improved fiscal conditions and stronger fiscal discipline.

## **PFM and Budget Credibility**

In addition to improving fiscal discipline, strengthened PFM should also lead to improved budget execution and credibility. A credible budget should reflect minimal deviations from approved allocations, both in aggregate and in composition (Mustapha 2019a). While fewer studies have focused on this relationship, the limited evidence available points to an association between improved quality of the PFM system and a more credible budget.

Extracting data on expenditure deviations from PEFA reports for a sample of 45 countries, Addison (2013) finds a correlation between more accurate budget composition and improved PFM quality, but a small correlation between PFM quality and aggregate expenditure deviation. Meanwhile, Fritz, Sweet, and Verhoeven (2014) find a positive and significant relationship between PFM quality and overall budget credibility.

Drawing from PEFA assessments to study 116 countries, Mustapha (2019a) presents mixed evidence, but overall finds that improving the quality of a PFM system can have a positive impact even in fragile countries by improving budget credibility and reducing the variance in expenditure composition. Meanwhile, in a comparison case study of Kenya and Rwanda, Omollo (2018) finds that PFM reform improves revenue in Rwanda, but not in Kenya. Omollo (2018) illuminates the role of different policy factors in shaping reform outcomes. Other studies also find that factors beyond the type of reform—such as capacity, political commitment, and political economy—can affect the quality and performance of PFM systems and the success of reforms (Andrews 2010; Baudienville 2012; Fritz, Verhoeven, and Avenia 2017; Lawson 2012; Mishra 2014; World Bank 2016).

## **PFM, Budget Allocations, and Allocative Efficiency**

A core objective of a strong PFM system is effective resource allocation—that is, resources should be allocated based on evidence of program effectiveness and reflect government priorities as laid out in strategies (Cangiano, Curristine, and Lazare 2013b). Public investment and recurrent spending needs should align with sector strategies and program objectives (Cangiano, Curristine, and Lazare 2013b). Projects are selected accordingly, and resources subsequently shift to more productive sectors (Cangiano, Curristine, and Lazare 2013b).

Despite their potential relevance for PIM and public investment performance, few works examine the effects of a strengthened PFM system on sectoral expenditure levels. Studies that analyze these effects focus primarily on the health and education sectors. World Bank (2013) finds that having an MTF framework increases health spending as a share of total expenditure; meanwhile, Fukuda-Parr, Guyer, and Lawson-Remer (2011) find that increased budgetary transparency is associated more broadly with increased spending per capita on health and education. However, the bulk of this limited literature focuses on the impact of PFM quality on sectoral budget credibility and allocative efficiency.

The literature once again uses expenditure deviation to assess budget credibility at the sectoral level. Analyzing education and health expenditure deviations for 73 countries, Sarr (2015) finds that increased budget transparency increases the likelihood of having a reliable and credible budget and is associated with higher budget execution rates in the health and education sectors.

The literature defines allocative efficiency as resource allocation that is aligned to government priorities. Studies such as Fritz, Sweet, and Verhoeven (2014), Vlaicu et al. (2014), and World Bank (2013) use expenditure deviation to derive a measure of sectoral expenditure volatility, which in turn serves as a proxy for allocative efficiency.<sup>2</sup> Vlaicu et al. (2014) and World Bank (2013) focus on MTFs and the health sector, finding that the adoption of a medium-term framework is associated with a decrease in the volatility of health spending. Fritz, Sweet, and Verhoeven (2014) also find a positive relationship between improved PFM quality and allocative efficiency, using intersectoral credibility (as captured by PI-2, Composition of expenditure outturn) as an alternative proxy for allocative efficiency.

In short, while existing studies suggest that a strong PFM system can positively affect sectoral spending—in particular by improving allocative efficiency and reducing budget deviations—more research is needed.

## **PFM and Higher-Level Outcomes**

Beyond affecting budget processes, strengthening PFM systems can have higher-level outcomes. Improved PFM systems should increase the transparency of public spending, facilitate its monitoring and oversight, and increase the reliability of service delivery (World Bank 2012). While this area also remains understudied, the literature can be divided into two strands: (1) studies focusing on the technical efficiency of sectoral spending and (2) studies examining the impact of PFM quality on the provision of economic and social infrastructure and other services.

The literature defines technical efficiency as “the ability to produce the maximum possible output from a given set of inputs” (Vlaicu et al. 2014, 11). In other words, it reflects the cost-effectiveness of sectoral expenditures (Fritz, Sweet, and Verhoeven 2014; Vlaicu et al. 2014; World Bank 2013). Several studies derive this measure using stochastic frontier modeling (Fritz, Sweet, and Verhoeven 2014; Vlaicu et al. 2014; World Bank 2013), with life expectancy at birth and primary school enrollment rates as outputs. Vlaicu et al. (2014) and World Bank (2013) focus on a specific component of PFM: medium-term frameworks. Through econometric analysis, Vlaicu et al. (2014) find that only the most developed form of MTF—medium-term performance frameworks—significantly and positively affect the health sector. In an econometric analysis, World Bank (2013) similarly finds that only the adoption of a medium-term performance framework is correlated with an improvement in technical efficiency of public health expenditures. The event studies, however, show minimal evidence of significant change in

---

<sup>2</sup> This study tests the correlation between government expenditure and various measures of PFM performance to assess the relationship between improved PFM and expenditure levels. While some studies use expenditure deviation to assess budget credibility, this study does not use this variable because doing so gives rise to endogeneity issues, given the selection of PEFA indicators as measures of PFM performance (see section 4 for more details). See Vlaicu et al. (2014) for a succinct explanation of the link between spending volatility, allocative efficiency, and improved PFM.



technical efficiency with the adoption of MTFs, while a medium-term budgetary framework is associated with improvements (World Bank 2013). In both studies, results are hampered by small sample size, limited variation in life expectancy, and a short sample period (World Bank 2013, 50). Meanwhile, using a PEFA-based variable, Fritz, Sweet, and Verhoeven (2014) measure the overall quality of PFM systems and find no clear evidence that improved PFM quality increases technical efficiency for health and education expenditures.

Alternatively, Fonchamnyo and Sama (2016) employ data envelopment analysis to calculate efficiency scores. With CPIA-13 as a measure of PFM quality, they study the impact of CPIA-13 scores on health and education outcomes and find a positive and significant association between PFM quality and relevant sector outcomes: life expectancy at birth, infant mortality rates, immunization against measles, and school enrollment rates.

Another part of the literature assesses the direct relationship between improved PFM and service delivery outcomes. While these linkages remain largely unexplored (Goryakin et al. 2017; Piatti-Funkfkirchen and Schneider 2018; Pretorius and Pretorius 2009; Quak 2020; Rao 2013; Welham et al. 2017), the limited empirical evidence indicates that improved PFM has a positive impact on both health and education outcomes (Goryakin et al. 2017; Piatti-Funkfkirchen and Schneider 2018; Quak 2020; Welham et al. 2017).

Using a PEFA-based measure of PFM quality, Welham et al. (2017) find that a stronger PFM system yields positive health outcomes, measured by under-five and infant mortality rates as well as life expectancy at birth. Piatti-Funkfkirchen and Smets (2019) find that, as PFM quality increases, the under-five mortality decreases. Fukuda-Parr, Guyer, and Lawson-Remer (2011) find that budget openness more broadly reduces the under-five mortality rate.

However, some studies find no clear evidence that improved PFM quality leads to better service delivery outcomes (Bandypadhyay 2016; Baudienville 2012; Fritz, Sweet, and Verhoeven 2014; World Bank 2012). This mixed evidence reflects, in part, the difficulty of establishing causal relationships between a PFM system and service delivery (World Bank 2012), which is further affirmed by the lack of reliable tools for isolating and assessing the impact of PFM reforms and PFM quality on service delivery (Rao 2013).

In short, the literature on the effects of strengthened PFM systems focuses largely on assessing aggregate fiscal outcomes and, to a lesser extent, budget credibility. Studying these areas, however, is critical for understanding the extent to which strengthening PFM practices results in improved budget process outcomes, but also wider outcomes related to public investment performance. As Cangiano, Curristine, and Lazare (2013a) explain, sound PIM should adhere to the established goals of public financial management: (1) improved fiscal discipline and sustainability and long-term consistency with total public investment spending; (2) allocative efficiency, in which selected projects are aligned with sectoral priorities and resources are shifted to productive sectors; and (3) technical efficiency, in which projects deliver outcomes and outputs in a more cost-effective manner. Empirical evidence establishes that strengthening PFM can lead to these outcomes. Nevertheless, theoretical links and empirical evidence of the relationship between public financial management, public investment management, and public investment outcomes remain limited.



This study seeks to address this gap by contributing to the evidence establishing the relationship between PFM system strength and public investment performance in terms of improving both budget process outcomes and wider outcomes. It focuses on public investment performance, measured through a set of infrastructure-related outcomes, because most of the existing literature has addressed health and education. It analyzes peculiarities of the capital budget process (see *PIM and the Capital Budgeting Process* in section 3) that differ from the wider budget. In doing so, it examines the link between PFM system strength and public investment performance. The following section presents a conceptual framework for assessing the relationship between strengthened PFM, PIM, and public investment performance.

### 3. Conceptual Framework

This study contributes to the currently limited research examining the links between public financial management (PFM) system strength and public investment performance. Public investment performance can be measured through the impact of public investment management (PIM) on public infrastructure quality (IMF 2015, 27), which is studied here through a set of infrastructure-related outcomes (see *Measures of Performance* in section 4).

The framework underpinning the model is that improved PFM system performance is connected to increased and better managed public investment, which in turn is linked to better provision of economic and social infrastructure—that is, improved public investment performance (figure 3.1). This section explains the model used to examine the links between strengthened PFM and public investment performance. It begins by justifying use of the PEFA framework as the best available measure of PFM system strength, sets out which indicators are the most relevant to public investment, and aligns them with an ideal capital budget process. It ends with a review of how this all fits together in a model for understanding the effects of PFM-strengthening efforts.

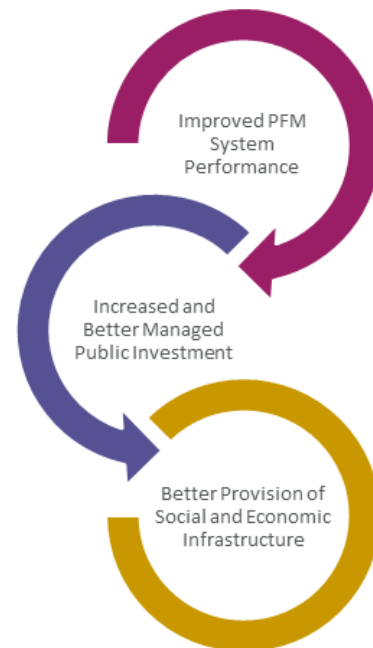
#### Measuring the Quality and Strength of PFM Systems

As countries around the world have implemented changes to improve their PFM and PIM systems, donors and other stakeholders have developed various tools and frameworks for measuring the effectiveness of these reforms.

The literature assessing PFM performance and reform frequently uses the PEFA framework. PEFA organizes key PFM processes into pillars and connects their quality to budgetary outcomes (Kristensen et al. 2019a). Each pillar comprises different indicators of performance. Each indicator has one to four dimensions, which measure different aspects of the PFM system against a four-point ordinal scale from D to A, indicating the level of compliance with international best practices of PFM (see section 4 for further details).

In addition to PEFA, several other diagnostic tools are available for assessing PFM performance. Tools such as the International Monetary Fund’s Fiscal Transparency Evaluations (FTEs), the International Budget Partnership’s Open Budget Index (OBI), and the World Bank’s Public Expenditure Reviews (PERs) offer broad diagnostics (Kristensen et al. 2019a). PERs are often used in case studies or qualitative analyses (see, for instance, Andrews 2010; Baudienville 2012; Hedger and de Renzio 2010; World Bank 2010). Other diagnostic tools evaluate particular elements of PFM; these tools include the International Monetary Fund’s Tax Administration

**Figure 3.1. Linking Public Financial Management (PFM) System Performance, Public Investment Management, and Public Investment Performance**



Diagnostic Assessment Tool (TADAT), the World Bank’s Debt Management Performance Assessment (DeMPA), and the International Monetary Fund’s Public Investment Management Assessment (PIMA). PEFA Secretariat (2018) provides a comprehensive mapping and assessment of the various PFM diagnostic tools, indicating how PEFA complements other diagnostic tools.

Additional frequently used donor-created diagnostic tools include the International Monetary Fund and World Bank’s Highly Indebted Poor Countries (HIPC) Assessment and Action Plan (AAP) and the World Bank’s Country Policy and Institutional Assessment (CPIA) indicators. While the HIPC AAP is only available for the years 2001–06, several studies use it with PEFA as a way to extend data coverage for years prior to the start of PEFA (see, for instance, de Renzio, Andrews, and Mills 2010; de Renzio and Dorontinsky 2007; Lawson 2012; Prakash and Cabezón 2008). Empirical studies often use indicator 13 from the CPIA as a measure of the quality of PFM systems, either on its own or as a robustness check for studies using PEFA scores (see, for instance, Fonchamnyo and Sama 2016; Fritz, Sweet, and Verhoeven 2014; Mustapha 2019a, 2019b; Welham et al. 2017). CPIA-13 measures “the quality of budgetary and financial management” (Mustapha et al. 2019b, 42). Mustapha (2019b, 43) finds that CPIA-13 and PEFA are highly correlated and that PEFA provides comparable overall scores to the scores of both CPIA-13 and OBI (Kristensen et al. 2019b, 15).

PEFA has several advantages over other frameworks. First, it is the most comprehensive measure, as it covers the entire budget cycle as well as other key areas of public financial management (Hadley and Miller 2016; Kristensen et al. 2019b). Second, it is standardized, allowing assessments to be repeated and changes to be tracked over time (Kristensen et al. 2019b). Third, the PEFA Secretariat provides quality assurance, which is an obvious advantage for any data set or analytical framework (Kristensen et al. 2019b, 33).

Last, and perhaps most important, a large body of PEFA data is available for analysis spanning 15 years and more than 600 assessments across more than 150 countries. This coverage represents the most significant advantage of PEFA over PIM-focused frameworks such as PIMA. Although PIMA provides a more thorough and comprehensive examination of PIM systems, as of this writing, fewer than 100 PIMA assessments have been conducted, and results have been published for just 20 of those assessments. PIMA’s relative newness (with the first assessments completed in 2016) also precludes the kind of repeat measurements for individual countries that a study such as this one requires; PEFA’s larger and longer-term data set includes multiple assessments for individual countries, making it possible to determine if, and when, countries’ scores have improved over time. Considering this and the other advantages outlined above, PEFA has become the standard for measuring progress in PFM reform and strengthening (Hadley and Miller 2016).

PEFA, however, does have certain limitations. It does not cover all aspects of the PFM system, such as the political economy of a country or the strategic interactions among actors (Asian Development Bank 2017; Dabla-Norris et al. 2010; Fritz, Verhoeven, and Avenia 2017; Hadley and Miller 2016). Second, PEFA largely measures the “form” of PFM systems—in other words, whether governments have introduced certain formal processes that comply with and follow international best practice. It does not, however, fully assess how these processes work in

practice or whether they are delivering outcomes (Andrews et al. 2014; Hadley and Miller 2016). These limitations not only can incentivize “isomorphic mimicry” by governments seeking to improve their scores, but also can make it difficult to discern whether (for instance) a change in policy or legislation led to a significant change in practice and function. In other words, does the change captured by the PEFA framework reflect a meaningful improvement or just the appearance of improvement? Moreover, within the PEFA framework, certain indicators, such as those tied to de jure, deconcentrated, and upstream PFM functions, are more likely to be amenable to isomorphic mimicry, as they more easily assess the form versus the function of an element of PFM (Andrews 2011; Kristensen et al. 2019b). Andrews et al. (2014) and Hadley and Miller (2016) also provide examples of elements of the PFM system that PEFA measures poorly. These elements include the quality of budget execution and transactions more generally, as the reliability of cash flows, procurement transactions, and wage and salary payments are not assessed explicitly (Andrews et al. 2014, 10). Hadley and Miller (2016, 11) also point out how the indicators of public investment management do not indicate whether “budgeted projects were implemented, if they were implemented at cost, or if they were focused on the right areas.” Critically, as highlighted by Hadley and Miller (2016, 10) “A PEFA assessment does not reveal *how* overall the PFM system is working and *why* it is working that way.”

Lastly, quantifying PEFA scores has posed challenges, including potential time inconsistency and endogeneity concerns when used for quantitative regressions as well as for determining how to enumerate ordinal ranks. Enumerating PEFA scores requires judging the value of the ordinal rankings—that is, should progressing one rank carry the same weight across different intervals (Kristensen et al. 2019b)? In other words, are all changes equal and equally comparable and meaningful (for example, a change from D to C versus a change from C to B)?

Kristensen et al. (2019b) provide a detailed analysis of quantifying PEFA and the different methodologies applied throughout the literature. While studies take different nuanced approaches to calculating and enumerating PEFA scores, the literature overall assumes equal weights between categorical scores (Kristensen et al. 2019b, 28).<sup>3</sup> Identifying and analyzing four main scoring methodologies, Kristensen et al. (2019b) find that all the methodologies share similar descriptive statistics and are highly correlated with each other. They conclude that the methodologies are therefore the same from a statistical perspective, making the choice of methodology largely academic. Moreover, the PEFA Secretariat recently recommended the methodology employed by de Renzio (2009) for quantifying PEFA scores, which provides equal weighting across ranks and assigns numerical values from 1 to 4 for the ordinal score (D to A) of each indicator (Kristensen et al. 2019b, 26).

These challenges emphasize the natural difficulty of assessing the quality of PFM systems. Despite its limitations, PEFA still provides the best framework for assessing the quality and progress of PFM reforms.

---

<sup>3</sup> Some studies omit certain indicators from their PEFA score calculations (de Renzio Andrews, and Mills 2010; Fritz, Sweet, and Verhoeven 2014; Fritz, Verhoeven, and Avenia 2017; Haque et al. 2012; Kristensen et al. 2019b), while others aggregate scores at different levels—either for indicators or underlying PEFA dimensions. See Kristensen et al. (2019b, 28–29) for an overview of different nuanced approaches.

## Selection of the PEFA Framework 2011 Indicators

PEFA has changed its questions and scope over time. In 2016 the original assessment questionnaire (the 2011 framework) was supplemented and changed to expand its areas of measurement. The 2016 PEFA assessment framework expanded the examination of public investment processes to measure more directly the stages of the capital budget process. This study uses the 2011 framework rather than the 2016 framework because the event study methodology requires at least five years after each assessment for any measurement of improvement in PFM or public investment. Currently, insufficient data are available to apply the methodology used here to the 2016 framework. Additionally, while recent assessments have been conducted using the 2016 framework, these assessments cannot be included in the analysis using the 2011 framework. As highlighted by the PEFA Secretariat (2018), while some of the 2011 indicators are carried over into the 2016 framework by name, the 2016 indicators are not directly comparable to the 2011 indicators due to changes in the evaluation criteria underlying the indicators. Hence, a simple cross-walking to include countries with more recent PEFA assessments under the 2016 framework was not feasible for this study. Table 3.1 summarizes the relevant 2016 indicators and how they compare to the original 2011 indicators, as analyzed by the PEFA Secretariat.

**Table 3.1. Comparison of Relevant 2011 and 2016 PEFA Indicators**

2011 PEFA indicator	2016 PEFA indicator	Comparison of scores
<b>PI-1 (i)</b>	PI-1.1	Indirectly comparable
<b>PI-2 (i)</b>	PI-2.1	Indirectly comparable
	PI-2.2	New 2016 indicator
<b>PI-2 (ii)</b>	PI-2.3	Directly comparable
<b>PI-12 (i)</b>	PI-16.1	Not comparable (subject only)
	PI-16.2	New
	PI-16.4	New
<b>PI-19 (i)</b>	PI-24.1	New
<b>PI-19 (ii)</b>	PI-24.2	Not comparable (subject only)
<b>PI-19 (iii)</b>	PI-24.3	Indirectly comparable
<b>PI-19 (iv)</b>	PI-24.4	Indirectly comparable
<b>PI-20 (i)</b>	PI-25.2	Directly comparable
<b>PI-20 (ii)</b>	PI-25.2	Indirectly comparable
<b>PI-20 (iii)</b>	PI-25.2	Indirectly comparable

Source: PEFA Secretariat.

The 2011 PEFA framework contains a total of 28 performance indicators (PIs), which measure several elements of planning, budgeting, and execution, but not all indicators measure those elements directly in relation to public investment. This study selected a set of five indicators most relevant to public investment and the capital budget process without employing a statistical method that might create selection bias (see *PIM and the Capital Budgeting Process* in section 3).

Multiyear budgeting, in theory, ensures provisioning and availability of funds for large investment projects; however, most countries run arrears requiring the need to scale back or stall multiyear capital projects. Often, the process of recording and continuing commitments evolves

across years, and, until a multiyear budgeting framework is fully in place, countries often do not finance public investment projects adequately (and may adjust the cap for public investment to the available resources, regardless of the outstanding commitments). This practice is deemed to have a direct relation to the adequate execution of capital. Revenue raising (for example), while important, has less direct impact, as capital projects are often financed by debt. The following indicators are relevant to PIM under the 2011 framework:

- *Multiyear perspective in fiscal planning, expenditure policy, and budgeting (PI-12)*. Without a genuine multiannual budget planning framework, countries often face challenges in credibly setting aside funds for and ultimately recognizing the multiyear commitments that are often required for public works contracts and infrastructure projects. Lack of a multiyear perspective can also inhibit governments from properly planning for the budgetary adjustments needed to transition projects smoothly from the development stage to the operational stage (that is, capital costs to recurrent costs).
- *Aggregate expenditure outturn compared to original approved budget (PI-1)*. Failure to execute planned expenditure can have an obvious impact on public investment and service delivery outcomes. Poor performance in this PEFA indicator may be a sign of PFM weakness, which could inhibit government's ability to process payments for contracts, maintenance, or operating costs. It may also indicate inadequate revenues to support planned expenditure due to inaccurate projections or insufficient collection. Capital budgets often bear the brunt of budget cuts due to greater political sensitivities involving cutting recurrent costs such as wages or subsidies.
- *Composition of expenditure outturn compared to original approved budget (PI-2)*. Well-planned, multiyear budgets and detailed capital spending plans are only useful so far as they are adhered to. Likewise, improving aggregate expenditure outturn is of limited utility for achieving PIM outcomes if those funds are spent on interventions other than the investments and projects for which they were intended. Accordingly, bringing the composition of expenditure more in line with approved budgets should have a positive impact on PIM outcomes.
- *Competition, value for money, and controls in procurement (PI-19)*. Governments procure goods and services from private sector vendors for the construction, maintenance, and operation of public infrastructure. Governments with better systems and practices in place for ensuring fair, open, and transparent processes for procurement and contracting are less susceptible to corrupt and inefficient dealings and thus more likely to obtain "value for money" and superior PIM results.
- *Effectiveness of internal controls for nonsalary expenditure (PI-20)*. Effective internal controls help to prevent governments from spending funds on unplanned, unauthorized costs, without inhibiting authorized expenditures. When government funds are expended improperly, the funds available for paying legitimate costs are reduced, which can lead to work stoppages, lack of required equipment and supplies, and other issues detrimental to successful PIM projects and service delivery. Accordingly, stronger performance on this

indicator should have a positive impact on PIM outcomes by preventing waste and inefficiencies in the use of funds.

## **PIM and the Capital Budgeting Process**

Public investment spending comes primarily from capital budgets, although recurrent budgets can also contribute to public investments. Thus, to examine the effect of PFM system strength on public investment, it is necessary to understand how the capital budgeting process and the PIM cycle are intertwined. The line between capital and recurrent budgeting, however, can be hard to define. Capital and recurrent expenditures often need to be evaluated as part of the same proposal. Capital spending generally encompasses physical assets with a useful life of more than one year, but it may also include rehabilitation of those assets in future years. One way that capital and recurrent expenditures are differentiated is in the source of their funding. In some countries, capital budgets are funded primarily by borrowing, as opposed to funding for recurrent costs, which comes from tax receipts. Although this may still be the case in some countries, capital budgets can also be funded in different ways, including from general resources, borrowing, or even partnership with the private sector. In the United States, for example, the Highway Trust Fund for road construction and maintenance is financed by excise taxes on gasoline and other fuel products.

Fiscal frameworks can constrain capital budgeting processes. For example, several countries have implemented the “golden rule” that public debt should not exceed net public investment. Furthermore, as capital budgets create assets, they also create a need to monitor and account for these assets and thereby entail balance sheet considerations.

In low- and middle-income countries, donors may set up a specific capital account to fund projects through either grants or concessionary loans. Thus, a significant share of infrastructure and public investment may not be funded by the capital budget, but rather supported by donors. It therefore remains outside the scope of the capital budget process. The methodology used here neither distinguishes nor treats these countries differently, as the focus is on the capital budgeting process. This type of funding is nevertheless an area for further study that could be explored in country-specific contexts in the future.

Now that it is understood that capital expenditures must be treated differently, it is necessary to look at the ways in which it can be planned, costed, and expended differently. The development cycle for public investment projects runs through several steps, set out in table 3.2. The steps examined can be divided into a “precommitment” phase, where the options for the public investment project are developed, appraised, and financed, and a “postcommitment” phase, where the project is procured, implemented, and monitored. These steps are aligned with PEFA indicators that are relevant for that portion of the capital budget process. While some of the steps, such as project identification and option appraisal, are outside the scope of this analysis, the chosen PEFA indicators are otherwise matched to the stages of budget execution that they assess.

**Table 3.2. Stages of Public Investment in Relation to PEFA Performance Indicators and Stages of Budget Execution**

Timing (phase) and capital budgeting activity	Budget execution stage	PEFA indicator(s)
<i>Precommitment (planning and budgeting)</i>		
Strategic objectives, strategies, planning, detailed objectives	Project identification: Sector ministries prepare list of potential multiyear projects.	
Option development	Cost estimation: Sector ministries prepare detailed costs for potential projects.	<b>PI-12:</b> Multiyear perspective in fiscal planning, expenditure policy, and budgeting
Option appraisal	Cost evaluation: Realism of and return on costs are ranked and prioritized.	
Planning financing	Identification of financing sources: Ministry of Finance reviews options to finance appraised projects.	<b>PI-12; PI-1:</b> Aggregate expenditure outturn compared to original approved budget (aggregate outturn)
Budgetary allocation	Allocation: Legislature allocates financing to project.	<b>PI-12</b>
Financing realized	Reservation: Financing is approved and set aside for project.	<b>PI-1; PI-2:</b> Composition of expenditure outturn compared to original approved budget
<i>Postcommitment (implementation)</i>		
Procurement	Obligation: Sector ministry incurs a liability to pay for a project.	<b>PI-19:</b> Competition, value for money, and controls in procurement
Project management	Verification: Project implementing unit ensures that goods and services have been delivered.	<b>PI-2; PI-20:</b> Effectiveness of internal controls for nonsalary expenditure
Budget monitoring and control	Warranting and monitoring: Progress of project is monitored and remunerated through warranted payments.	<b>PI-1; PI-2; PI-20</b>
Asset management	Asset monitoring: Assets are monitored continuously and assessed for quality.	
Performance measurement	Performance audit: Effectiveness of investment is assessed and informs future planning.	

Source: Adapted from Jacobs 2008.

## A Model for Understanding Effects

While PEFA is the most in-depth analysis available of a country's PFM capacity,<sup>4</sup> it is ultimately an assessment of something hard to quantify: What is the quality of a PFM system? *Measuring*

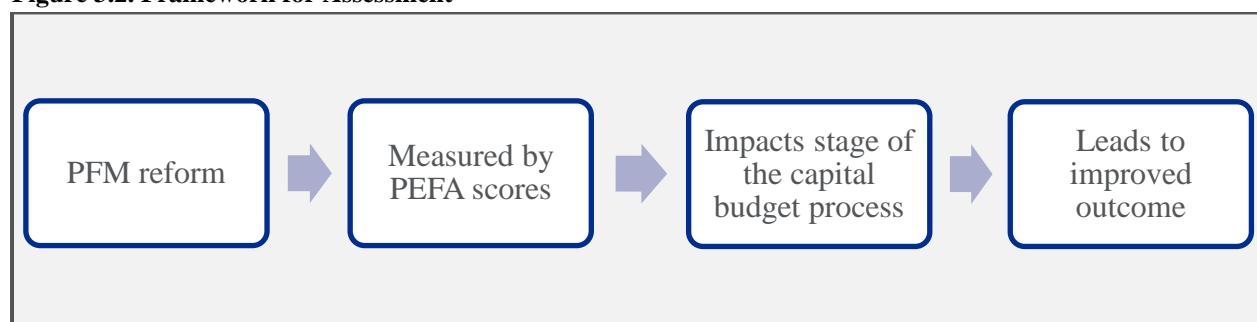
<sup>4</sup> Other measures exist, including PIMAs, which are few; Reports on the Observance of Standards and Codes (ROSCs), which are more bespoke; or CPIA scores, which are more high level. See *Measuring the Quality and Strength of PFM Reforms* in section 3 for a fuller discussion of other tools.



*the Quality and Strength of PFM* in section 3 discusses the degree to which PEFA can be said to speak to the occurrence of an actual PFM reform.

The framework for assessment ultimately assumes that (1) PFM improves or deteriorates; (2) PEFA can adequately measure that improvement or deterioration; (3) the framework aligns with particular stages of the capital budget process; and finally (4) the improvements in this process lead to improved outcomes (figure 3.2).

**Figure 3.2. Framework for Assessment**



PEFA indicators reflect the quality of different stages of the capital budget process (table 3.3). The question asked in reviewing results is, does that stage of the capital budget process matter for this measure? If so, then a relationship is expected between the measure and the PEFA indicators of quality at that stage. Given that the 2016 framework has additional indicators that are related directly to the capital budget process, this analysis will become more refined in the future (once sufficient data are available to use the 2016 framework).

**Table 3.3. PEFA Measurement of Indicators**

Indicator	Description of Measure <sup>a</sup>	Scores
PI-1	<p>“The indicator [measures] the actual total expenditure compared to the originally budgeted total expenditure (as defined in government budget documentation and fiscal reports) but excludes two expenditure categories over which the government has little control. Those categories are (a) debt service payments, which in principle the government cannot alter during the year and which may change due to interest and exchange rates movements, and (b) donor-funded project expenditure, the management and reporting of which are typically under the donor agencies’ control to a high degree.”</p> <p>(PEFA Secretariat 2011, 13)</p>	<p><b>A:</b> Actual expenditure deviated from budgeted expenditure by an amount equivalent to more than 5% of budgeted expenditure in no more than one out of the last three years.</p> <p><b>B:</b> Actual expenditure deviated from budgeted expenditure by an amount equivalent to more than 10% of budgeted expenditure in no more than one out of the last three years.</p> <p><b>C:</b> Actual expenditure deviated from budgeted expenditure by more than an amount equivalent to 15% of budgeted expenditure in no more than one of the last three years.</p> <p><b>D:</b> Actual expenditure deviated from the budgeted expenditure by an amount equivalent to more than 15% of budgeted expenditure in two or all of the last three years.</p> <p>(PEFA Secretariat 2011, 13)</p>

Indicator	Description of Measure <sup>a</sup>	Scores
PI-2	<p>“This indicator requires an empirical assessment of expenditure outturns against the original budget at a subaggregate level. As budgets are usually adopted and managed on an administrative (ministry/department/agency) basis, this is the preferred basis for assessment, but a functional or program basis is acceptable, provided that the same basis is used for both appropriation and reporting execution” (PEFA Secretariat 2011, 14). The first measure looks at the percentage variance. The second measure looks at actual expenditure.</p>	<p><b>A:</b> (i) Variance in expenditure composition exceeded 5% in no more than one of the last three years. (ii) Actual expenditure charged to the contingency vote was on average less than 3% of the original budget.</p> <p><b>B:</b> (i) Variance in expenditure composition exceeded 10% in no more than one of the last three years. (ii) Actual expenditure charged to the contingency vote was on average more than 3% but less than 6% of the original budget.</p> <p><b>C:</b> (i) Variance in expenditure composition exceeded 15% in no more than one of the last three years. (ii) Actual expenditure charged to the contingency vote was on average more than 6% but less than 10% of the original budget.</p> <p><b>D:</b> (i) Variance in expenditure composition exceeded 15% in at least two of the last three years. (ii) Actual expenditure charged to the contingency vote was on average more than 10% of the original budget.</p> <p>(PEFA Secretariat 2011, 14)</p>
PI-12	<p>Multidimensional indicator on whether forecasts are provided for multiple years, if a DSA (debt sustainability analysis) is done, if sector strategies exist and are costed, and the strength of links between outer year forward estimates and the investment budget costings. (PEFA Secretariat 2011, 26)</p>	<p><b>A:</b> (i) Forecasts are done for at least three years (economic and sectoral), and links are clear and explained. (ii) DSA is undertaken annually, strategies are fully costed for 75% of expenditure and consistent with aggregates. (iii) Investments are selected consistently on the basis of sector strategies and recurrent costs and are incorporated in forward estimates (FEs).</p> <p><b>B:</b> (i) Forecasts are done for at least two years (economic and sectoral), and links between FEs and future budgets are clear. (ii) DSA is undertaken for external and domestic debt at least once in three years, and fully costed sector strategies are in line with fiscal aggregates for at least 25%–75% expenditure. (iii) The majority of investments are selected on the basis of sector strategies and recurrent costs and are incorporated in FEs.</p> <p><b>C:</b> (i) Forecasts of fiscal aggregates are done for at least two years (rolling). (ii) DSA is undertaken for external debt at least once in three years, and sector strategies exist but are costed for sectors covering 25% of expenditures. (iii) Links between investment decisions and recurrent costs are weak.</p> <p><b>D:</b> (i) There are no forward estimates, no DSA for three years, and no substantially costed sector strategies. (ii) The capital and recurrent budgets are separate.</p> <p>(PEFA Secretariat 2011, 26)</p>

Indicator	Description of Measure <sup>a</sup>	Scores
PI-19	<p>This assesses “(i) transparency, comprehensiveness, and competition in the legal and regulatory framework, (ii) use of competitive procurement methods, (iii) public access to complete, reliable, and timely procurement information, (iv) existence of an independent administrative procurement complaints system.”</p> <p>(PEFA Secretariat 2011, 38)</p>	<p><b>A:</b> (i) Legal framework meets six listed requirements for transparency. (ii) In all cases contracts not subject to open competition are justified in law. (iii) All of the key procurement information is complete and reliable for 90% of procurement (by value) and is public. (iv) The procurement complaint system meets seven listed criteria.</p> <p><b>B:</b> (i) Legal framework meets five of six listed requirements for transparency. (ii) In 80% of cases contracts not subject to open competition are justified in law. (iii) All of the key procurement information is complete and reliable for 75% of procurement (by value) and public. (iv) Procurement complaint system meets criteria 1 and 2 plus 3 of the others.</p> <p><b>C:</b> (i) Legal framework meets 2 of 6 listed requirements for transparency. (ii) In 60% of cases contracts not subject to open competition are justified in law. (iii) All of the key procurement information is complete and reliable for 50% of procurement (by value) and public. (iv) Procurement complaint system meets criteria 1 and 2 plus 1 of the others.</p> <p><b>D:</b> (i) Legal framework meets 1 or none of 6 listed requirements for transparency. (ii) In less than 60% of cases contracts not subject to open competition are justified in law. (iii) Not available to public. (iv) Procurement complaint system does not meet criteria 1 and 2 or 1 of the others or there is no body for complaints.</p> <p>(PEFA Secretariat 2011, 39)</p>
PI-20	<p>This indicator assesses “(i) effectiveness of expenditure commitment controls, (ii) comprehensiveness, relevance, and understanding of other internal control rules and procedures, (iii) degree of compliance with rules for processing and recording transactions.”</p> <p>(PEFA Secretariat 2011, 40)</p>	<p><b>A:</b> (i) Comprehensive expenditure commitment controls are in place and effectively limit commitments to actual cash availability and approved budget allocations (as revised). (ii) Other internal control rules and procedures are relevant and incorporate a comprehensive and generally cost-effective set of controls, which are widely understood. (iii) Compliance with rules is very high, and any misuse of simplified and emergency procedures is insignificant.</p> <p><b>B:</b> (i) Expenditure commitment controls are in place and effectively limit commitments to actual cash availability and approved budget allocations for most types of expenditure, with minor areas of exception. (ii) Other internal control rules and procedures incorporate a comprehensive set of controls, which are widely understood, but may in some areas be excessive (through duplication in approvals) and lead to inefficiency in staff use and unnecessary delays. (iii) Compliance with rules is quite high, but simplified or</p>

Indicator	Description of Measure <sup>a</sup>	Scores
		<p>emergency procedures are used occasionally without adequate justification.</p> <p><b>C:</b> (i) Expenditure commitment control procedures exist and are partially effective, but they may not comprehensively cover all expenditures, or they may occasionally be violated. (ii) Other internal control rules and procedures consist of a basic set of rules for processing and recording transactions, which are understood by those directly involved in their application. Some rules and procedures may be excessive, while controls may be deficient in areas of minor importance. (iii) Rules are complied with in a significant majority of transactions, but the use of simplified or emergency procedures in unjustified situations is an important concern.</p> <p><b>D:</b> (i) Commitment control systems are generally lacking or they are routinely violated. (ii) Clear, comprehensive control rules and procedures are lacking in other important areas. (iii) Core rules are not complied with on a routine and widespread basis due to direct breach of rules or unjustified routine use of simplified and emergency procedures.</p> <p>(PEFA Secretariat 2011, 40)</p>

a. PEFA measures are drawn directly from the *Public Financial Management Performance Measurement Framework* in PEFA Secretariat 2011.

The indicators fit into two molds: (1) immediate outcomes of PFM processes and (2) direct measurement of PFM processes. In the former, PI-1 and PI-2 are looking at expenditure performance against budget (which itself is a function of the credibility of the budgeting processes); in the latter, they are looking at the other three indicators. These other indicators cover the form and existence of multiyear budget processes, procurement processes and commitment controls, and recording (that is, the existence or otherwise of the processes themselves, not their effect).

## 4. Methodology

---

### Measures of Performance

To assess the links between a public financial management (PFM) system's strength and public investment performance, this study uses the full PEFA data set, including nonpublic assessments, in order to expand the set of possible results. As a result, discussion of individual countries' experiences is limited (which is an area for further study recommended in section 6). The study selects key PEFA indicators aligned to the capital investment process and measures the extent to which improvements in their scores are associated with (1) expansions in the level of public investment and (2) improvements in public infrastructure outcomes.

As explained in section 3, public investment spending comes primarily from capital budgets. Capital is generally considered to consist of assets with a “useful life of more than one year” (Jacobs 2008, 4); however, governments have defined capital and investment in multiple ways. Capital is often conflated with development projects and with investment, which makes it difficult to define measures consistently across countries. The International Monetary Fund (IMF) measures public investment as general government gross fixed capital formation (GFCF), including the total net value of general government acquisitions of fixed assets during the accounting period plus variations in the valuation of nonproduced assets (for example, subsoil assets) (IMF 2015). The analysis presented here examines two measures of spending on capital: (1) estimated public GFCF<sup>5</sup> and (2) cash transactions in nonfinancial assets.<sup>6</sup> Nevertheless, limited information exists across countries, even with the existence of Government Finance Statistics (GFS) standard reporting.

Public investment performance can be measured through the impact of public investment management (PIM) on the quality of public infrastructure (IMF 2015, 27). For the purposes of this study, public infrastructure refers to both economic infrastructure (roads, electricity, airports, railways, water and sewage systems, telecommunications) and social infrastructure (schools, hospitals) (IMF 2015; Miller and Hart 2017).

Key indicators are used to measure infrastructure-related outcomes, starting with a range of measures (set out in appendix A) that generally have significant gaps in the time series. When narrowed further to the sets of countries with both PEFA assessments (at least two) prior to 2013<sup>7</sup> and available data for the measures, several indicators had to be dropped. The results are broken down across the remaining measures of performance, summarized in table 4.1. They include indicators for water and sanitation, electricity and communication, and transport as well as the African Infrastructure Development Index (AIDI). Developed by the African Development Bank, the AIDI uses multiple dimensions and measures of public infrastructure quality. Data in the index typically represent a three-year lag (that is, the index for 2019 represents the conditions in 2016). The index has been relatively consistent in construction, although a review of the

---

<sup>5</sup> Defined by World Bank, World Development Indicators.

<sup>6</sup> Defined by the IMF.

<sup>7</sup> This year was selected to ensure that events were not included where there would be limited or no data following assessment.

components is currently under way,<sup>8</sup> and components are weighted in inverse to their volatility (that is, the more volatile a measure, the less it is weighted in the composite index).

Limited availability of information on some types of infrastructure has limited the scope of examination in certain areas (road networks). Similarly, given data constraints and the availability of appropriate measures, the selected measures do not include indicators for social infrastructure (health and education outcomes). Given these limitations on the scope of analysis, infrastructure could be an area for further study (as data become available or, for instance, through use of a case study methodology).

**Table 4.1. Measures of Public Investment Performance**

<b>Indicator</b>	<b>Measure</b>
<i>Level of public investment</i>	<ul style="list-style-type: none"> <li>• General government total expenditure (% of GDP)</li> <li>• Net acquisition of nonfinancial assets (% of GDP)</li> <li>• Estimated public gross fixed capital formation (% of GDP)</li> </ul>
<i>Infrastructure-related outcomes</i>	
Water and sanitation indicators	<ul style="list-style-type: none"> <li>• People using at least basic drinking water services (% of population)</li> <li>• People using at least basic sanitation services (% of population)</li> <li>• People using safely managed drinking water services (% of population)</li> <li>• People using safely managed sanitation services (% of population)</li> <li>• Water and sanitation composite index (AIDI)</li> </ul>
Electricity and communication indicators	<ul style="list-style-type: none"> <li>• Mobile cellular subscriptions (per 100 people)</li> <li>• Access to electricity (% of population)</li> </ul>
Transport indicators	<ul style="list-style-type: none"> <li>• Air transport, registered carried departures</li> <li>• Propensity to travel (number of passengers per population)</li> <li>• Container port traffic</li> <li>• Transport Composite Index (AIDI)</li> </ul>
Composite indicator (index comprising electricity, information and communication technology, transport, and water and sanitation measures)	<ul style="list-style-type: none"> <li>• African Infrastructure Development Index (AIDI)</li> </ul>

These outcomes are all, in theory, tied to capital investment, which is the effect of interest. However, distinguishing the impact of capital investment from a more general concept of government spending is hard. For instance, while a government may spend capital building schools, schools do not function without teachers. This array of measures is examined to track any possible improvement in expenditure or any more immediate and more long-term improvements in outcomes.

The remainder of this section presents two tools used to examine key outcomes:

---

<sup>8</sup> Team discussions with the African Development Bank Statistics Department.

1. *An event study.* An event study framework is used to examine a cross section of countries with improvements in measured outcomes of public investment, the results of which are presented in detail in appendix B, including sample sizes. Building on the event studies, the significance of the size of changes is tested in countries with and without PFM improvements, explaining shifts in infrastructure measures and running a series of fixed-effects regressions as a robustness check.
2. *Curve fitting.* An attempt is made to fit sigmoid curves to some of the results in order to provide another method of analyzing the scale of difference. This effort is very preliminary and should be extended further in the future. It is done to see if the structural form of the effects of PFM improvements can be approximated and if they lead to sustainable improvements in infrastructure over time.

## Event Study

Event studies are useful in contexts with limited data where the aim is not to show exact causality but to understand the behavior of variables around an event. The event study framework was used initially to avoid imposing any assumptions on the structure of response. In addition, the event study framework has the benefit of not making a strong assumption about the treatment of PEFA scores as continuous and equivalent (the aim is to find an “improvement”). In *Fitting a Curve to the Data* in this section, some of the observed results are taken and fitted into a possible theoretical framework. The event study work takes two points of comparison: (1) Is an increase observed over time in an improving country? (2) Is that increase significantly greater than the increase in nonimproving countries?

A series of event studies are estimated by using measures of both longer-term outcomes of infrastructure performance (access to electricity) as well as more medium-term effects demonstrating changing government focus and ability with regard to infrastructure planning (government expenditure on capital).

## Transitions

Two types of transition are defined:

- Improving countries, defined as countries having a change in score of 0.5 or higher
- Nonimproving countries, defined as countries having a change in score of 0 or lower.

In defining transitions, there is a trade-off between precision and availability of data. The transitions are defined fairly widely in order to capture as many event instances as possible.

## Defining Events

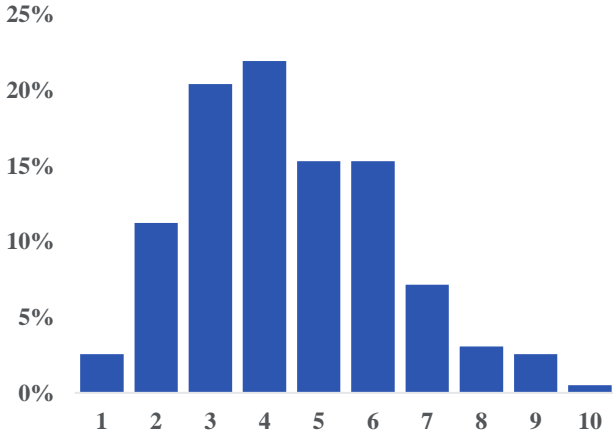
PEFA assessments are carried out at varied periods. The event studied here is defined as “a change between two assessments in the quality of PFM processes.” As noted in the discussion of event study methodology, the event is not necessarily a traditional “event”; rather, it occurs during a period of time rather than at a discrete point in time.



PEFA measures score on an ordinal scale ranked from A to D (as well as some dimensions that are not scored when insufficient information is available). “A” represents the highest performance against the criteria, “D” the lowest. Scores are assessed by a team of specialists during an assessment mission and are finalized through discussions with the country authorities, who have an opportunity to present evidence and challenge the scores given. The conversion to numerical values from de Renzio (2009) is used (where A = 4, down to D = 1, with half steps for scores appended with a plus), in line with the recommendations of the PEFA Secretariat (Kristensen et al. 2019b).

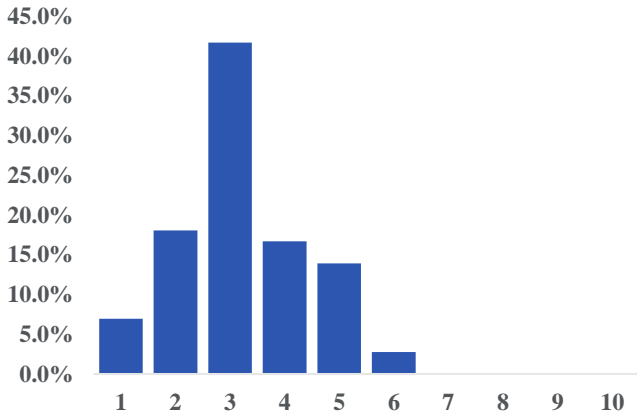
The bulk of the assessments occur within two to four years of each other (figure 4.1). Assessments with insufficient intervening time are dropped; in practice, this is any assessment taking place after 2013.

**Figure 4.1. Years in between PEFA Assessments**



Note: Frequency 196, representing changes between assessments.

**Figure 4.2. Years in between PEFA Assessments, Excluding Post-2013 Assessments**



Note: Frequency 72, representing changes between assessments.

This winnowing reduces the set of periods between assessments and concentrates it further, typically to around a two- to four-year period (figure 4.2). Further qualification in the sample occurs depending on the available data for the measures.

Overlapping changes in assessment are treated as discrete events.

Any “event,” even if measured at the exact moment, is imperfect for PFM. For instance, if a country “introduces” a medium-term fiscal framework in one year, is this really an introduction at a single point, or did a gradual improvement lead to this formal introduction?

Additionally, PEFA is measured in hindsight (Did the budget over the preceding three years meet a given criteria?); hence it is expected that improvements will occur on, around, or just after the test period. The lag, however, should not differ substantially between countries, as much of the lag is caused by the measurement terms of the indicator (see Table 3.3).



Events are measured by the average across all countries, with at least four years post and two years prior to the initial assessment. Averages are taken across index values as well as actual values to weight effects evenly across countries, where doing so may be germane. Median and mean averages are taken across event countries in order to indicate the prevalence of the effect.

### Time and Measurement

Time  $t$  is the time of the first study, and the period  $t$  to  $t + 4$  encompasses the period of reforms or, otherwise, the time between PEFA assessments (that is, the “test period”).

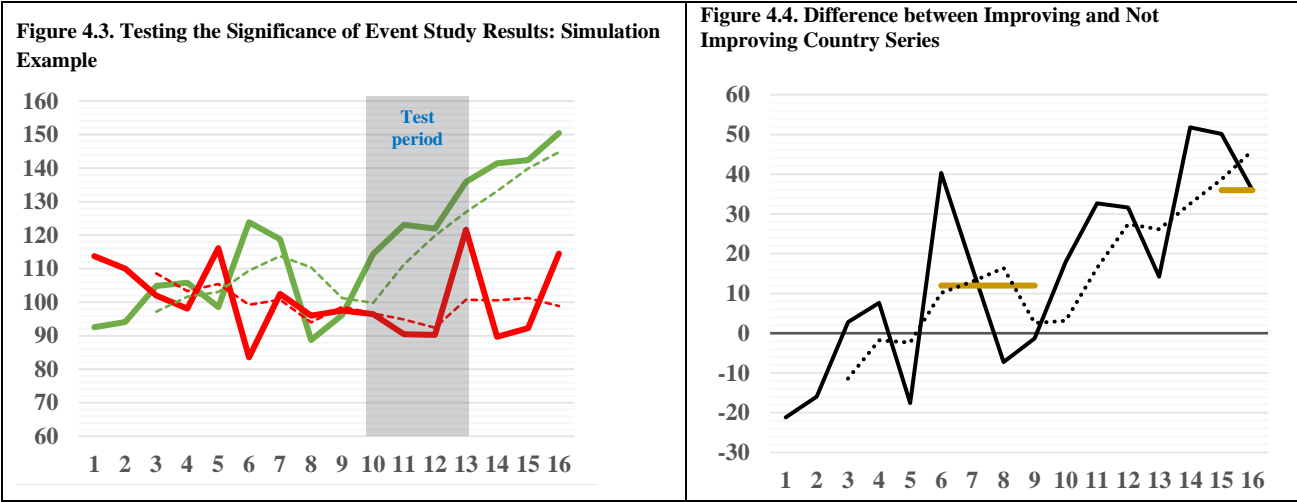
Time  $t + 5$  onward is the period during which it is possible to start seeing most improvements in infrastructure outcomes, although, as outlined previously, there could be a lag before any of these improvements are significant.

Countries that do not see any measured PFM improvements are used as a form of comparison group; in other words, they are used to see if losses in PFM capacity, relative to gains, improve public investment measures.

### Testing Significance in Results Sheets

In the example simulation in figure 4.3, an increase is evident during the test period and after in the improving series (green) compared to the nonimproving series (red). Around the actual improvement is an element of noise reflecting the difficulty of accurately measuring some of the data and other elements. Dotted lines show three-year moving averages.

This simulation appears to show an improvement during the test period and after, consistent with differences between countries with improving PFM and countries without improving PFM.



The two series presented in figure 4.4 show the difference between the improving and nonimproving countries. The gold lines mark the final point, and the average of the series for the four years prior to the test period (and therefore the period in which the PFM process improvement occurs).

This difference is calculated as follows:

$$d = (x_{t=16,i=imp} - \sum_{t=6}^{10} x_{t,i=imp}) - (x_{t=16,i=nimp} - \sum_{t=6}^{10} x_{t,i=nimp}) \quad (4.1)$$

Ultimately, the calculation is testing to see if this difference is significant. A variance is generated from the individual country measures:

$$\sigma_{i=imp,j}^2 = \sum_{j=1}^J (x_{t=16,j} - \sum_{t=6}^{10} x_{t,j}) \quad (4.2)$$

The pooled standard deviation of differences is calculated as follows:

$$\sigma_d = \sqrt{\frac{\sigma_{i=imp}^2}{n_{i=imp}} + \frac{\sigma_{i=nimp}^2}{n_{i=nimp}}} \quad (4.3)$$

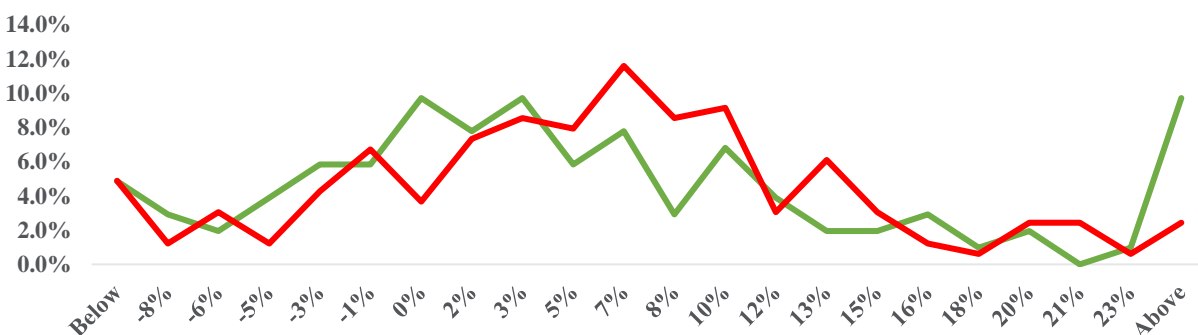
This calculation is then tested against a *t*-distribution with  $(n_{i=imp} + n_{i=nimp} - 2)$  degrees of freedom and standard deviation  $\sigma_d$ . The test threshold is 10 percent.

## Looking at the Adjustment in Growth

In order to indicate the universality, or otherwise, of improvements following reform, the distribution after the test period is examined for those countries that see an improvement in PEFA scores and those that do not. This comparison indicates whether there is a shift in rates of growth across countries. The percentage changes are taken from periods 13, 14, 15, and 16 for both improving and nonimproving countries.

For example, figure 4.5 shows a distinct leftward skew of the improving countries relative to the nonimproving ones, with a sizable number of outliers having significant growth.

**Figure 4.5 Growth Rates of Countries, Disaggregated by Improvement in PEFA Scores**



Finally, as a last check on the robustness of results, the results in a difference-in-differences regression framework are examined, building a data set from countries showing an improvement in PEFA scores and countries not showing an improvement. The treatment therefore becomes the

period after improvement has occurred in the countries that show improvement. This examination uses the following formulation:

$$Y_t = c + T + T * D_t + e_t \quad (4.4)$$

## Mapping Results to the Capital Budget Process

Results from these assessments are then examined in the same framework, looking at each of the PEFA indicators against the capital budget process (set out in *PIM and the Capital Budgeting Process* in section 3). A color code is applied to the overall strength of the relationship between the stage of the capital budget process, on the one hand, and the measured outcomes in public investment performance and PFM improvement, on the other:

- **Red.** There is no clear relationship between PFM improvements and measurement improvements.
- **Yellow.** One (or more) indicator suggests a relationship between PFM improvements and measurement improvements.
- **Green.** All or most indicators clearly indicate a substantial relationship between PFM improvements and measurement improvements.

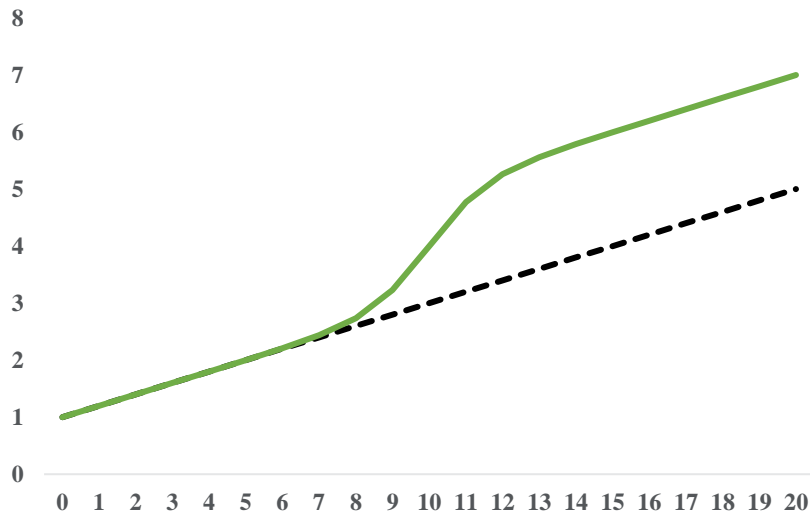
The classification results are presented in section 5, with a discussion of the reasons for a classification.

## Fitting a Curve to the Data

The curve around the improvement in performance, controlled for other factors, is in theory an S-shaped curve, showing an improvement during or just after a period of PFM improvement before returning to trend.

In essence, at a point at which PFM improvement occurs, the improvement shown in figure 4.6 should be evident.

**Figure 4.6 PFM Improvement Resulting in Performance Improvement**

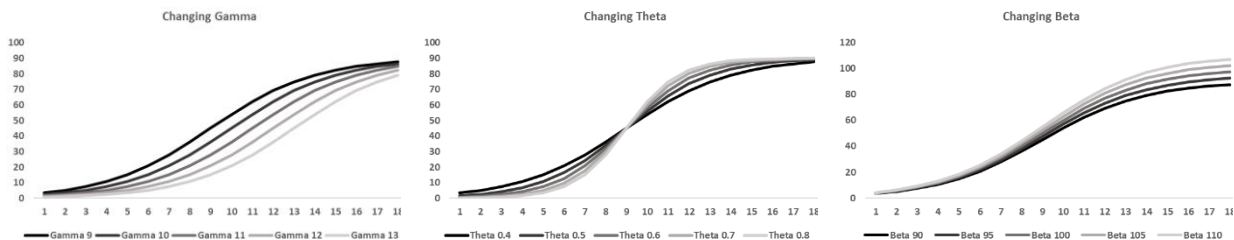


Therefore, a simplified model of four parameters is being solved:

$$c + \rho T + \frac{\beta}{1+e^{-\theta(T-\gamma)}} \tag{4.5}$$

where  $T$  is a measure of the year,  $c$  is the constant,  $\beta$  is the end point of the  $s$  function,  $\theta$  is the slope and the starting point of the  $s$  function.  $P$  is the time trend, where  $C$  would shift the curve up and down, determining the value at  $t = 0$ . This calculation makes it possible to take into account circumstances where different countries would have different starting points in a fixed-effects framework. The analysis is looking for beta and theta values more sizable than those in the countries that do not see an increase in their PEFA scores (figure 4.7). The gamma value should be such that the improvement occurs during or after the test period.

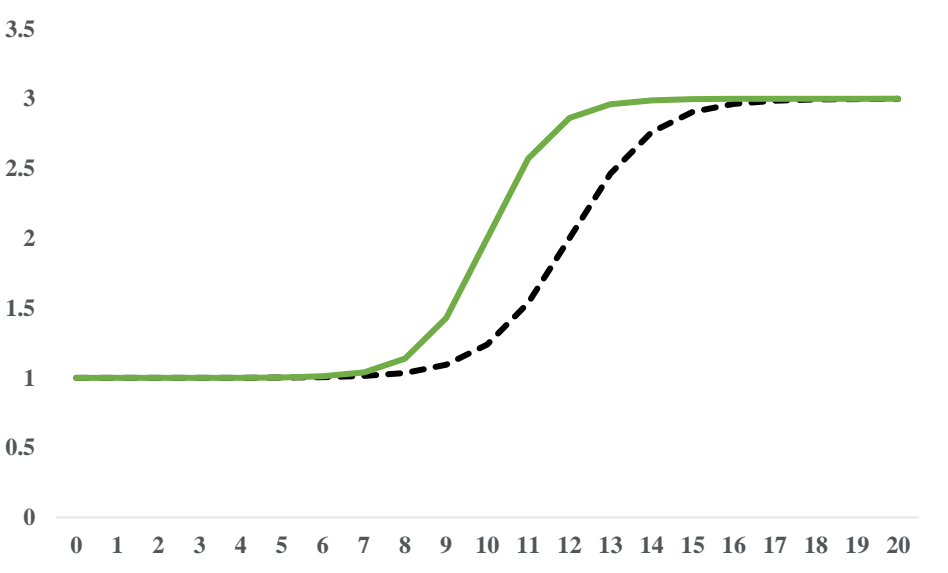
**Figure 4.7. Changing Gamma, Theta, and Beta Values**



This approach entails some issues, including the muddying of results if multiple relevant improvements occur during the period, which would also have a nonlinear impact. However, this approach can be refined further over time.

Although reforms may be more complex in practice, measured infrastructure reforms have an obvious cap in certain cases (once 100 percent of people have access to electricity, it is impossible to expand further) and may endure for long periods. Therefore, a model—with lower or similar gamma and beta values and higher theta values—should not be surprising (figure 4.8).

**Figure 4.8. PFM Improvement Resulting in Performance Improvement, Revised Model**



It is possible to examine other outcomes, such as a permanent change in the rate of improvement (from reforms that have potential ongoing effects on the ability to add to infrastructure). Doing so would require a slightly different theoretical model, with faster rates of change following reform. This model is examined to a degree in the event study assessments. In the theoretical model, the forms that infrastructure improvements take are more in levels: an improvement in PFM competency allows a jump to the next level of infrastructure performance. A case in point is road construction: the rate of construction of new roads is limited by more than the government’s ability to procure; it is also limited by facets like the ability to design projects and non-PFM factors like the ability of firms to operate in the country, number of qualified workers, and state of the labor market. The theoretical model then assumes that removing one PFM constraint is sufficient to create a sudden expansion in activity (in a simplified world, improving procurement makes it possible to clear a backlog of planned projects). However, to improve the rate permanently, many cross-cutting reforms would have to occur.

A minimization model built in Excel is used to estimate this relationship. The model uses a range of starting points to test convergence on a similar outcome (and uses the starting values for each country in the data set as the starting point for the intercept). To assess the divergence from this outcome, the time trend, where used, is fixed as the time trend prior to the test period.

This model can be extended controlling for other variables.

## Limitations and Trade-offs

Several circumstances could cause this work to diverge from a simple event study framework:

1. *Short time series.* There is a risk that reform initiatives will happen at the same time in multiple countries. If multiple countries carry out the same reform at the same time, the risk is that some external co-temporaneous event will be measured that affected all of them. The PEFA assessments have been run from 2005 to 2020 under the 2011 framework. This is a total of 345 individual assessments or 197 changes between assessments.
2. *Nondiscrete events.* Improvements in PFM systems often occur gradually over time, as new processes and reforms are rolled out, practices are institutionalized, and capacity is built. Similarly, the measurement of those improvements occurs during a period of time, as opposed to in a single instance.
3. *Grouping.* PFM reforms are not binary; rather, they can have varying degrees of intensity. This variability holds true for most events for which these forms of studies are used (recessions, devaluations, stock splits).
4. *Lag.* The reforms captured in instances of improved PFM quality are likely to lead to eventual outcomes with a potentially substantial lag, in certain cases. There is also potentially a variable lag—that is, in one country, improvement may take longer, and the form of the reform itself may change the lag (for example, improvements in procurement could even speed the rate of improvement in public investment outcomes).

Event studies have been used to examine seemingly discrete events ranging from devaluations (Acevedo-Espinoza et al. 2015) to financial crises (Hemming, Kell, and Schimmelpfennig 2003) to PFM reforms (World Bank 2013). The methodology was initially developed to examine asset and stock prices, particularly behavior around stock splits (Fama et al. 1969). PFM reforms fit in naturally with the methodology: a clear, discrete event, high availability of data, few gaps in the series, a sizable cross section, and limited risk of concurrent events at the same time. Event studies have also been used for events covering time periods, such as studying the effects of armed conflict or trade and financial liberalization reforms, despite the difficulty of clearly defining these events. Given this wide range of applications, looking at instances in which the quality of PFM improved seems in line with historic usage.

### Short Time Series

The risk in economic and governance contexts comes primarily from the potential for the cross section of events to collapse to a single point in time. For example, many countries experience recessions at the same time. In the context of PFM improvements, there is a small risk that countries were coordinating improvements as fundamental as internal controls or medium-term budgeting. However, such concurrence should not be dismissed when using this methodology in PFM contexts. A reform that becomes a donor focus, for example, may see substantial improvement across countries during the same time period. In addition, ideally a study would

include a constant sample across all examined years; data limitations, however, can increase the difficulty of achieving such a sample.

Studies such as Chen, Loayza, and Reynal-Querol (2007) use a control group comparison with the hopes of separating out events other than peace (for example, a wave of democratization or introduction of a new vaccine) that may affect their variables. Such a comparison allows them to look at whether countries that experienced conflict improve both with respect to their prewar levels and with respect to gains obtained by the “control group”—in this case, a sample of nonconflict low- and middle-income countries. Chen, Loayza, and Reynal-Querol (2007) use an event study methodology to study the effects of civil war on a variety of indicators, including health, education, political development, demographic trends, and economic performance. Their study only includes a country in the sample if it has at least five years of observations both before and after the war when comparing prewar and postwar periods. When only analyzing the aftermath of war, a country is included in the sample if it has at least five years of observations after the war has ended (that is, there is no need to have sufficient prewar data).

Similarly, this study uses nonimproving countries as a comparison group and restricts the requirements for including before and after data.

## Nondiscrete Events

The methodology, as set out in section 4, recognizes that events that may capture instances of reforms or other improvements are not necessarily discrete, and given the periodic nature of PEFA assessments, neither is their measurement. PFM reforms are not the only type of reform that faces the issue of an unclear event period; event studies in other areas, such as trade and debt reforms, face similar difficulties. For instance, Ferry (2015) uses an event study framework to assess the impact of debt relief on government tax efforts. The article looks specifically at the provision of debt relief at different stages of the Heavily Indebted Poor Countries (HIPC) initiative.

Salinas and Aksoy (2006, 10) employ event studies to examine the impact of trade liberalization on economic growth, recognizing that “the exact timing of major trade reform would always be debatable.” Lederman and Rojas (2018), looking at armed conflict and macroeconomic outcomes, and Manova (2008), looking at changes in exports after equity market liberalization, use various definitions for their events to vary the starting point of the event as robustness checks.

This study looks at events phased and measured over time as well, and so the events are presented as occurring over a test period.

## Grouping

Ferry (2015) also faces this issue, as different countries have variable sizes of debt burden prior to the HIPC initiative and thus face varied sizes of impact. Similarly, Chen, Loayza, and Reynal-Querol (2007) define a civil war by the occurrence of more than 1,000 battle-related deaths per year (to focus on major conflicts) and consider both internal and internationalized internal

conflicts as internal wars. The postconflict period is defined by requiring a minimum of 10 years of peace after the war has ended to ensure that time periods considered for the event study are relatively peaceful. As with HIPC debt forgiveness and PFM improvements, there is a grouping on the intensity of war because only a minimum level of intensity is set. Manova (2008) and Chen, Munasib, and Roy (2012) have a variable that captures the “intensity”—that is, the level of liberalization or reforms.

This study sets a minimum level of intensity as well and examines different degrees of intensity of reforms as well as different starting points, as checks against assertions.

## Lag

Finally, PFM reforms face a lag. The results are unlikely to be immediately visible after the event period, except in items more immediately under the country’s direct control (government spending). Chen, Munasib, and Roy (2012) look at the easing of financial constraints on trade flows and examine whether reforms have lagged effects or whether their effects could occur before the reform itself has been initiated due to anticipating the reform. Salinas and Aksoy (2006) look at 12 years before and 9 years after reforms to cover a wide range of effects. Similarly, Ferry (2015) only considers HIPCs for which data 6 years before and 6 years after the “debt relief point” are available.

PEFA assessments themselves work with a lag, which can be variable across countries depending on the fiscal year and availability of information. Therefore, this study takes as long a period as is feasible after the event (where there are sufficient data).

The curve-fitting work presupposes a form of improvement, assuming that infrastructure reform is a linear process over time, sped up by one-off improvements in the capacity of the government. While this is a nice conceptual framework and simplifying assumption, in many countries infrastructure improvements could be more sudden and the result of one-off investments. An attempt has been made to mitigate this situation using a measure of outcome that is more likely to see small, progressive changes. Even then, there are likely to be instances that work against this. For example, countries with a small base of electricity production are likely to see large shifts with the completion of a single power plant. At the same time, the event study work does not assume a structure of the form of improvement and is used to indicate the improvement curve.

Another potential limitation is that the definition of a reform is subjective. While PEFA as a framework sets out clear guiding questions in a standard framework, tries to evidence scores with documentation, and allows a challenge function by authorities, doing all of this does not guarantee consistent treatment. Even on something as arguably clear as internal controls or a medium-term budget framework, the actual practice could differ significantly from the appearance. Finding some relationship between improvements in PEFA indicators and improvements in PIM measures provides at least some assurance that PEFA indicators do potentially seem to be measuring an actual change in government performance.

Lastly, the PEFA framework has some weaknesses and limitations, and whether PEFA is an appropriate framework for this assessment is discussed at length in *Measuring the Quality and*



*Strength of PFM* in section 3. Nevertheless, it is the most extensive and relevant mechanism for assessing PFM quality currently available.

## 5. Results

This section provides the results across the various methodologies, beginning with the descriptive statistics. The full results tables generated by the event study model are set out in appendix B.

### Descriptive Statistics

**Table 5.1. Summary Descriptive Statistics**

Indicator	Status	Average change	Difference
Composition of expenditure outturn compared to original approved budget (raw sample: improving, 47; nonimproving, 85)	Improving	1.25	1.82
	Not improving	(0.57)	
Effectiveness of internal controls for nonsalary expenditure (raw sample: improving, 50; nonimproving, 82)	Improving	0.81	1.06
	Not improving	(0.25)	
Multiyear perspective in fiscal planning, expenditure policy, and budgeting (raw sample: improving, 59; nonimproving, 73)	Improving	0.80	0.94
	Not improving	(0.14)	
Aggregate expenditure outturn compared to original approved budget (raw sample: improving, 49; nonimproving, 83)	Improving	1.43	2.1
	Not improving	(0.67)	
Competition, value for money, and controls in procurement (raw sample: improving, 52; nonimproving, 80)	Improving	1.31	1.81
	Not improving	(0.50)	

Note: Green = average improvements in score for improving countries. Pink = average fall for nonimproving countries. The last column = the difference between the two.

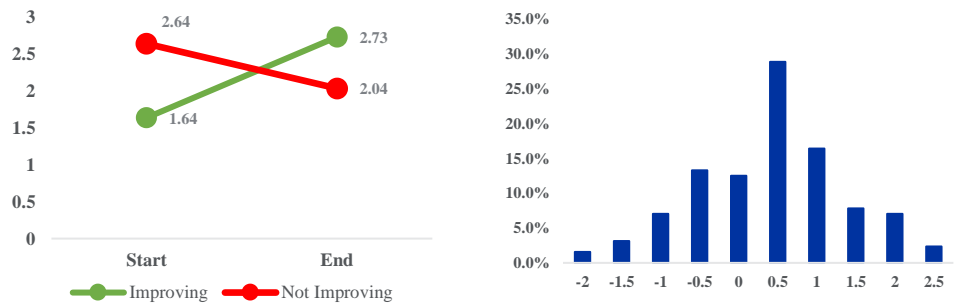
Generally, the improvements are more substantial than the falls in value. However, this study looks at the difference between effects in improving and nonimproving countries, and so the difference between nonimproving and improving is the key metric. On average, controls and multiyear budgeting are the indicators with the smallest differences between improving and nonimproving countries. Again, while PEFA scores are difficult to assign to clear quantitative metrics, this result indicates the intensity of reforms, if not the exact measure.

This section looks at the scale of improvements and the average starting and ending points. The degree to which countries are starting with advanced systems obviously limits the degree to which improvements in measures will be evident from improvements in PEFA indicators. Intuitively, introducing basic controls may be transformative for the government's ability to function, but having a marginally better warranting or cash control system, while beneficial, may not be as impactful.

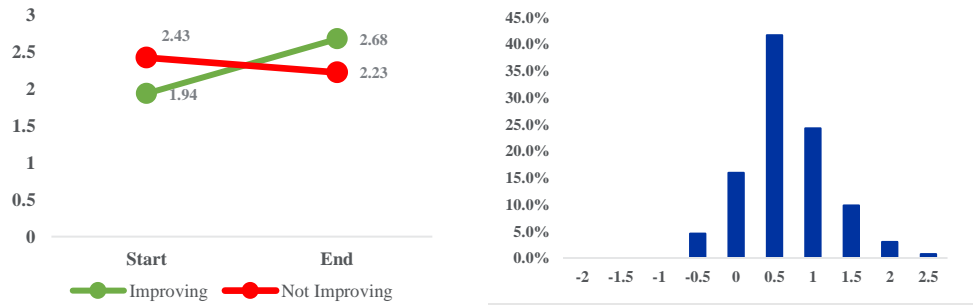
PEFA scores across the chosen indicators generally skew toward showing an improvement between assessments (see the distribution diagrams in figures 5.1–5.5, panel b). *This trend does not reflect final samples, as it is affected by the availability of data under each measure as well.*

Figures 5.1 to 5.5 show the average starting and ending points of each indicator and the distribution of changes across countries. Notably, nonimproving countries tend to show a decline in scores rather than simply a lack of progress. This finding may reflect that scores can be reassessed over time, that factors like expenditure outturn vary year to year, and that poor management can result in the deterioration of systems (among other possible reasons).

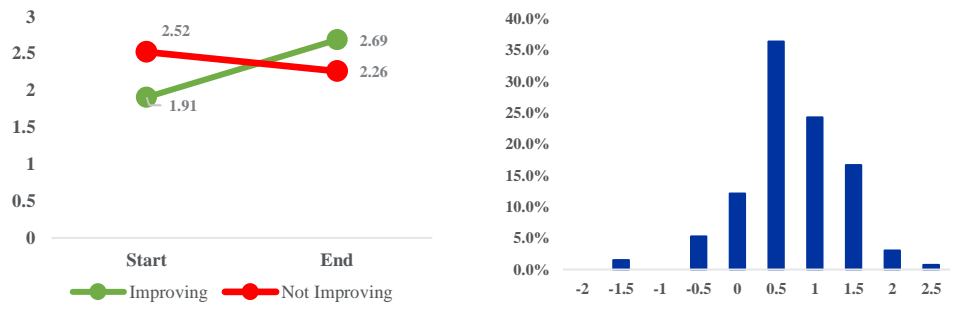
**Figure 5.1. Composition of Expenditure Outturn Compared to Original Approved Budget**



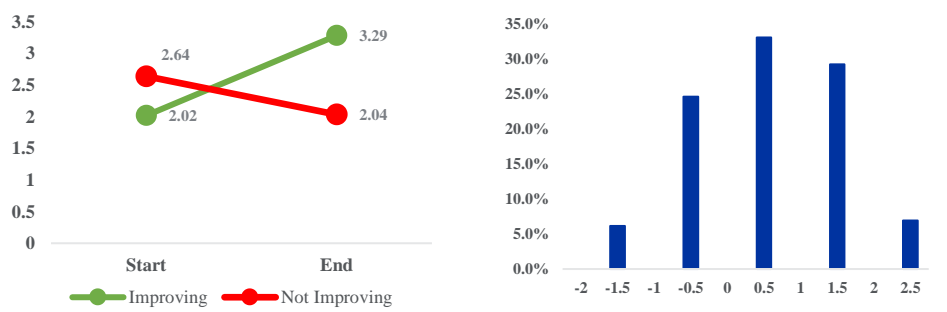
**Figure 5.2. Effectiveness of Internal Controls for Nonsalary Expenditure**



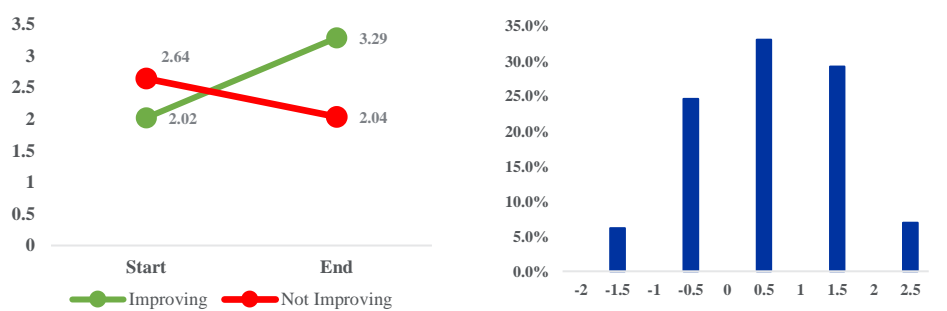
**Figure 5.3. Multiyear Perspective in Fiscal Planning, Expenditure Policy, and Budgeting**



**Figure 5.4. Aggregate Expenditure Outturn Compared to Original Approved Budget**



**Figure 5.5. Competition, Value for Money, and Controls in Procurement**



## Main Results

The results give rise to the following questions:

1. Does the pattern appear to follow that which would be expected, at least part of a sigmoid function showing an improvement in performance in countries that see an increase in their PEFA scores, relative to countries that do not see an improvement in their PEFA scores?
2. Does the change show significance, in line with the test above?
3. Does the distribution of percentage growth change? In other words, is this change seen across multiple countries?

For each public investment or infrastructure-related measure, charts show only the indicators that are useful based on the measure itself. For example, the actual values for propensity to travel (passengers per population) are not useful, given that they are extremely skewed by small, highly tourism-dependent states. For this indicator, the index value is less distorted.

Table 5.2 shows the breakdown of each measure across the capital budget process, flagging whether improvement in that stage of the process can be linked to improvement in measures. The key below shows the color coding. In certain cases, kernel density estimation (KDE) is used to examine changes in the value at the end of the test period ( $t - 13$ ) and at the end of the sample period ( $t - 16$ ) to see if there is a shift in the smoothed distribution of actual values.

Many countries with improving PEFA scores for the selected indicators do see subsequent improvements across a range of public investment and infrastructure-related outcomes. While those outcome improvements are often greater than those seen in the comparator countries (countries whose PEFA scores do not improve), few of these tests show statistically significant results. Following the summary in table 5.2, detailed breakdowns are provided for several measures that show statistically significant results. Table 5.3 presents regression results for key measures. The similar scale of effects (for example, on cell phones) could indicate that these changes in score represent a more broad-based improvement in PFM, with multiple changes occurring concurrently. Annex B presents further figures to highlight the relationships between the five selected PEFA indicators and measures that have statistically significant outcomes.

**Table 5.2. Summary of Results**

Indicator	Option development	Planning financing	Budgetary allocation	Financing realized	Procurement	Project management	Budget monitoring and control
Total government expenditure (% GDP)							
Gross fixed capital formation (% GDP)							
Investment in nonfinancial assets (Government Finance Statistics [GFS]) (% GDP)							
People using at least basic drinking water services (% of population)							
People using at least basic sanitation services (% of population)							
People using safely managed drinking water services (% of population)							
People using safely managed sanitation services (% of population)							
Electricity access (% of population)							
Mobile cellular subscriptions (per 100 people)							
Air transport carrier departures (number)							
Propensity to travel (number of passengers per population)							
Container port traffic (number)							
AIDI Transport Index							
AIDI Water and Sanitation Index							
AIDI Composite Index							

Note: Green = effect. Yellow = limited effect. Pink = no effect.

**Table 5.3. Regression Results for Key Measures and Indicators**

Measure and indicator	Constant	Time trend	Treatment	R <sup>2</sup>
<i>Electricity access</i>				
Multiannual	42.7807*** (0.000)	1.168629*** (0.000)	-.1661895 (0.220)	0.4453
Composition outturn	43.24697*** (0.000)	1.056155*** (0.000)	.005232 (0.975)	0.4390
<i>Basic sanitation</i>				
Controls	40.10465*** (0.000)	.4293573*** (0.000)	.2133495** (0.034)	0.3926
Procurement	41.34186*** (0.000)	.4690021*** (0.000)	.1324313 (0.186)	0.3882
Composition outturn	41.26348*** (0.000)	.485488*** (0.000)	.138848 (0.166)	0.3882
<i>Air carriage</i>				
Multiannual	10,987.38 (0.127)	1066.267*** (0.008)	1,694.145 (0.256)	0.1011
Controls	10,163.69 (0.145)	1290.832*** (0.001)	1,923.831 (0.382)	0.1095
<i>Safe water</i>				
Controls	38.92496*** (0.000)	.6413421*** (0.001)	.1012977 (0.621)	0.2630
Multiannual	38.99757*** (0.000)	.6250286*** (0.002)	.0899312 (0.603)	0.4634
Procurement	39.39002*** (0.000)	.5391969*** (0.003)	.2833052 (0.161)	0.5000
<i>Cell phones</i>				
Multiannual	-6.102297** (0.015)	5.774277*** (0.000)	.5318694* (0.097)	0.8009
Composition outturn	-6.139585** (0.015)	5.774677*** (0.000)	.6776489* (0.051)	0.8027
Aggregate outturn	-5.784346** (0.025)	5.709845*** (0.000)	.6970245** (0.031)	0.8031
Controls	-7.002941*** (0.007)	5.980291*** (0.000)	.2702393 (0.441)	0.7983
<i>Basic water</i>				
Controls	62.79172*** (0.000)	.695724*** (0.000)	.0717763 (0.532)	0.5144
Multiannual	62.08892*** (0.000)	.8595071*** (0.000)	-.1938* (0.059)	0.5330
<i>Investment in acquisition of nonfinancial assets</i>				
Composition outturn	4.220448*** (0.000)	.0103712 (0.793)	.0095212 (0.819)	0.0012
Aggregate outturn	4.257693*** (0.000)	.002148 (0.961)	.0200912 (0.667)	0.0021

Note: Regression sample sizes are noted in the results sheets in appendix B.

\*<0.1 \*\*<0.05 \*\*\*<0.01

## Select Tables for Measures with Statistically Significant Results

### *Acquisition of Nonfinancial Assets*

**Table 5.4 Investment in nonfinancial assets (% of GDP)**

Stage	Indicator	Relevance to measure
2. Option development	PI-12 (Multiyear)	No statistically significant results
4. Planning financing	PI-12 (Multiyear) + PI-1 (Aggregate outturn)	Aggregate: significant improvements in the index and mean actual values. Multiyear: no significant improvements
5. Budgetary allocation	PI-12 (Multiyear)	No statistically significant results
6. Financing realized	PI-2 (Composition outturn) + PI-1 (Aggregate outturn)	Aggregate: significant improvements in the index and mean actual values. Composition: improvements, but not significant
7. Procurement	PI-19 (Procurement)	No statistically significant results
8. Project management	PI-20 (Controls) + PI-2 (Composition outturn)	Controls: significant improvement in mean actual values. Composition: improvements in index and median values, but not statistically significant
9. Budget monitoring and control	PI-20 (Controls) + PI-1 (Aggregate outturn) + PI-2 (Composition outturn)	Aggregate: significant improvements in the index and mean actual values. Composition: improvements, but not significant

Source: World Bank (GC.AST.TOTL.GD.25).

Note: Green = effect. Yellow = limited effect. Yellow = no effect.

Strong expenditure management practices can help to ensure that public investments receive the resources required to bring projects to completion such that new capital becomes usable. Accordingly, positive relationships are evident between improving aggregate outturn, composition of outturn, and internal controls with net acquisition of nonfinancial assets (table 5.4). While the results for composition are not significant, the other two indicators do show statistically significant improvements for some measurements. Notably, the aggregate outturn measure pulls in countries with gaps in the data series (Afghanistan and Vanuatu); if these countries are omitted, there is an additional statistically significant improvement. The indicator for procurement shows no statistically significant correlation with this measure. This result may reflect the indicator's focus on transparency and value for money, while the measure in question



(net acquisition of nonfinancial assets) can just as easily reflect the acquisition of assets in nontransparent transactions with poor value for money.

### *Water and Sanitation Services*

**Table 5.5. People Using at Least Basic Drinking Water Services**

Stage	Indicator	Relevance to measure
2. Option development	PI-12 (Multiyear)	No statistically significant results
4. Planning financing	PI-12 (Multiyear) + PI-1 (Aggregate outturn)	No statistically significant results
5. Budgetary allocation	PI-12 (Multiyear)	No statistically significant results
6. Financing realized	PI-2 (Composition outturn) + PI-1 (Aggregate outturn)	Aggregate: no significant improvements. Composition: significant results for median values. KDE: results not meaningful
7. Procurement	PI-19 (Procurement)	No statistically significant results
8. Project management	PI-20 (Controls) + PI-2 (Composition outturn)	Both indicators: significant results for median actual values. KDE: results not meaningful
9. Budget monitoring and control	PI-20 (Controls) + PI-1 (Aggregate outturn) + PI-2 (Composition outturn)	Aggregate: no significant improvements. Composition and controls: significant results for median actual values. KDE: results not meaningful

Source: World Health Organization and United Nations Children’s Fund data.

Note: Pink = no effect. KDE = kernel density estimation.

Well-functioning public water utilities and sanitation systems require proper planning and design. These systems, particularly urban water and sanitation systems in rapidly expanding cities, can require sizable investments. Large, complex, multiyear water and sanitation projects require proper budgetary allocation over multiple years to ensure project completion and proper delivery of services once construction is complete (for example, transitioning from capital to recurrent costs). Strong expenditure management processes are also required to ensure that projects are managed to completion and that operations and maintenance are executed properly and deliver services over the long term.

Four measures indicative of the quality of water and sanitation infrastructure and services are examined. Due to their interrelated nature, the relationships between public financial management (PFM) and public investment management (PIM)—and thus the results of the tests

performed—are expected to be similar across the four measures. Results for three of these measures (the last measure, covering safely managed sanitation services, is not presented because the sample size is too small for meaningful results).

Table 5.5 presents the results for basic drinking water services. These measures show no meaningful relationships between the tested indicators and service delivery outcomes. Indicators for composition of outturn and internal controls do show some small improvement, including statistically significant results for median actual values; however, KDE of those tests points to very weak, likely not meaningful, relationships.

**Table 5.6. People Using at Least Basic Sanitation Services**

Stage	Indicator	Relevance to measure
2. Option development	PI-12 (Multiyear)	Significant improvement for median actual values
4. Planning financing	PI-12 (Multiyear) + PI-1 (Aggregate outturn)	Aggregate: significant improvement for mean actual values. Multiyear: significant improvement for median actual values
5. Budgetary allocation	PI-12 (Multiyear)	Significant improvement for median actual values
6. Financing realized	PI-2 (Composition outturn) + PI-1 (Aggregate outturn)	Both indicators: broad improvements. Aggregate outturn: significant results for mean actual values. Composition: significant results for mean and median actual values
7. Procurement	PI-19 (Procurement)	Broad improvements, with significant results for mean and median values
8. Project management	PI-20 (Controls) + PI-2 (Composition outturn)	Controls: significant improvements across-the-board. Composition: broad improvements, with significant results for mean and median actual values
9. Budget monitoring and control	PI-20 (Controls) + PI-1 (Aggregate outturn) + PI-2 (Composition outturn)	Controls: significant improvements across-the-board. Composition: broad improvements, with significant results for mean and median actual values. Aggregate: significant improvement for mean actual values

Source: World Health Organization and United Nations Children’s Fund data.

Note: Green = effect. Yellow = limited effect.

The measures for basic sanitation services in table 5.6 show strong results, particularly on the execution side. All indicators have at least one positive statistically significant result, while internal controls, procurement, and composition of outturn each have two or all three results showing positive significant correlations.

As with basic sanitation services, the measure for safe drinking water services in table 5.7 shows some evidence of improvement linked to all indicators, but stronger and more significant results on the execution side. Procurement and controls both show large and significant improvements. Aggregate outturn and multiyear budgeting show small improvements across multiple measures, but none is significant. Composition shows significant improvement for median actual values, but other measures indicate no positive relationship.

**Table 5.7. People Using Safely Managed Drinking Water Services**

Stage	Indicator	Relevance to measure
2. Option development	PI-12 (Multiyear)	Small improvements across all measurements, but none significant
4. Planning financing	PI-12 (Multiyear) + PI-1 (Aggregate outturn)	Some improvements in each indicator, but none significant
5. Budgetary allocation	PI-12 (Multiyear)	Small improvements across all measurements, but none significant
6. Financing realized	PI-2 (Composition outturn) + PI-1 (Aggregate outturn)	Mixed results for both indicators. Composition: significant results for median actual values, but other measurements showed opposite results
7. Procurement	PI-19 (Procurement)	Improvements across-the-board, with median actual values showing large and significant improvement
8. Project management	PI-20 (Controls) + PI-2 (Composition outturn)	Controls: improvements across-the-board, with median actual values showing large and significant improvements. Composition: mixed results
9. Budget monitoring and control	PI-20 (Controls) + PI-1 (Aggregate outturn) + PI-2 (Composition outturn)	Controls: improvements across-the-board, with median actual values showing large and significant improvements. Aggregate outturn: mixed results, with none significant. Composition: significant improvement for median actual values, but no improvement for other values

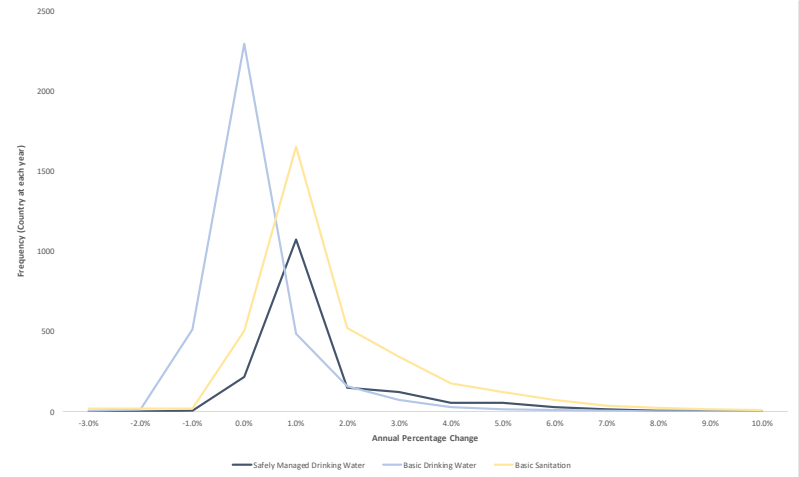
Source: World Health Organization and United Nations Children’s Fund data.

Note: Green = effect. Yellow = limited effect. Pink = no effect.

The results of water and sanitation tests are not consistent across the four measures examined. While measures for basic sanitation services and safely managed drinking water services do show similar and statistically significant results, basic drinking water services do not show any meaningful positive correlations. Further examination of the data suggests a possible reason for this inconsistency: while the two comparable measures show notable changes in countries' scores over the test years, scores for basic drinking water services show little to no change across most of the data set. With such limited change observed in the measure, it would be difficult to generate any statistically significant results.

Figure 5.6 charts the number of country observations corresponding to annual percentage changes in their scores for three water and sanitation measures. Country observations for basic drinking water are clustered tightly around 0 percent, while the other two measures cluster around 1 percent annual change. The latter two measures also show consistently more observations with higher annual percentage changes in their scores. The likely reason for this disparity is that most countries in the world had already attained very high levels of access to basic drinking water by the time they entered the PEFA scoring period (the data bear this out); as such, this measure shows relatively little upward mobility. In contrast, basic sanitation services have generally lagged behind basic access to drinking water worldwide, and safely managed drinking water represents a further step up in quality of service over basic drinking water. Accordingly, both of those measures had greater potential to increase during the PEFA scoring period. Analysis of the data suggests that this was the case.

**Figure 5.6. Water and Sanitation Curves**



## Cell Phone Access

**Table 5.8. Mobile Cellular Subscriptions (per 100 People)**

Stage	Indicator	Relevance to measure
2. Option development	PI-12 (Multiyear)	Multiyear budgeting: improvement across index and mean actual values
4. Planning financing	PI-12 (Multiyear) + PI-1 (Aggregate outturn)	Multiyear budgeting: improvement across index and mean actual values. Aggregate: potential improvement, with significance on mean actual values
5. Budgetary allocation	PI-12 (Multiyear)	Multiyear budgeting: improvement across index and mean actual values
6. Financing realized	PI-2 (Composition outturn) + PI-1 (Aggregate outturn)	Composition of outturn: potential improvement, with significance on median actual values. Aggregate: potential improvement, with significance on mean actual values
7. Procurement	PI-19 (Procurement)	Mean actual values: significant, with rise occurring during test period. Median actual values: close to significant, with rise occurring during test period
8. Project management	PI-20 (Controls) + PI-2 (Composition outturn)	Internal controls: significant or nearly significant improvement for measures following test period. Composition of outturn: potential improvement across median average actual values
9. Budget monitoring and control	PI-20 (Controls) + PI-1 (Aggregate outturn) + PI-2 (Composition outturn)	Internal controls: significant or nearly significant improvement for measures following test period. Composition of outturn: potential improvement across median average actual values. Aggregate: potential improvement, with significance on mean actual values

Source: World Bank (CEL.SETS).

Note: Green = effect. Yellow = limited effect.

Cell phone subscriptions are influenced by the availability of infrastructure, including towers, which can be affected by government policy and direct government investment, but cell phone

service is often managed through public corporations and not directly by government. While all of these investments are individually relatively small, master planning is likely to be key to infrastructure rollout, and multiyear budgeting and forecasting can support the long-term financial planning aspects. In table 5.7, step 6 of the capital budget process—financing realized—shows a weaker relationship, which implies that reforms in this area are less integral to these improvements. This result makes intuitive sense because more sizable road projects are most likely to stop if financing is not realized, while smaller, higher-priority investments, like expanding phone infrastructure, may continue.

For further analysis of this measure, see the discussion in *Curve-Fitting Results* in this section.

### *Propensity to Travel*

**Table 5.9. Propensity to Travel (Number of Passengers per Population)**

Stage	Indicator	Relevance to measure
2. Option development	PI-12 (Multiyear)	Multiyear: no significant relationship, but some follow pattern for index
4. Planning financing	PI-12 (Multiyear) + PI-1 (Aggregate outturn)	Aggregate: no significant relationship, but with some pattern of movement, though not significant. Multiyear: no significant relationship, but some follow pattern for index
5. Budgetary allocation	PI-12 (Multiyear)	Multiyear: no significant relationship, but some follow pattern for index
6. Financing realized	PI-2 (Composition outturn) + PI-1 (Aggregate outturn)	Composition: no significant relationship. Aggregate: no significant relationship, but with some pattern of movement, though not significant
7. Procurement	PI-19 (Procurement)	No significant impact
8. Project management	PI-20 (Controls) + PI-2 (Composition outturn)	Controls: significant impact on index and follows pattern, with actuals showing shift in distribution of values (see KDE chart). Composition: no significant relationship
9. Budget monitoring and control	PI-20 (Controls) + PI-1 (Aggregate outturn) + PI-2 (Composition outturn)	Controls: significant impact on index and follows pattern, with actuals showing shift in distribution of values (see KDE chart). Composition: no significant relationship. Aggregate: no significant relationship, but with some pattern of movement, though not significant

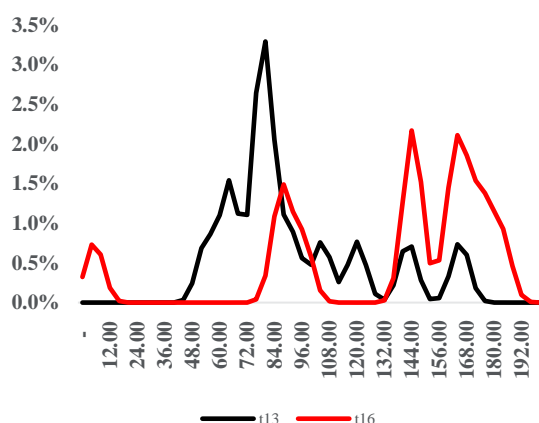
Source: Based on author's calculations, drawing from the World Bank (IS.AIR.PSGR)

Note: Yellow = limited effect. Pink = no effect.

The index value is more relevant for propensity to travel (table 5.9), given that the results can be easily skewed by small, tourism-dependent countries (particularly, in this instance, Seychelles). This measure consists of the number of passengers divided by the country's total population. Small island states, especially tourism-dependent ones, therefore exercise a significant influence over the pattern of results.

Figure 5.7 shows the KDE of the distribution of index values in improving countries, before and after the improvement in PEFA indicators (for controls).

**Figure 5.7. Kernel Density Estimation of the Distributions of Index Values**



### Water and Sanitation Index

**Table 5.10. AIDI Water and Sanitation Index**

Stage	Indicator	Relevance to measure
2. Option development	PI-12 (Multiyear)	Multiyear: some significant increase for median average actual value, with some shift in the curve
4. Planning financing	PI-12 (Multiyear) + PI-1 (Aggregate outturn)	Multiyear: some significant increase on median average actual values, with some shift in the curve. Aggregate outturn: some increase, but not significant, with substantial shift in the curve
5. Budgetary allocation	PI-12 (Multiyear)	Multiyear: some significant increase on actual values for median average, with some shift in the curve

6. Financing realized	PI-2 (Composition outturn) + PI-1 (Aggregate outturn)	Composition: some significant increase on actual values of median average, with some shift in the curve. Aggregate outturn: some increase, but not significant, with substantial shift in the curve
7. Procurement	PI-19 (Procurement)	Some significant increase for actual values of median average, but not clear in test period
8. Project management	PI-20 (Controls) + PI-2 (Composition outturn)	Composition: some significant increase for actual values for median average, with some shift in the curve. All controls: some increase, registering as significant for mean actual values
9. Budget monitoring and control	PI-20 (Controls) + PI-1 (Aggregate outturn) + PI-2 (Composition outturn)	Composition: some significant increase for actual values for median average, with some shift in the curve. Aggregate outturn: some increase, but not significant, with substantial shift in the curve. All controls: some increase, registering as significant for actual mean values

Source: African Development Bank (African Infrastructure Development Index)

Note: Yellow = limited effect. Pink = no effect.

The water and sanitation index reported in table 5.10 shows more substantial relationships between improvements in PFM processes and improvements in the measured outcomes. This may occur because it is a more focused index, but also is in line with results seen across water and sanitation measures. Realization of financing has less of an immediate effect, potentially reflecting the priority nature of this spending: even in instances where government has failed to finance the budget fully, this area of expenditure typically would not be reduced.



## Curve-Fitting Results

Curve fitting works best with data series that are more complete and have discrete observations as opposed to those with extrapolated data points. Extrapolated data (for example, AIDI and multiple sanitation indexes) typically draw a straight line between measured points. Although growth before and after the test period may be greater in improving than in nonimproving countries, the exact point of growth is less clear.

A smaller set of measures can be examined within this framework, mainly the measures of government expenditure, cell phone access,<sup>9</sup> electricity access, and carrier departures. In addition, a sufficient period of time is needed for a return to trend activity.

### *Carrier Departures*

Carrier departures provide a good option for testing the other framework for examination. This measure has continuous data and does not have large portions of the series that are extrapolated.

The data are the log of carrier departures, which is used to linearize the data. In addition, the data are cleaned to remove countries that do not have data for all years from period 1 through to period 16.

Several restrictions are imposed to limit the scope of estimation that has to be undertaken. The time trend is set as the linear time trend prior to period 9. The theoretical assumption is that this kind of infrastructure improvement has an underlying growth rate (especially measures without a natural cap—for example, 100 percent of population). Constants are established for each country separately in a fixed-effects framework based on the initial value in the data set as a starting point (for the minimization needed to fit the nonlinear curve). With these restrictions and starting points in place, the examination yields the results in table 5.11 and figure 5.8.

**Table 5.11. Carrier Departures: Curve-Fitting Results**

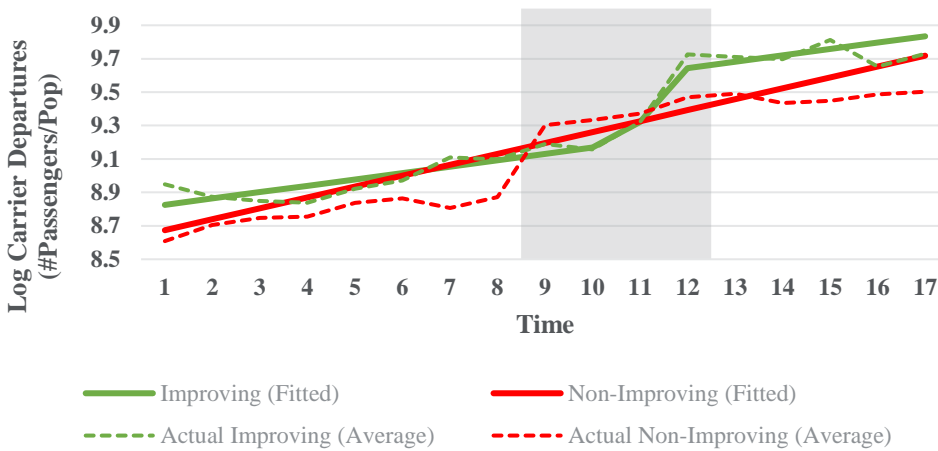
Indicator	Improving	Nonimproving
<b>Beta</b>	0.399	—
<b>Theta</b>	14.600	—
<b>Gamma</b>	11.061	—
<b>Constant (average)</b>	8.787	8.609
<b>Time trend</b>	0.038	0.065

Note: — = not available.

---

<sup>9</sup> There are some heteroskedasticity concerns for cell phone access, because all countries start at zero and likely do not exhibit much of a spread of varied values early on.

Figure 5.8. Level of Carrier Departures, Disaggregated by Countries with Improved PEFA Scores

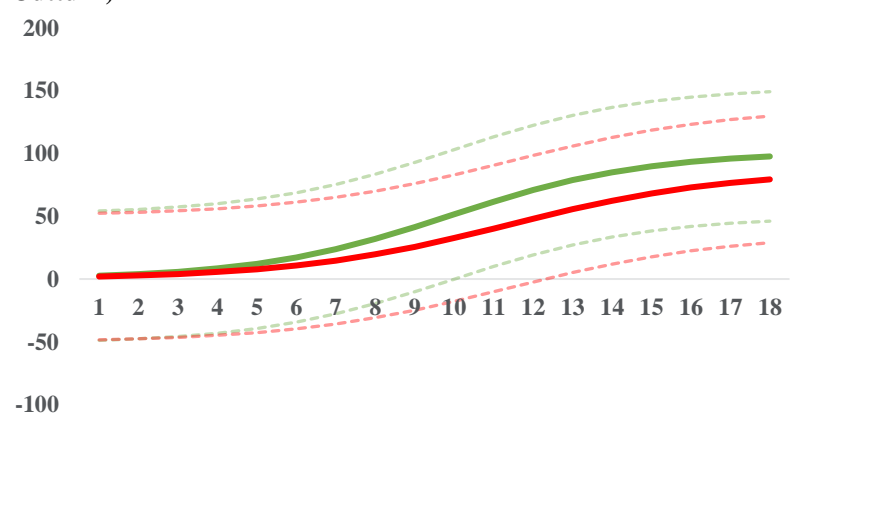


However, nonimproving countries do see some jump in levels prior to the test period. Even though the pattern of improving countries matches that of the theoretical model of improvement, this finding suggests that the effect has limited significance.

### Cell Phone Access

Cell phone access is an unusual case, as the entire growth in access to cell phones follows an S-shaped curve. In this case, the theory is adapted slightly, and the form of the S-shaped growth function across the time period is tested both for countries with improving and for countries without improving PFM (figure 5.9).

Figure 5.9. Cell Phone Access: Nonlinear Fitted Curve, Ordinary Intercept (Indicator: Composition of Outturn)



The results in table 5.12 are broadly expected: a higher beta if the reform improves outcomes that seem to be permanently above those of countries that do not undertake reform; a higher theta if the reforms speed outcomes, and a lower gamma if the improvement occurs earlier (as the

starting point for improvement—the higher this gets, the later the improvements began). The time trend is not included in this estimation, since it is a measure of infrastructure capacity with a cap (that is, cell phone penetration cannot exceed 100 percent).

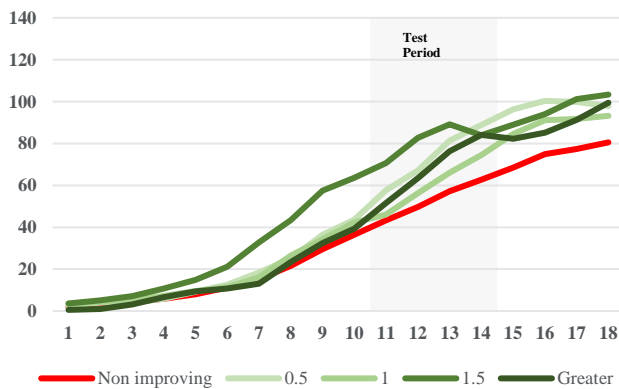
**Table 5.12. Cell Phone Access: Nonlinear Fitted-Curve Results**

Indicator	Common intercept		Fixed effects	
	Nonimproving	Improving	Nonimproving	Improving
Beta	86.7	101.4	73.1	79.5
Theta	0.36	0.41	0.50	0.58
Gamma	10.4	8.9	11.1	10.1
Constant	0.0	0.0	10.7	20.5

The data are now examined to see if a consistent pattern emerges. Figure 5.10 shows the change in cell phone access during an improvement in the composition of outturn scores. As the scale of improvement in the scores gets greater, there is some, but not a clear, shift in the form of the improvement. In this case, the scale of the increase in PEFA score does not necessarily yield far greater improvement.

However, the scope of greater growth is not universal and, for a sizable group, falls below the growth rates of the set of nonimproving countries (figure 5.11).

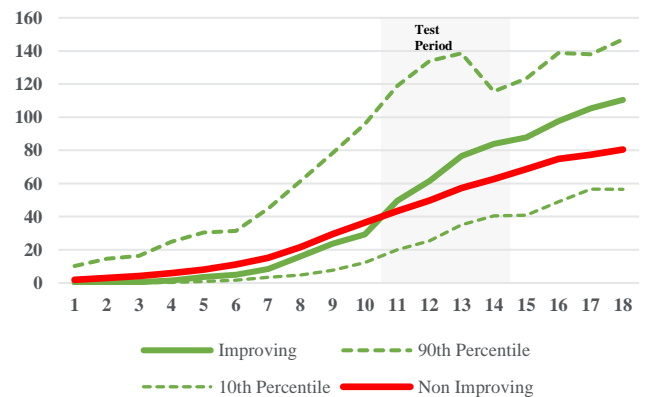
**Figure 5.10. Cell Phone Access during Improvements in Composition of Outturn, at Varied Scales of Improvement**



**Sample size:** 0.5 improvement (7); 1 improvement (7); 1.5 improvement (4) and greater (3).

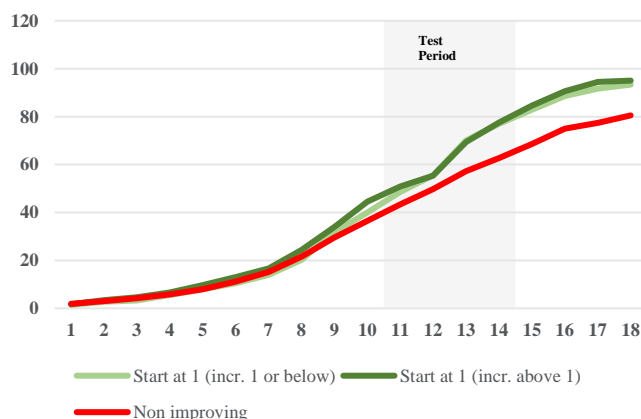
**Note:** Countries have been removed from improvement if they do not have data from at least period 3 to the end.

**Figure 5.11. Cell Phone Access during Improvements in Composition of Outturn, Percentiles**



Similarly, the starting point for this indicator (composition of outturn) does not change the scope of the improvement in the measure. This finding is telling, as it again seems to indicate that it is neither the point of start or end nor the scale of improvement that yields or is concurrent with faster improvement in the measure, but merely the fact that there has been an improvement in the ability to execute the budget according to plan (that is, in the composition of outturn compared to budget) (figure 5.12).

**Figure 5.12. Cell Phone Access during Improvements in Composition of Outturn, at Varied Starting Points**



**Sample size:** Start at 1 and improve 1 or below (5); start at 1 and improve above 1 (7).

**Note:** Countries have been removed from improvement if they do not have data from at least period 3 to the end

The composition indicator in table 5.13 speaks to the ability of government to cost realistically and execute effectively its budget component by component. That this element shows an improvement is in line with experience across countries, where capital budgets are cut most often following poor execution (see another demonstration of this in *Shocks and Spending* in this section).

**Table 5.13. Cell Phone Access: Improvements in Procurement**

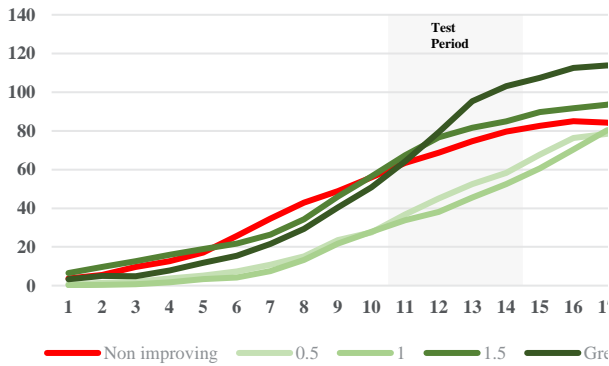
Indicator	Common intercept		Fixed effects	
	Nonimproving	Improving	Nonimproving	Improving
<b>Beta</b>	86.2	97.4	71.1	77.0
<b>Theta</b>	0.39	0.38	0.53	0.56
<b>Gamma</b>	9.1	9.9	10.2	10.8
<b>Constant</b>	0.0	0.2	13.4	16.8

Procurement reforms are potentially expected not only to improve outcomes (that is, beta), but also to improve the rate of improvement, shortening the time span required to expand infrastructure (in this case, cell phone access). Such an improvement would be equivalent to a rise in the theta value, which appears to be the case when cross-country starting points are taken into account.

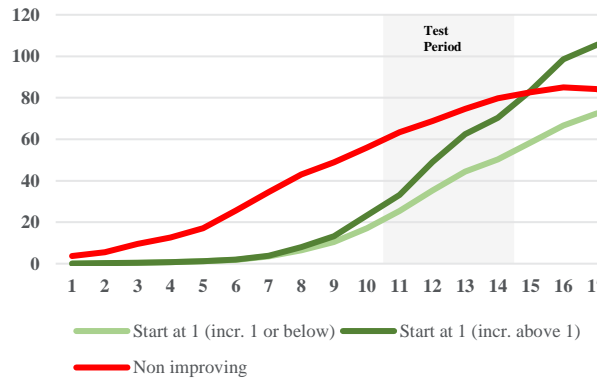
The improvement becomes potentially more distinct the higher the threshold (that is, a larger-scale improvement seems to correspond to a larger jump in outcomes).

Similarly, the starting point seems to define the speed at which an improvement in scores is evident (figures 5.13 and 5.14). This result makes more intuitive sense than it does for composition. In this case, differing levels of procurement capacity will influence the speed of the ability to improve infrastructure, and the implication is that higher-level improvements in indicators (or improvements from a higher starting base) mean more substantial improvements in measures.

**Figure 5.13. Cell Phone Access during Improvements in Procurement, at Varied Scales of Improvement**



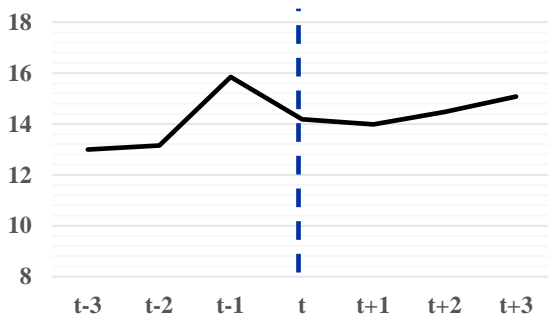
**Figure 5.14. Cell Phone Access during Improvements in Procurement, at Varied Starting Points**



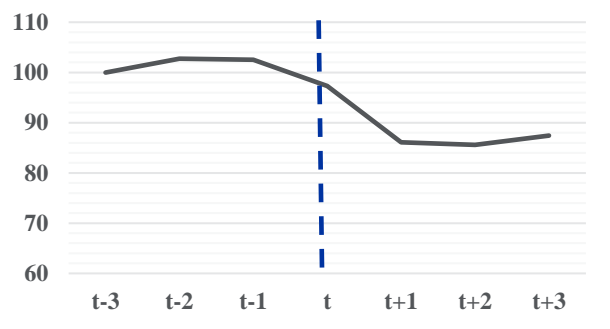
## Shocks and Spending

Some of the presumptions made about the performance of PFM reforms rely on the assumption that governments will typically cut expenditures on capital first in the event of a revenue shortfall. This assumption is tested using both the Government Finance Statistics (GFS) data and the gross fixed capital formation (GFCF) estimates in an event study framework. The event in this case uses International Monetary Fund (IMF) projections, and actuals, of growth from 2000 onward. An event is defined as a period during which the growth forecast of a country (that is, the estimated value for growth from the World Economic Outlook data release one year prior) underperforms by at least 6 percent. This approach assumes that an unexpected drop in growth, as evidenced by a significant one-year forecast error, will lead to a need to reduce expenditures. If this drop in growth is concurrent with a drop in capital expenditure, it would provide some evidence toward the assertion. Results are presented in figures 5.15–5.18.<sup>10</sup>

**Figure 5.15. WDI Data: Actuals (Median Average)**

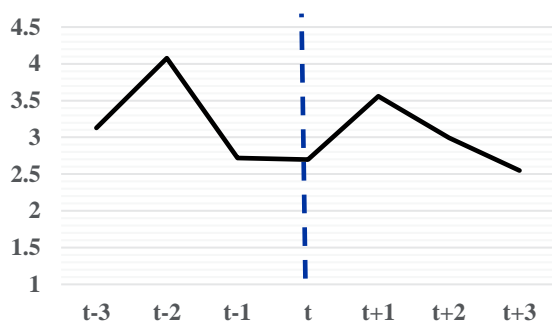


**Figure 5.16. WDI Data: Index (Median Average)**

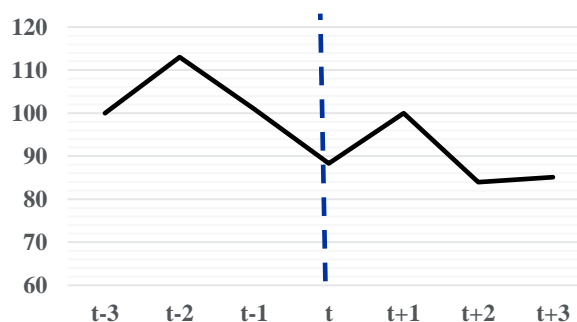


<sup>10</sup> Figures 5.15 and 5.17 present the trends of the WDI and GFS data as presented originally in the database, whereas the 5.16 and 5.18 present the trends of the values, when the data is indexed. The index value is the average of values across the period.

**Figure 5.17. GFS Data: Actuals (Median Average)**



**Figure 5.18. GFS Data: Index (Median Average)**



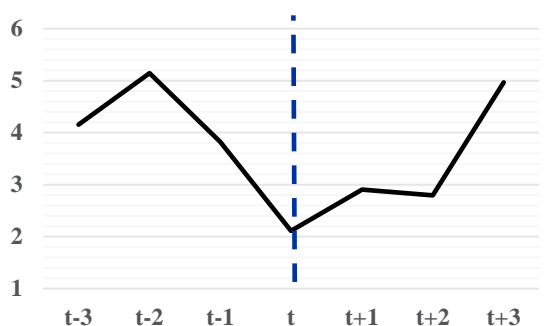
This approach does not present a clear relationship; however, less financing-constrained countries are likely to expand spending on capital (or at least not reduce it) in order to offset the downturn. For this reason, the set of event countries is limited to those that actually cut overall expenditure (that is, were more likely liquidity constrained), and then overall expenditure cuts are examined relative to the drop in capital spending (table 5.14 and figures 5.19 and 5.20).

**Table 5.14. Overall Expenditure Cuts Relative to the Drop in Capital Expenditure**

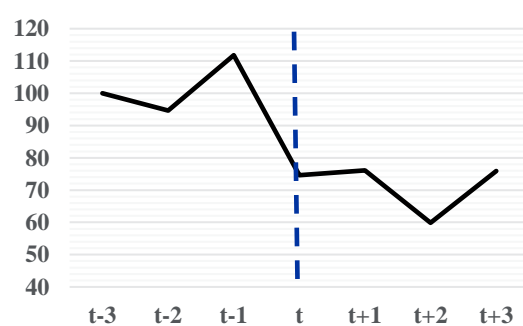
*Median average fall across the set of countries*

Indicator	GFS data	GFCF data
Capital fall (%)	16.3	28.1
Overall fall (%)	11.7	10.4

**Figure 5.19. GFS Data: Actuals (Median Average)**



**Figure 5.20. GFS Data: Index (Median Average)**



World Development Indicators data show a similar result and so are not presented here.

These results seem to indicate that countries do, in line with general perceptions, cut capital expenditure in the event of shocks to resources and that capital expenditure is more substantial than the fall in overall expenditure.

## 6. Conclusions and Recommendations

---

This study has sought to determine whether improving public financial management (PFM) performance (as measured by PEFA scores) is linked to subsequent improvements in public investment performance. Using an event study framework, a set of countries where PFM scores improved was compared to a set of countries where PFM systems did not improve or got weaker.

### Conclusions

In brief, *the analysis finds evidence of a positive link between improved PFM performance and improved public investment performance*. Many countries with improving PEFA scores for the selected indicators see subsequent improvements across a range of public investment and infrastructure-related outcomes. While those improvements in outcome are often greater than those seen in comparator countries (whose PEFA scores did not improve), few of these tests show statistically significant results. Although this finding limits the conclusions drawn, the results of this study nevertheless give rise to conclusions that can shape future investigations.

In particular, the study finds that *there is a stronger link between quantifiable improvements in PFM performance (for example, adherence to approved budgets) and public investment and infrastructure-related outcomes than there is between qualitative PFM improvements (for example, presence of commitment controls) and those same outcomes*.

The three qualitative PEFA indicators chosen for this study indicate the presence or absence of particular PFM structures or processes (Does a country have a multiyear framework? Does it have certain procurement and commitment processes in place?), whereas the two quantitative indicators reflect actual PFM performance (Does the government execute its budget in line with planned expenditure?). The tests conducted here show a stronger association between improvements in the quantitative indicators of performance and improvements in the level of public investment and infrastructure-related outcomes (Is there more air traffic? Have basic sanitation services improved?). However, within the time frames examined, qualitative indicators merely showing the presence or absence of structures or processes—such as those relating to multiyear budgeting or procurement processes—appear to have limited relationships with improvements in those outcomes.

These findings may result from the likelihood that quantitative indicators of PFM performance reflect the cumulative impact of multiple reforms and capacity-strengthening efforts—improvements in budget planning *and* procurement processes *and* internal controls—while the qualitative indicators are defined more narrowly. Additionally, the qualitative indicators may only signal the appearance of PFM improvements (that is, the *de jure* presence or absence of process improvements), while the quantitative indicators demonstrate actual (that is, *de facto*) changes in outputs and performance. *The results therefore highlight that the measurement of PEFA performance captures both de facto and de jure impacts*. This finding has important implications for the implementation of PFM programs as well as for the examination of PFM system strength.

This conclusion does not suggest that the implementation of process improvements captured by qualitative indicators is not important. It is possible, for instance, that implementing these “qualitative” process improvements could serve as a necessary condition for achieving a better outcome on the quantitative indicators. Instead, the findings highlight the importance of

considering the preconditions for successfully implementing qualitative reform measures and the quality of implementation of these reforms, to avoid the risks associated with “isomorphic mimicry.”

The one outlier is PI-20, a qualitative indicator focused on the implementation of effective controls. This indicator shows statistical significance against various infrastructure-related outcomes. It looks primarily at whether commitment controls exist, whether they are cost-effective, and whether they are adhered to. Commitment controls are very influential in achieving, and much more directly tied to, a particular immediate outcome—that is, the degree to which the government makes payments (most likely measured by an immediate outcome like the accumulation of arrears). While other factors can influence this immediate outcome (for example, access to financing, overdraft facilities, T-bills), the ability to conduct adequate cash planning and control expenditures (through a viable and efficient commitment control process) is fundamental. Commitment controls are equally essential in countries with easier access to liquidity. Measuring controls therefore is akin to measuring an immediate outcome rather than the occurrence of a reform or process change. Internal controls, unlike the indicators of both composition and aggregate outturn, show a scale of change similar to that of the other direct measures of PFM improvement.

These results also point to the possibility that the quality of implementation of reforms and processes captured by qualitative indicators may be measured through quantitative indicators. Understanding the links between qualitative and quantitative indicators is important to screen out instances of isomorphic mimicry that could be overlooked if the focus is solely on qualitative indicators. Further analysis is recommended to understand this relationship better, as it falls beyond the scope of this report.

The study has two key implications:

- First, individual reforms should be linked to broader intermediate outcomes measured quantitatively (for example, improved budget preparation should lead to better budget execution, as measured by the variance between budget and outturn). While the immediate outcomes likely indicate improvements in underlying PFM systems, they are easier to link to public investment and infrastructure-related outcomes. This formulation would be easier to assess, allowing evaluators to examine not only whether the work was completed and outputs were delivered, but also whether immediate outcomes improved (and whether other complementary or countervailing reforms could be listed at the start as potential risks, to explain why immediate outcomes did not improve).
- Second, these intermediate outcomes should be the focus of future analysis linking reforms to higher-level outcomes (like public investment performance outcomes), as individual components of reforms appear to have limited measurable effect but can significantly influence the quality of public investment outcomes in the medium to long term (also suggested in this report).

In short, while not unique, this study highlights the importance of measuring reform effort and assessing its performance vis-à-vis public investment performance outcomes. This analysis also contributes to the existing literature by highlighting that country context plays an important part in determining the influence of PEFA reforms on PIM. While the results of this analysis do not provide a definitive answer to the research question posed—Do countries whose PFM performance is improving see greater subsequent improvements in public investment



performance outcomes than countries whose PFM performance is not improving?—this study provides the evidence that some PEFA indicators are influential in shaping PIM outcomes and informs future efforts exploring this question.

## Recommendations for programming

Based on these results, the following recommendations are offered for consideration in future PFM reform programming:

- Areas of PFM performance should ideally be linked to quantitative, intermediate measures that they can reasonably be evidenced to affect and then have to be linked to higher-level outcomes (for example, outcomes related to the provision of social and economic infrastructure) to assess the effects of PFM performance on higher-level outcomes.
- PFM-strengthening programs should establish expected, clear, quantifiable, and immediate outcomes from the reforms being undertaken (for example, better execution of the budget, reduced transfers and virements, reduced processing time for commitments, reduced arrears). These expected outcomes should then serve as a benchmark to determine whether a reform should be undertaken.
- PFM support programs should identify complementariness and prerequisites—that is, reform A will not influence the immediate outcome without reform B. For example, introducing multiyear budgeting may indeed, *ceteris paribus*, be a good measure for potentially improving the effectiveness of government spending and, eventually, development outcomes (for example, livelihoods). If, however, the government has no effective expenditure controls in place and does not adhere to budgets, then the improved budget planning and preparation could be rendered meaningless and unlikely to affect expenditure outcomes.
- Selection of further PFM-strengthening initiatives should be based on evidence of their impacts on specific PFM performance outcomes (for example, adherence to approved budgets) and on the links between those targeted performance outcomes and desired outcomes. Doing so is likely to favor reforms with clear goals and measurability (for example, the introduction of commitment controls) over reforms with more difficult interpretation and less clear immediate outcomes (for example, the introduction of multiyear budgets).

## Areas for Future Research

While this study contributes to the base of evidence examining the links between PFM system strengthening and public investment performance, various areas for further research remain. To understand the links between PFM performance and public investment performance, further research considering the following is recommended:

- *The relationship between quantitative versus qualitative indicators.* This study finds a stronger link between quantifiable PFM performance improvements and public

investment performance outcomes than between qualitative PFM improvements and those same outcomes. However, it is possible that the qualitative indicators could be necessary conditions for a better outcome of the quantitative indicators. Further research could explore whether improvements in qualitative indicators lead to an improvement in intermediate outcomes. Such research could be done by building off the current methodology and introducing a multiplicative binary variable in regression analysis (using, for instance, a logit or probit model).

- *Complementarity of PFM reforms.* Do reforms offer complementary improvements? In theory, reforms occurring at the same time (an improvement in procurement capacity and an improvement in controls) could yield better execution and greater outcomes, but only if both occur. An area of further study could be to focus on whether and which reforms offer complementary improvements and whether the timing of the implementation of these reforms matters (Should they be implemented at the same time? Do certain reforms serve as a prerequisite?).
- *Expanding the scope of indicators.* In particular, PI-11 (Orderliness and participation in the annual budget process), PI-4 (Stock and monitoring of capital expenditure arrears), PI-6 (Comprehensiveness of information included in the budget documentation), PI-16 (Predictability in the availability of funds for commitment of expenditures), PI-23 (Availability of information on resources received by service delivery units), PI-24 (Quality and timeliness of in-year budget reports) could be included to examine other indicators of PFM performance that may affect public investment management and performance.
- *Disaggregation of data and country-level analysis.* This study uses global data at a highly aggregate level. Future studies may include an additional analysis by (1) disaggregating the data by country income level to create an added layer of cross-country analysis and (2) analyzing specific country cases. Country-specific cases would shed light on specific dynamics and factors that may be shaping the links between the performance of a country's PFM system and the quality of the provision of certain social and economic infrastructure. Specific cases could also focus on characteristics that are beyond the scope of this study—such as the potential effects of direct budget support from donors as well as the effects of the presence of off-budget activities on PFM system strengthening and public investment performance.
- *Other methods for enumerating PEFA scores.* For certain types of improvement in PEFA scores, a more notable improvement is observed in some measures when starting from a higher PEFA score. This result could indicate that the approach is not measuring like for like (that is, a shift from B score to A is not the same as a shift from C score to B). While this does not create substantial issues for the event study framework, as improvements have been grouped, future research could attempt to draw more precise or numeric relationships between changes in PEFA scores and other indicators in order to examine further nuances of how the level of PFM system's strength shapes outcomes.
- *Further exploration of curve fitting.* The structural form of the improvement of infrastructure-related outcomes—that is, how and the rate at which outcomes are improving—remains unclear. One model is proposed based on observations from the

event studies: a sigmoid or S-shaped curve that indicates an improvement of an infrastructure-related outcome during or just after a period in which PFM systems also improve, before returning to trend. This S-shaped curve has utility in explaining how improvements occur, but it is also likely that PFM reforms could permanently shift the rate of improvement in outcomes for certain types of measures. Additional data across longer time horizons will likely be necessary for future research to make such a determination.

- *A policy-oriented study.* A study addressing policy implications for practitioners and development partners would be welcomed, but was beyond the scope of this research-focused study.

## Appendix A. Considered Measures

This appendix provides an overview of the original measures reviewed, which were then narrowed down to a smaller set to remove those with only very limited coverage.

Variable	Source	Coverage	Unit	Data concerns
Access to electricity	World Bank	For most countries, data begin in 1990 and go through 2017	% of population	Some lower- and middle-income countries do not have data until 2005 or later
Firms experiencing electrical outages	World Bank	Data begin in 2006 but only cover 28 countries; most countries have one to three data points between 2006 and 2019	% of firms	No major concerns
Mobile cellular subscriptions	World Bank	Data begin in 1982 and cover basically every country from 1996 to 2018	Cellular subscription per 100 people	Many countries, mostly in Latin America and the Caribbean, peaked above 100% in the early to mid-2010s and then settled back down to around 100. This may be related to factors that are not factored into the methodology (for example, tax laws, anti-money-laundering efforts in the 2010s, hurricanes)
Mobile network coverage	World Economic Forum (WEF) Global Information Technology Report, via World Bank	Data are only for 2012 to 2016 and are missing some countries	% of population	No major concerns besides incomplete coverage
Hospital beds	World Health Organization (WHO) data, supplemented by country data, via World Bank	Data cover almost every country from 2000 to 2015	Beds per 1,000 people	There is no global target for the number of hospital beds per country. The level of inpatient services is subject to country-specific factors, like burden of disease or demographic issues. The absence of a global target may make it difficult to derive any meaningful results, given the scope and methodology of this study
General government investment	International Monetary Fund (IMF)	Data begin in 1960 and go consistently through 2013. Data are missing for many Latin American and Caribbean countries and other small and island nations	Constant 2005 dollars	Since the data end right after many countries' fiscal responses to the Great Recession, many austerity and stimulus policies remain and have not had time to return to prerecession norms
Capital expenditure	World Bank	Data begin in 1998 and go through 2018 but do not	% of total expenditure in public institutions	This indicator may not be appropriate, since it refers to public educational institutions

Variable	Source	Coverage	Unit	Data concerns
		cover every country, with large gaps between years in some countries		
Source and uses of cash: net cash outflow on investment in nonfinancial assets	IMF	Data begin in 1990 in Maldives but do not cover most countries until about 2008 through 2016, 2017, or 2018, depending on the country	Unclear	No major concerns besides incomplete coverage
Transport and composite index	African Development Bank (AfDB), part of the African Infrastructure Development Index (AIDI)	2003–19	Composite index calculated as a weighted average of indicators for each component that comprises more than one indicator. Two components: (1) total paved roads (kilometers per 10,000 inhabitants) and (2) total road network in kilometers (per square kilometer of exploitable land area) <sup>a</sup>	No major concerns besides incomplete coverage; only covers countries in Sub-Saharan Africa
Rail lines	World Bank	1995–2018	Total route kilometers	No major concerns besides incomplete coverage
Railways, goods transported	World Bank	1996–2018	Million metric tons times kilometers traveled	No major concerns
Quality of road infrastructure	World Bank, originally from WEF	2007–17	Score 1–7. Answers: In your country, how would you assess the quality of roads? 1 is extremely underdeveloped	In addition to data coverage issues, this indicator may not be the most appropriate, since it draws from a survey question posed to business executives (WEF's Executive Opinion Survey) and is therefore subjective
Container port traffic	World Bank, originally from United Nations Conference on Trade and Development (UNCTAD)	2000–18	TEUs (twenty-foot equivalent units), a standard-size container	No major concerns
Logistics Performance Index (LPI)	World Bank	2007, 2010, 2012, 2014, 2016, 2018	Score 1–5 (5 is the highest). LPI 2018 ranks countries on six dimensions of trade, including customs performance, infrastructure quality, and timeliness of shipments, based on survey data	Insufficient data
Port infrastructure quality	World Bank, originally from WEF	2007–17	Score 1–7. Answers: In your country, how would you assess the quality of seaports? (For landlocked countries: How accessible are seaport facilities?) 1 is underdeveloped, among the worst in the world	In addition to data coverage issues, this indicator may not be the most appropriate, since it draws from a subjective survey question posed to business executives (WEF's Executive Opinion Survey)

Variable	Source	Coverage	Unit	Data concerns
Propensity to travel	Calculations based on World Bank data (air transport, passengers carried, and population indicators)	1970–2018	Number of passengers per population of the country	No major concerns
Quality of air transport	WEF via World Bank	2007–17	Score 1–7. Answers: In your country, how would you assess the quality of air transport infrastructure? 1 is extremely underdeveloped	In addition to data coverage issues, this indicator may not be the most appropriate, since it draws from a subjective survey question posed to business executives (WEF’s Executive Opinion Survey)
Air transport, registered carrier departures worldwide	International Civil Aviation Organization via World Bank	2000–18	Number of carrier departures, which are domestic takeoffs and takeoffs abroad of air carriers registered in the country	No major concerns
People using at least basic drinking water services	WHO/United Nations Children’s Fund (UNICEF) Joint Monitoring Program (JMP) via World Bank	2000–17	% of population	No major concerns
People using safely managed drinking water services	WHO/UNICEF JMP via World Bank	2000–17	% of population	Incomplete data; data are missing for half of the countries
Water and Sanitation Composite Index	AfDB, part of the AIDI	2003–19	Composite index calculated as a weighted average of indicators for each component that comprises more than one indicator. Two components: (1) improved water source (% of population with access); (2) improved sanitation facilities (% of population with access)	No major concerns besides incomplete coverage; only covers countries in Sub-Saharan Africa
Agricultural irrigated land	Food and Agriculture Organization (FAO) via World Bank	2001–16	Percentage of total agricultural land <sup>b</sup>	No major concerns
Total amount of municipal waste collected	United Nations data	1990–2016	1,000 tons	Incomplete data: only has consistent data for 1995–2015; even then, data are missing for almost half of countries
Infrastructure (part of the LPI)	World Bank	2007, 2010, 2012, 2014, 2016, 2018	Score 1–5 (5 is the highest); quality of trade and transport infrastructure	Insufficient data
Quality of overall infrastructure	WEF via World Bank	2007–17	Score 1–7. Answers: How would you assess general infrastructure (for example, transport, telephony, and energy) in your country? 1 is extremely underdeveloped	Insufficient data; in addition to data coverage issues, this indicator may not be the most appropriate, since it draws from a subjective survey question posed to business executives (WEF’s Executive Opinion Survey)

Variable	Source	Coverage	Unit	Data concerns
African Infrastructure Development Index (AIDI)	AfDB	2003–20	Composite index calculated as a weighted average of indicators for each component comprising more than one indicator. AIDI is composed of four indexes: transport composite index, electricity composite index, information and communication technology composite index, and water and sanitation composite index	No major concerns besides incomplete coverage

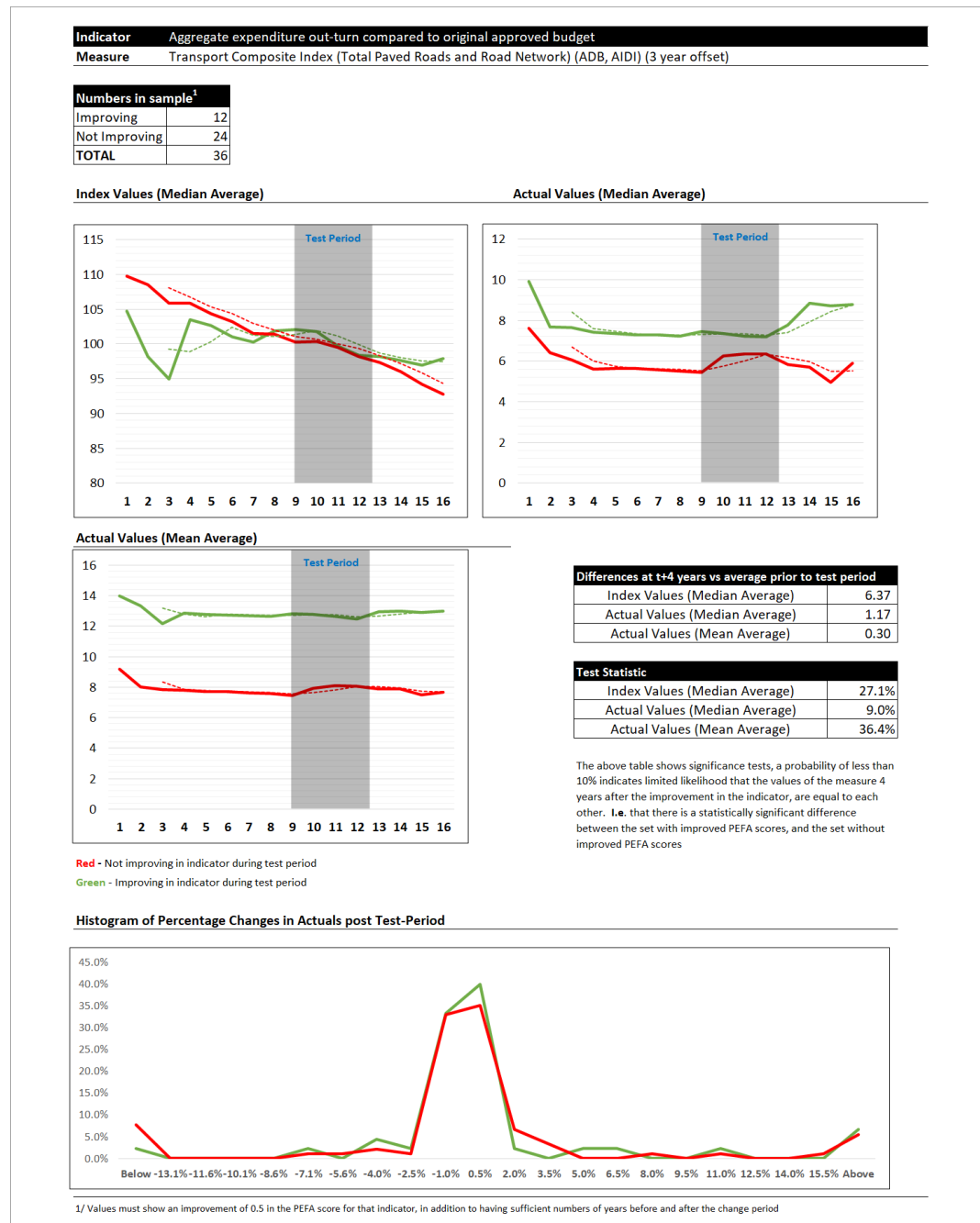
- a. Exploitable land area is the total surface area of a country minus the surface area of deserts, forests, mountains, and other inaccessible areas (AfDB).
- b. Agricultural irrigated land refers to agricultural areas purposely provided with water, including land irrigated by controlled flooding (FAO).

## Appendix B. Results Sheets

**NOTE: DATA PRIOR TO PERIOD 4 SHOULD BE TAKEN WITH CAUTION, AS SOME SERIES ARE LIMITED BEFORE THIS POINT**

### Aggregate Expenditure Outturn Compared to Original Approved Budget

#### Transport Composite Index



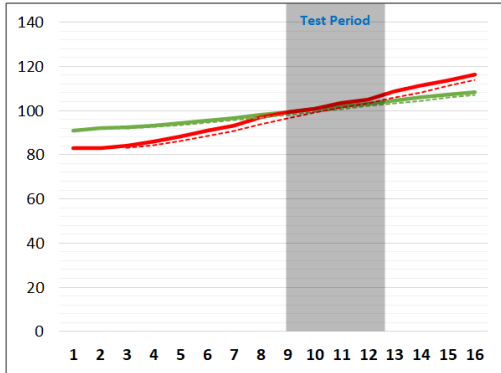


## People Using Safely Managed Sanitation Services

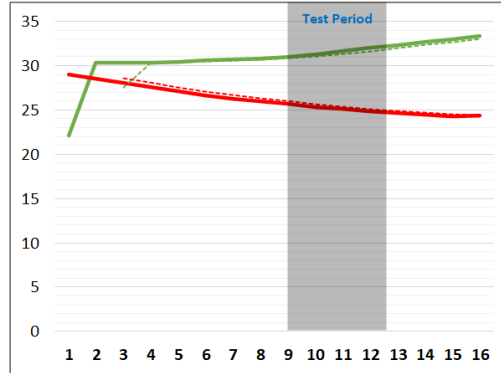
**Indicator** Aggregate expenditure out-turn compared to original approved budget  
**Measure** People using safely managed sanitation services (% of population) (WHO/UNICEF)

Numbers in sample <sup>1</sup>	
Improving	4
Not Improving	9
<b>TOTAL</b>	<b>13</b>

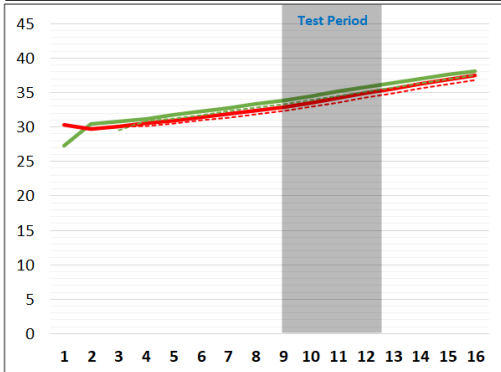
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



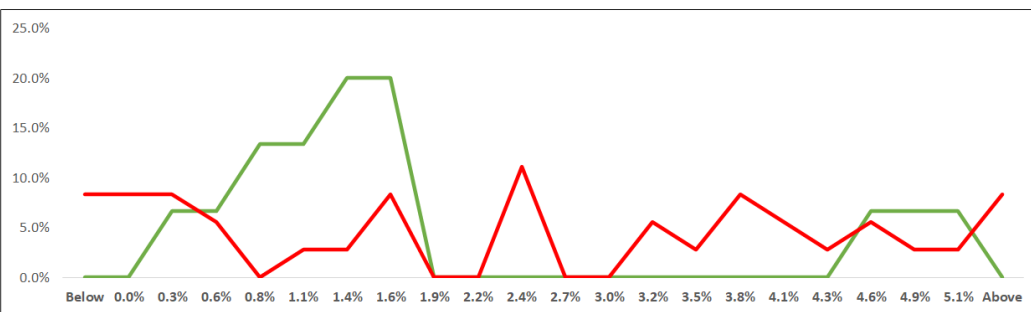
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	(11.74)
Actual Values (Median Average)	4.83
Actual Values (Mean Average)	(0.22)

Test Statistic	
Index Values (Median Average)	-
Actual Values (Median Average)	6.8%
Actual Values (Mean Average)	-

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



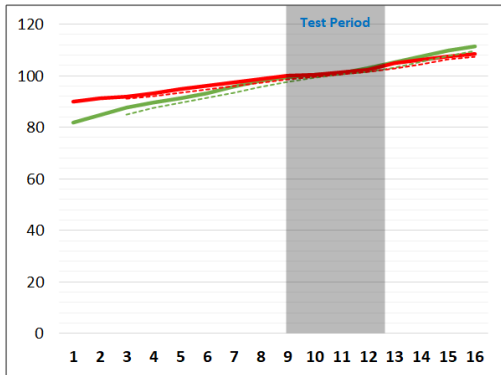
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## People Using At Least Basic Sanitation Services

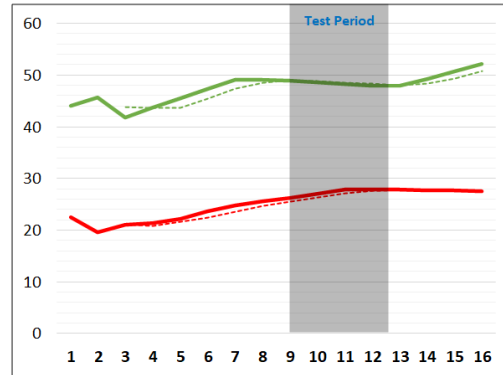
**Indicator** Aggregate expenditure out-turn compared to original approved budget  
**Measure** People using at least basic sanitation services (% of population) (WHO/UNICEF)

Numbers in sample <sup>1</sup>	
Improving	28
Not Improving	36
<b>TOTAL</b>	<b>64</b>

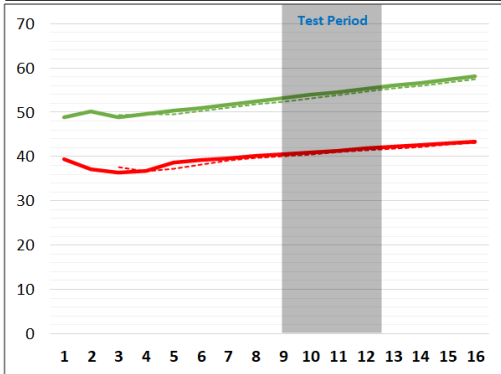
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



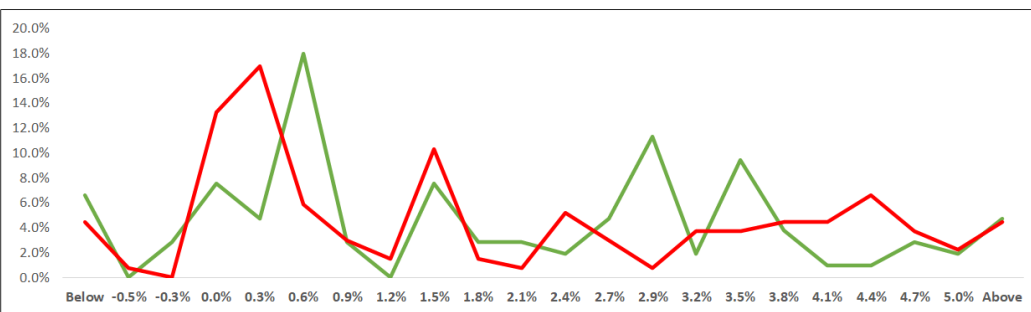
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	5.08
Actual Values (Median Average)	0.88
Actual Values (Mean Average)	2.72

Test Statistic	
Index Values (Median Average)	20.6%
Actual Values (Median Average)	31.5%
Actual Values (Mean Average)	7.0%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



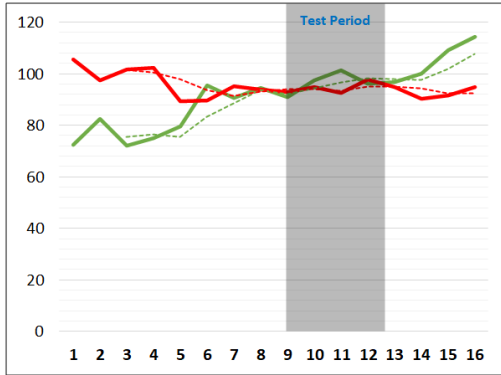
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## Estimated Gross Fixed Capital Formation

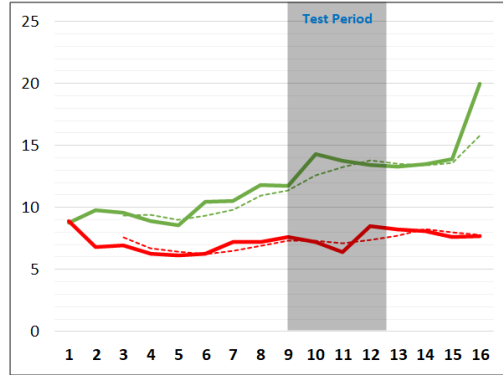
**Indicator** Aggregate expenditure out-turn compared to original approved budget  
**Measure** Estimated Public Gross Fixed Capital Formation (% GDP) (Calculated from World Bank data)

Numbers in sample <sup>1</sup>	
Improving	22
Not Improving	37
<b>TOTAL</b>	<b>59</b>

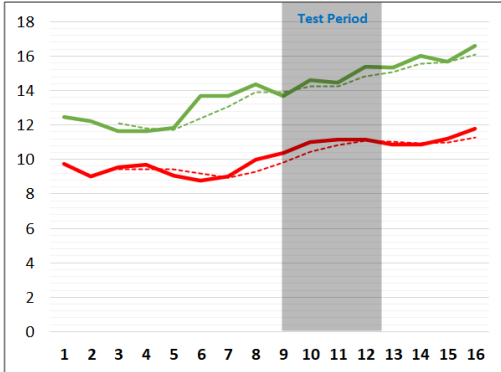
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



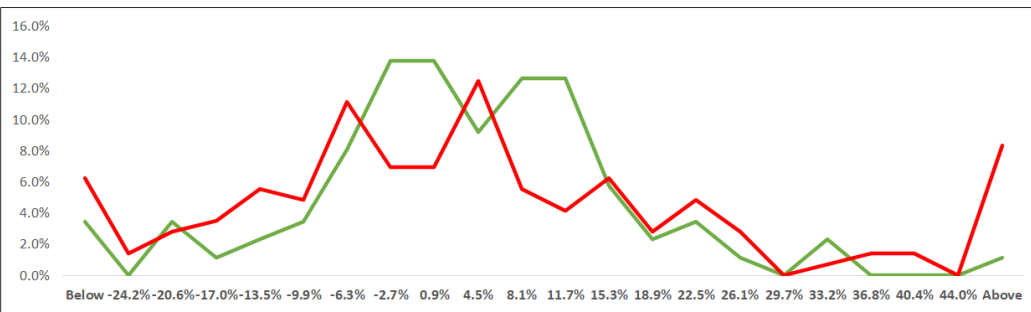
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	21.31
Actual Values (Median Average)	8.68
Actual Values (Mean Average)	0.64

Test Statistic	
Index Values (Median Average)	12.7%
Actual Values (Median Average)	0.0%
Actual Values (Mean Average)	39.3%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



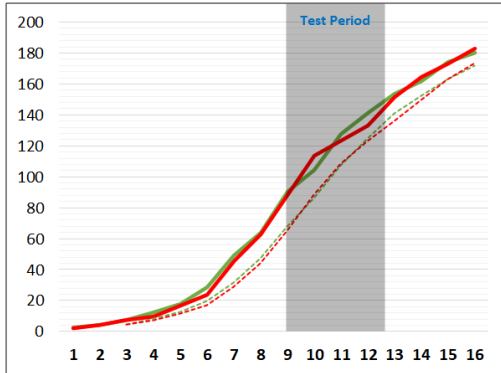
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## Mobile Cellular Subscriptions

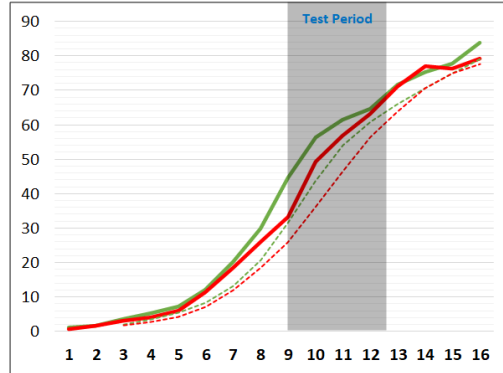
**Indicator** Aggregate expenditure out-turn compared to original approved budget  
**Measure** Mobile cellular subscriptions (per 100 people) (CEL.SETS) (World Bank)

Numbers in sample <sup>1</sup>	
Improving	28
Not Improving	40
<b>TOTAL</b>	<b>68</b>

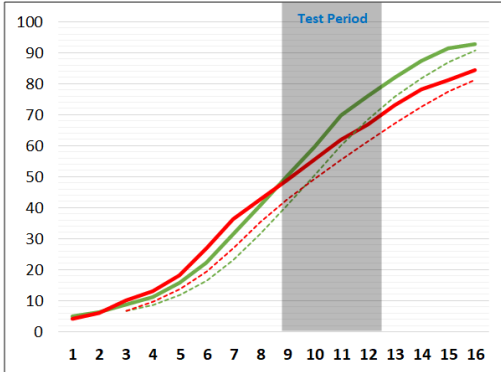
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



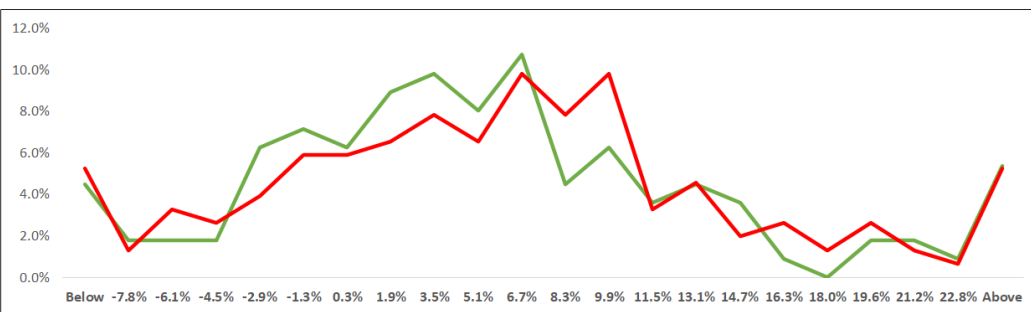
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	(5.07)
Actual Values (Median Average)	2.66
Actual Values (Mean Average)	11.81

Test Statistic	
Index Values (Median Average)	-
Actual Values (Median Average)	32.6%
Actual Values (Mean Average)	2.4%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



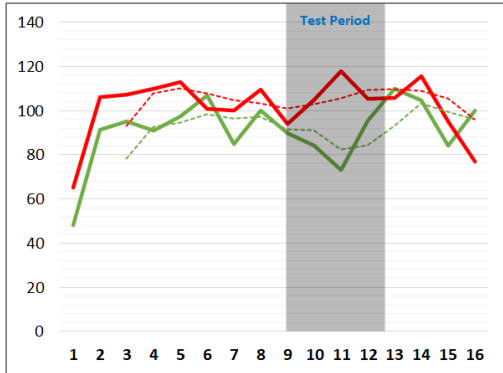
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## Net Acquisition of Non-Financial Assets

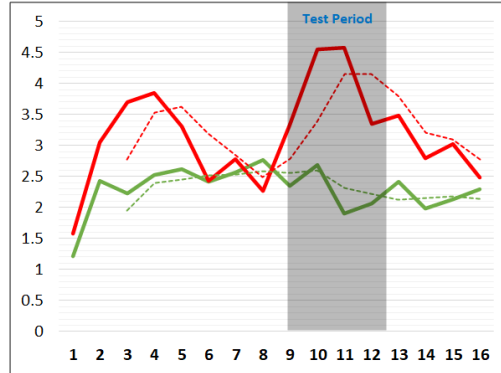
**Indicator** Aggregate expenditure out-turn compared to original approved budget  
**Measure** Net acquisition of non financial assets (% of GDP) (GC.AST.TOTL.GD.ZS) (World Bank)

Numbers in sample <sup>1</sup>	
Improving	11
Not Improving	11
<b>TOTAL</b>	<b>22</b>

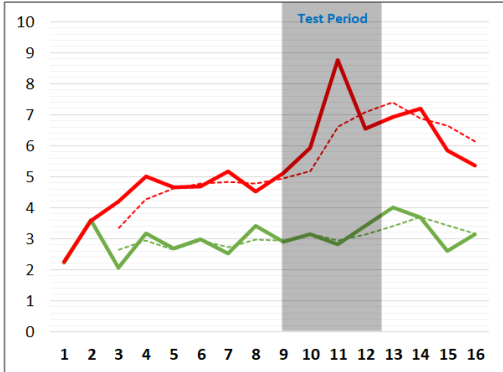
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



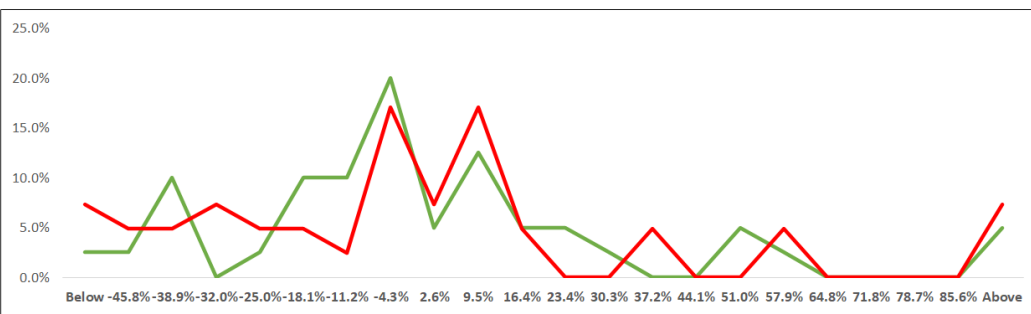
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	31.68
Actual Values (Median Average)	(0.09)
Actual Values (Mean Average)	(0.35)

Test Statistic	
Index Values (Median Average)	6.1%
Actual Values (Median Average)	-
Actual Values (Mean Average)	-

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

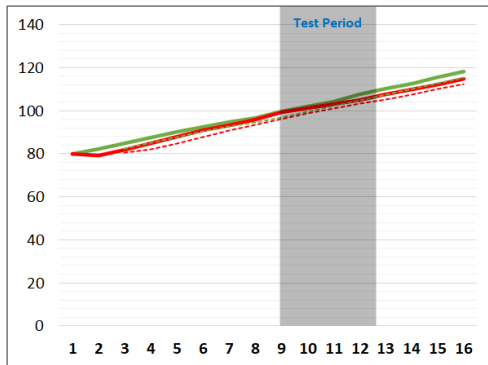
# Composition of Expenditure Outturn Compared to Original Approved Budget

## Water and Sanitation Composite Index

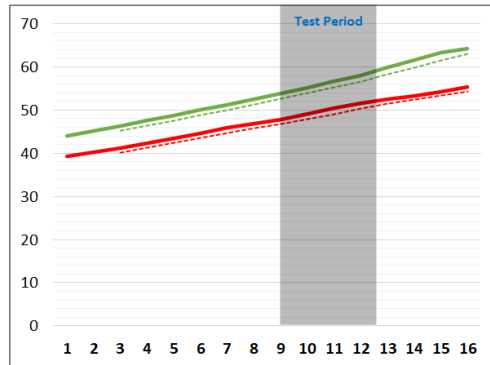
<b>Indicator</b>	Composition of expenditure out-turn compared to original approved budget
<b>Measure</b>	Water and Sanitation Composite Index (WSS) (ADB, AIDI) (3 year offset)

Numbers in sample <sup>1</sup>	
Improving	12
Not Improving	24
<b>TOTAL</b>	<b>36</b>

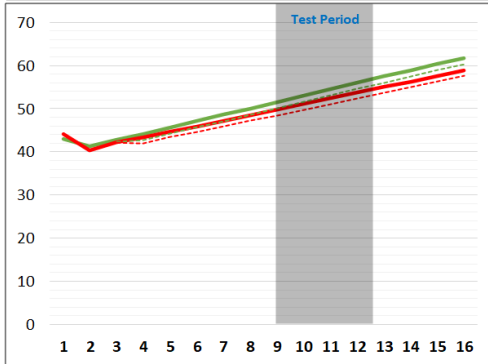
Index Values (Median Average)



Actual Values (Median Average)



Actual Values (Mean Average)



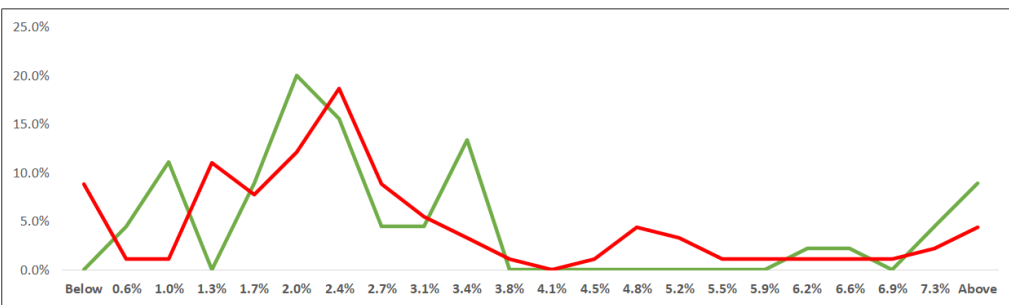
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	2.04
Actual Values (Median Average)	3.47
Actual Values (Mean Average)	1.67

Test Statistic	
Index Values (Median Average)	42.4%
Actual Values (Median Average)	6.0%
Actual Values (Mean Average)	22.4%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

Red - Not improving in indicator during test period  
Green - Improving in indicator during test period

Histogram of Percentage Changes in Actuals post Test-Period



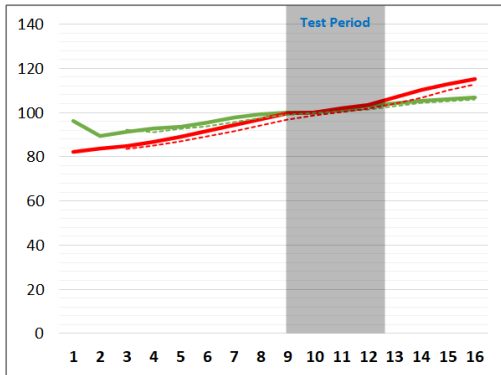
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## People Using Safely Managed Drinking Water Services

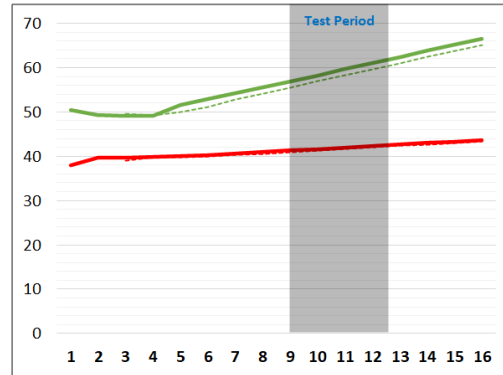
**Indicator** Composition of expenditure out-turn compared to original approved budget  
**Measure** People using safely managed drinking water services (% of population) (WHO/UNICEF)

Numbers in sample <sup>1</sup>	
Improving	10
Not Improving	14
<b>TOTAL</b>	<b>24</b>

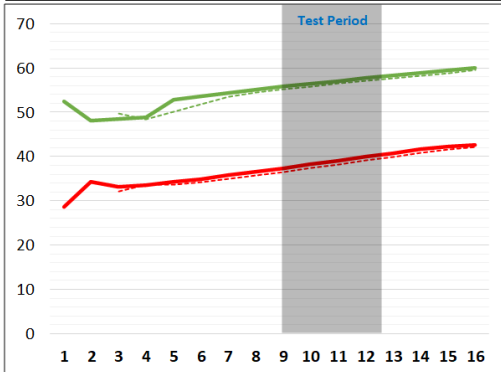
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



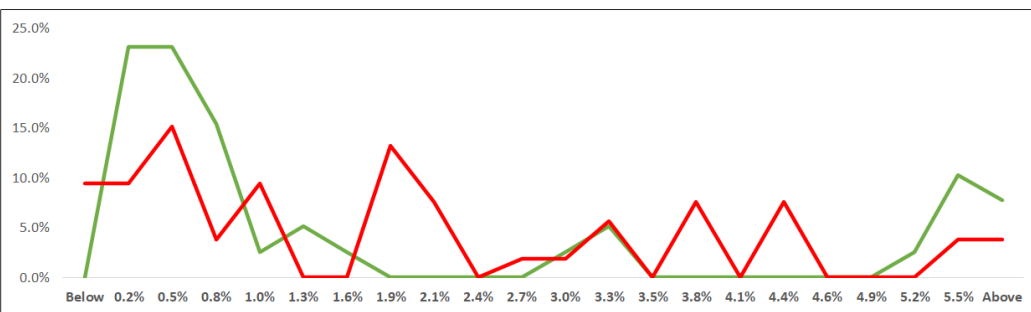
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	(12.01)
Actual Values (Median Average)	9.83
Actual Values (Mean Average)	(1.21)

Test Statistic	
Index Values (Median Average)	-
Actual Values (Median Average)	0.1%
Actual Values (Mean Average)	-

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



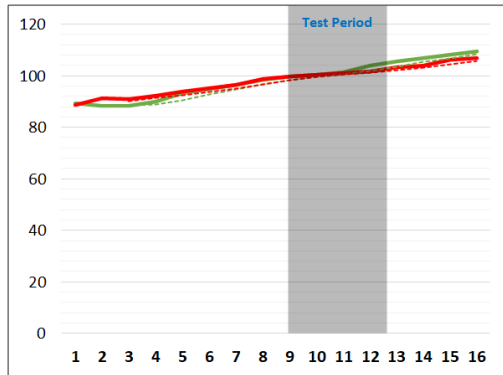
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## People Using At Least Basic Sanitation Services

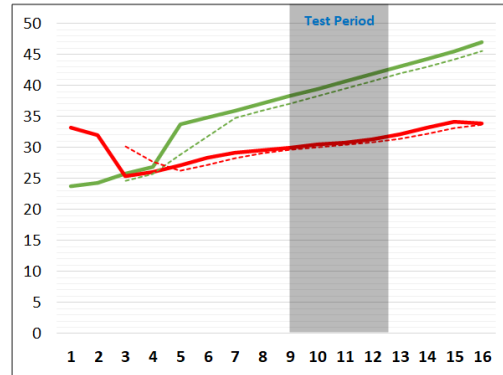
<b>Indicator</b>	Composition of expenditure out-turn compared to original approved budget
<b>Measure</b>	People using at least basic sanitation services (% of population) (WHO/UNICEF)

Numbers in sample <sup>1</sup>	
Improving	24
Not Improving	40
<b>TOTAL</b>	<b>64</b>

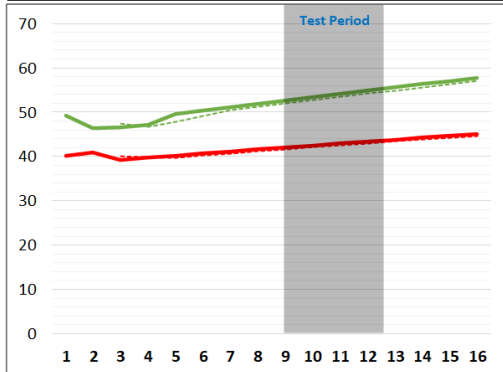
Index Values (Median Average)



Actual Values (Median Average)



Actual Values (Mean Average)



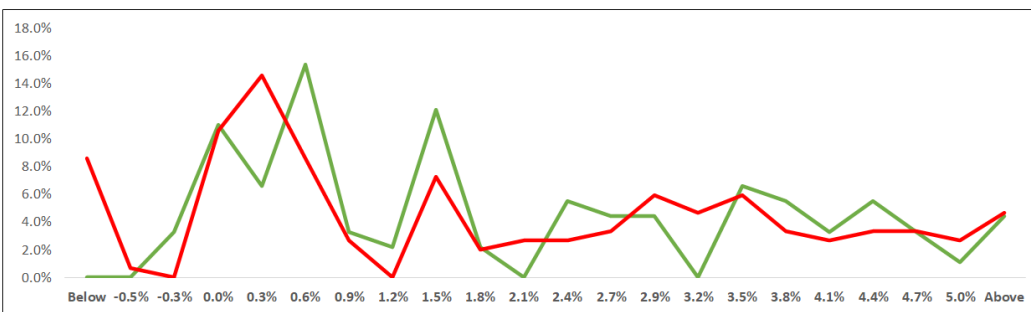
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	2.99
Actual Values (Median Average)	6.40
Actual Values (Mean Average)	2.70

Test Statistic	
Index Values (Median Average)	30.2%
Actual Values (Median Average)	0.0%
Actual Values (Mean Average)	4.8%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

Red - Not improving in indicator during test period  
Green - Improving in indicator during test period

Histogram of Percentage Changes in Actuals post Test-Period



1/ Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

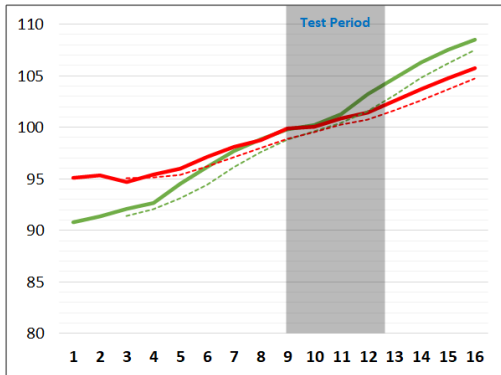


## People Using At Least Basic Drinking Water Services

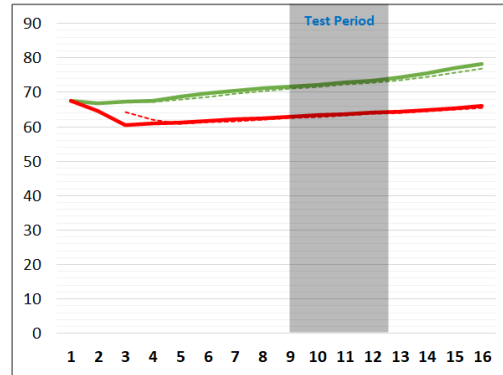
**Indicator** Composition of expenditure out-turn compared to original approved budget  
**Measure** People using at least basic drinking water services (% of population) (WHO/UNICEF)

Numbers in sample <sup>1</sup>	
Improving	24
Not Improving	40
<b>TOTAL</b>	<b>64</b>

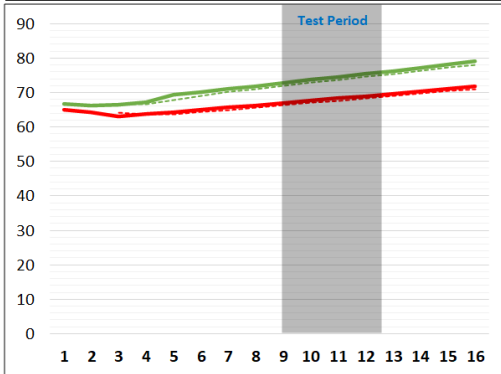
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



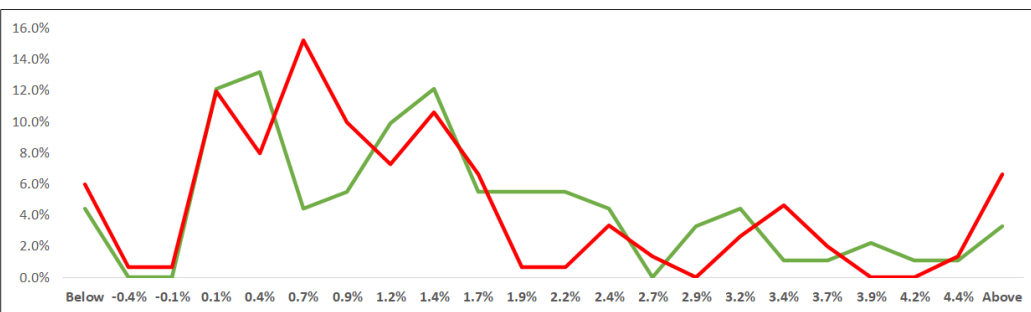
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	3.52
Actual Values (Median Average)	4.07
Actual Values (Mean Average)	1.91

Test Statistic	
Index Values (Median Average)	17.9%
Actual Values (Median Average)	1.2%
Actual Values (Mean Average)	14.2%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



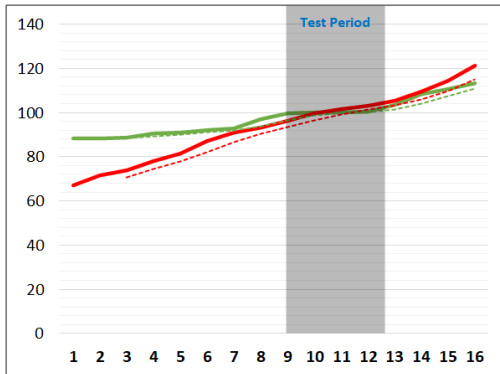
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## Access to Electricity

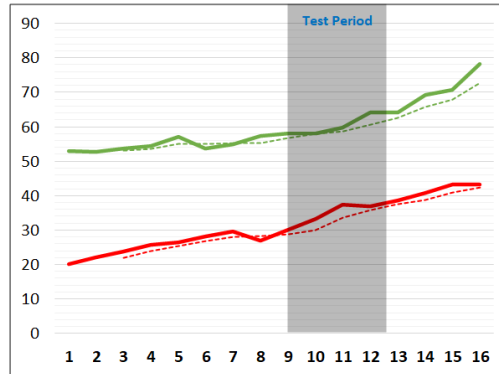
<b>Indicator</b>	Composition of expenditure out-turn compared to original approved budget
<b>Measure</b>	Access to electricity (% of population)(ELC.ACCS) (World Bank)

<b>Numbers in sample<sup>1</sup></b>	
Improving	25
Not Improving	41
<b>TOTAL</b>	<b>66</b>

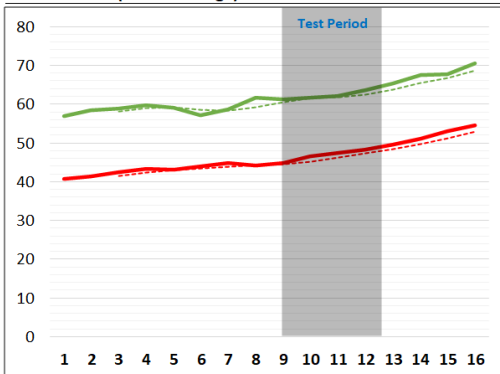
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



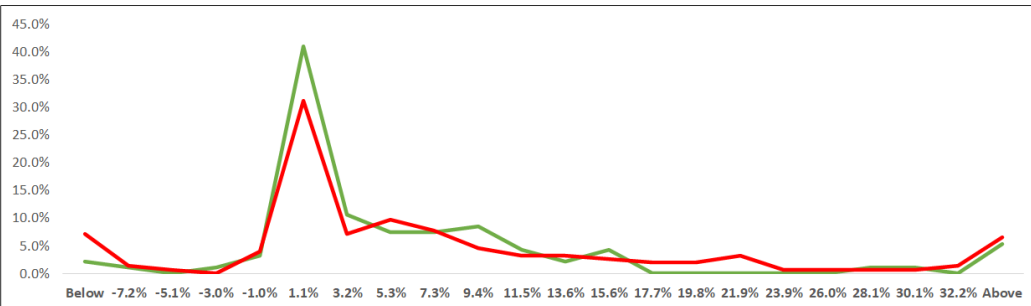
<b>Differences at t+4 years vs average prior to test period</b>	
Index Values (Median Average)	(13.01)
Actual Values (Median Average)	7.19
Actual Values (Mean Average)	0.76

<b>Test Statistic</b>	
Index Values (Median Average)	-
Actual Values (Median Average)	2.1%
Actual Values (Mean Average)	41.3%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



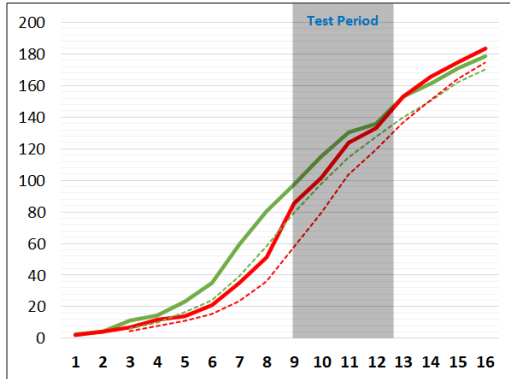
1/ Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## Mobile Cellular Subscriptions

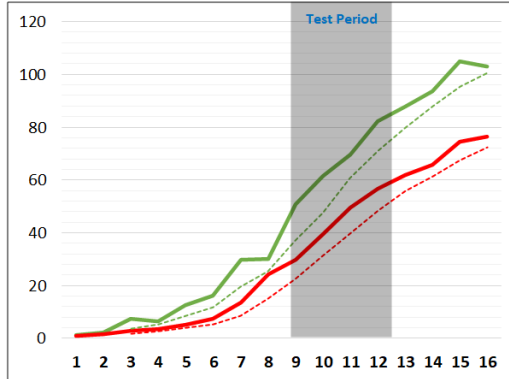
**Indicator** Composition of expenditure out-turn compared to original approved budget  
**Measure** Mobile cellular subscriptions (per 100 people) (CEL.SETS) (World Bank)

Numbers in sample <sup>1</sup>	
Improving	27
Not Improving	41
<b>TOTAL</b>	<b>68</b>

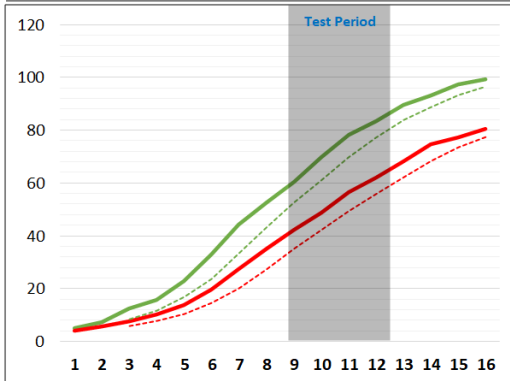
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



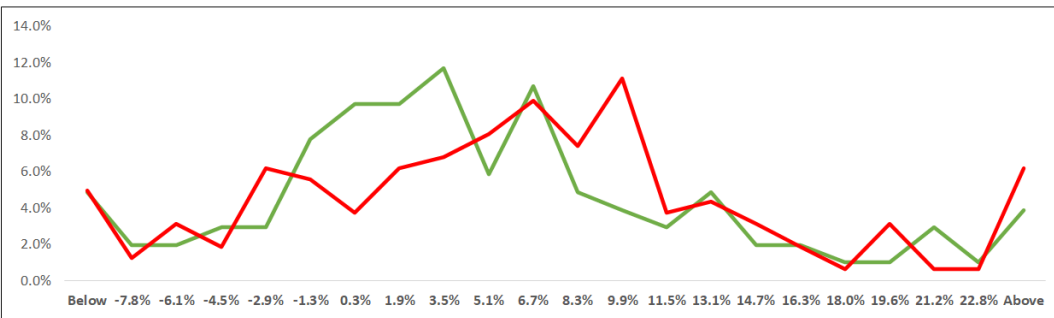
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	(23.98)
Actual Values (Median Average)	17.04
Actual Values (Mean Average)	4.51

Test Statistic	
Index Values (Median Average)	-
Actual Values (Median Average)	0.4%
Actual Values (Mean Average)	23.4%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

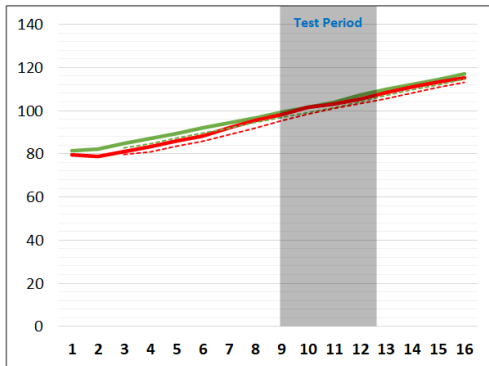
# Competition, Value for Money, and Controls in Procurement

## Water and Sanitation Composite Index

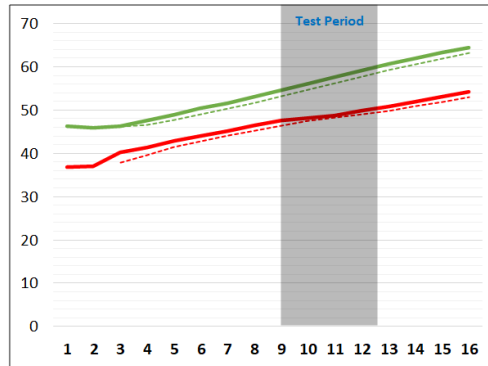
<b>Indicator</b>	Competition, value for money and controls in procurement
<b>Measure</b>	Water and Sanitation Composite Index (WSS) (ADB, AIDI) (3 year offset)

<b>Numbers in sample<sup>1</sup></b>	
Improving	19
Not Improving	17
<b>TOTAL</b>	<b>36</b>

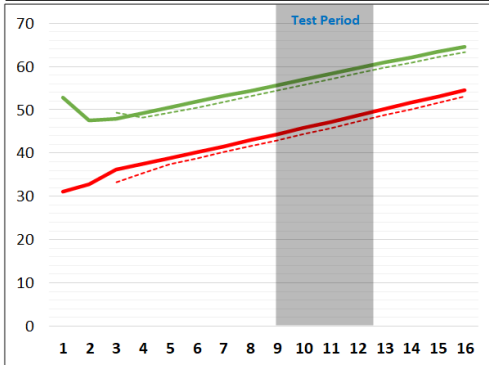
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



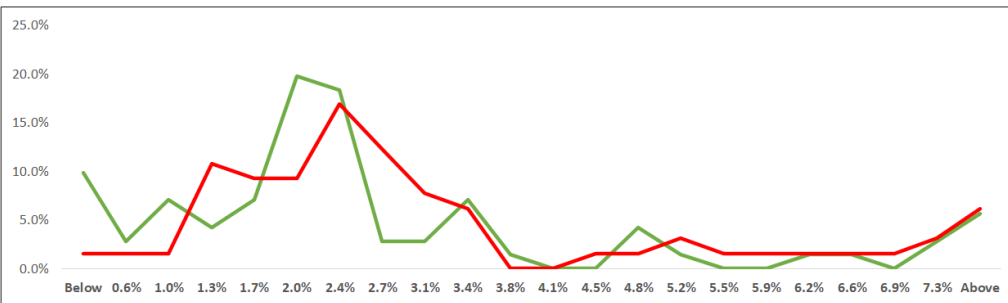
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	(1.20)
Actual Values (Median Average)	3.99
Actual Values (Mean Average)	(1.45)

Test Statistic	
Index Values (Median Average)	-
Actual Values (Median Average)	2.7%
Actual Values (Mean Average)	-

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. **I.e.** that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



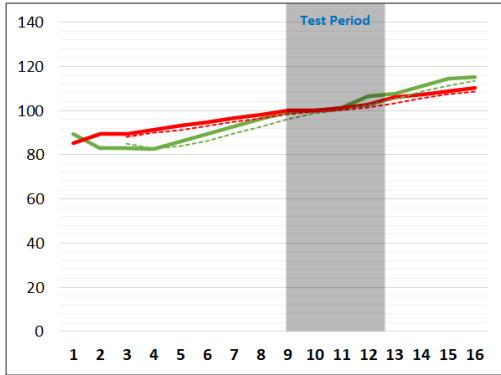
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## People Using Safely Managed Drinking Water Services

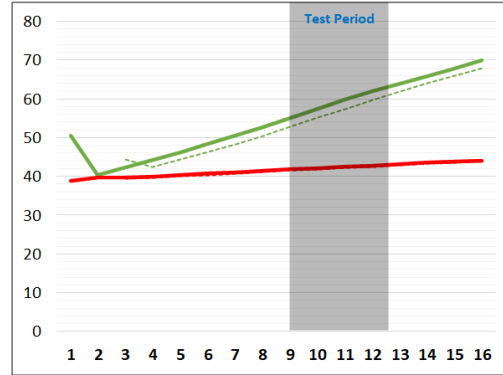
**Indicator** Competition, value for money and controls in procurement  
**Measure** People using safely managed drinking water services (% of population) (WHO/UNICEF)

Numbers in sample <sup>1</sup>	
Improving	9
Not Improving	15
<b>TOTAL</b>	<b>24</b>

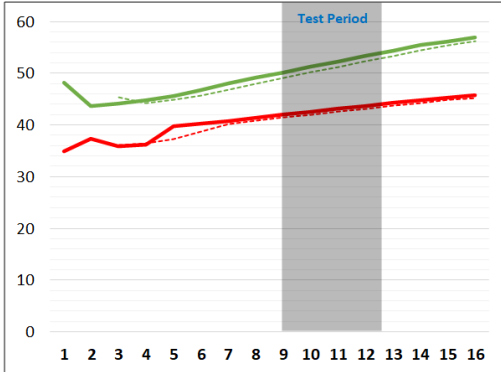
Index Values (Median Average)



Actual Values (Median Average)



Actual Values (Mean Average)



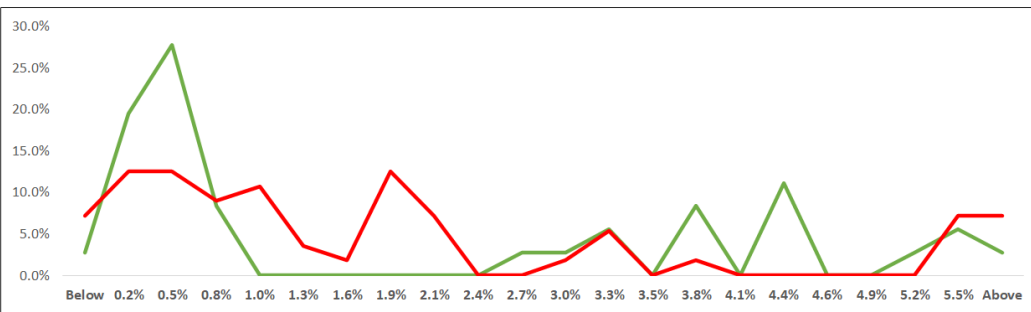
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	9.61
Actual Values (Median Average)	17.27
Actual Values (Mean Average)	4.41

Test Statistic	
Index Values (Median Average)	12.3%
Actual Values (Median Average)	0.0%
Actual Values (Mean Average)	9.5%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

Red - Not improving in indicator during test period  
Green - Improving in indicator during test period

Histogram of Percentage Changes in Actuals post Test-Period



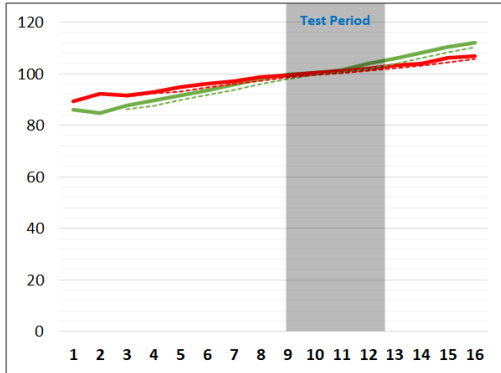
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## People Using At Least Basic Sanitation Services

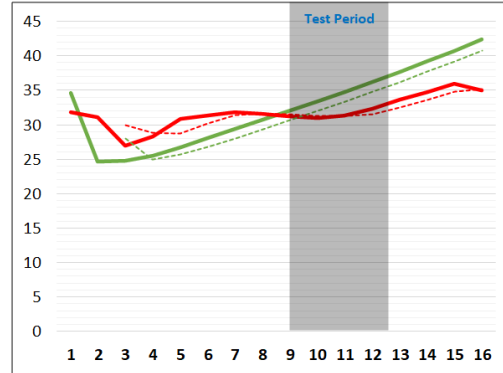
**Indicator** Competition, value for money and controls in procurement  
**Measure** People using at least basic sanitation services (% of population) (WHO/UNICEF)

Numbers in sample <sup>1</sup>	
Improving	30
Not Improving	34
<b>TOTAL</b>	<b>64</b>

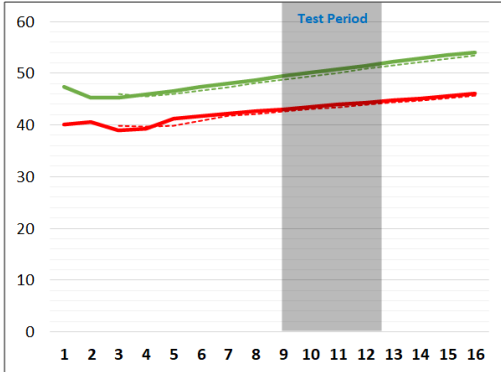
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



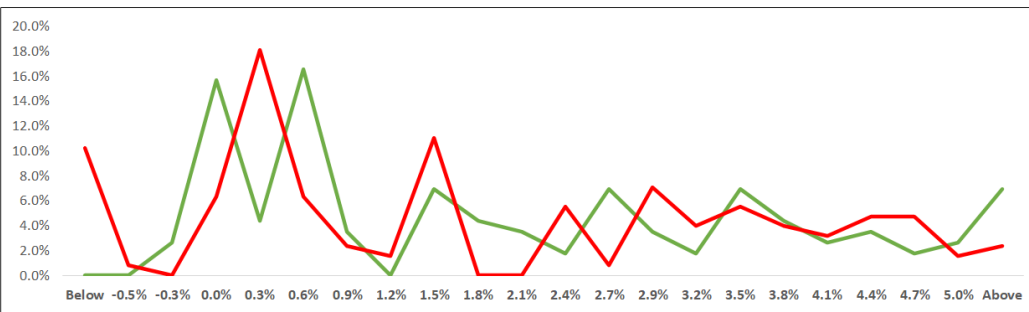
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	7.17
Actual Values (Median Average)	10.15
Actual Values (Mean Average)	2.30

Test Statistic	
Index Values (Median Average)	10.6%
Actual Values (Median Average)	0.0%
Actual Values (Mean Average)	7.6%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



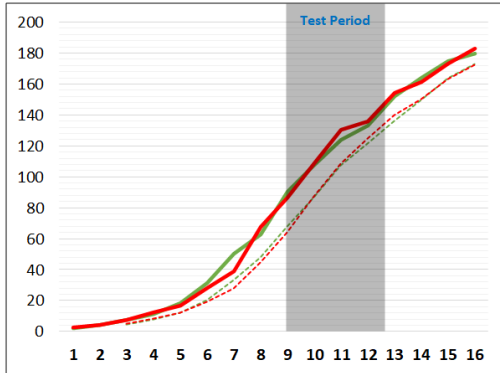
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## Mobile Cellular Subscriptions

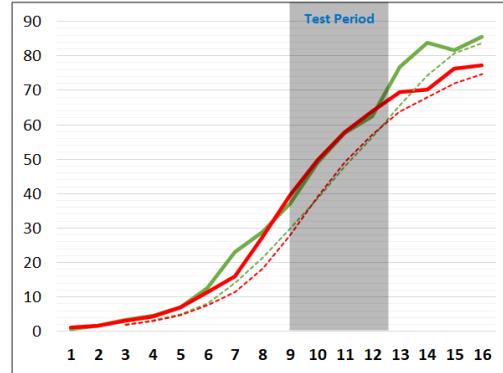
**Indicator** Competition, value for money and controls in procurement  
**Measure** Mobile cellular subscriptions (per 100 people) (CEL.SETS) (World Bank)

Numbers in sample <sup>1</sup>	
Improving	31
Not Improving	37
<b>TOTAL</b>	<b>68</b>

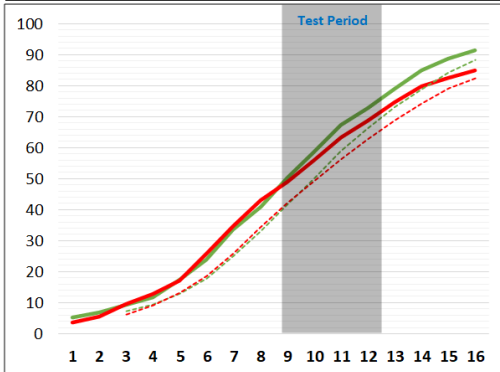
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



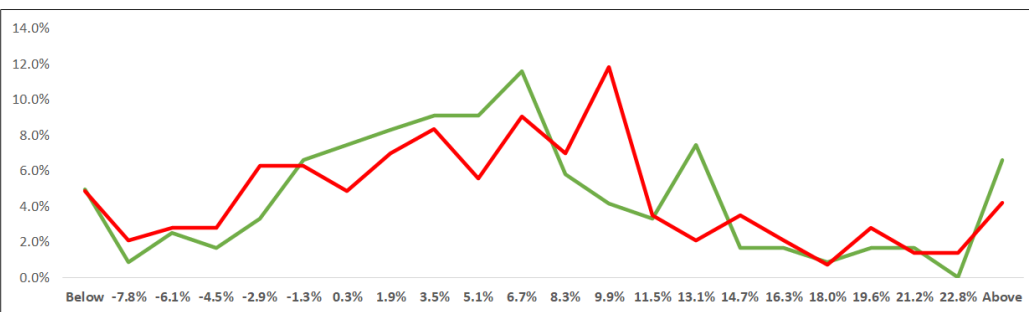
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	(6.02)
Actual Values (Median Average)	5.93
Actual Values (Mean Average)	7.53

Test Statistic	
Index Values (Median Average)	-
Actual Values (Median Average)	14.2%
Actual Values (Mean Average)	8.8%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

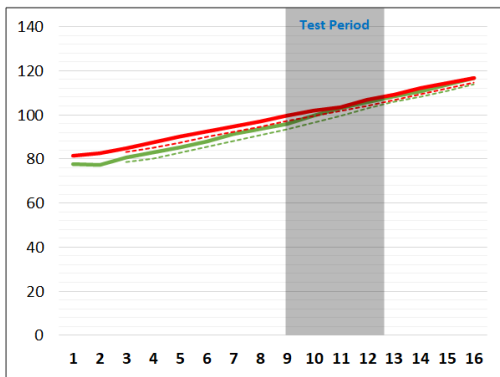
# Effectiveness of Internal Controls for Non-salary Expenditure

## Water and Sanitation Composite Index

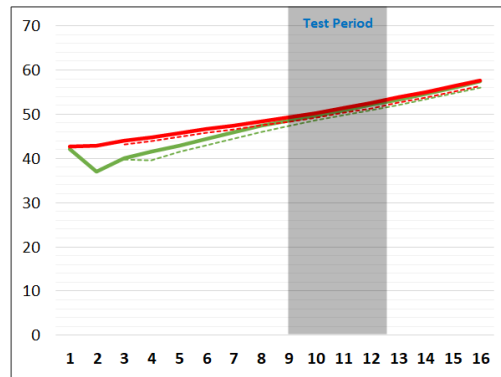
<b>Indicator</b>	Effectiveness of internal controls for non-salary expenditure
<b>Measure</b>	Water and Sanitation Composite Index (WSS) (ADB, AIDI) (3 year offset)

<b>Numbers in sample<sup>1</sup></b>	
Improving	12
Not Improving	24
<b>TOTAL</b>	<b>36</b>

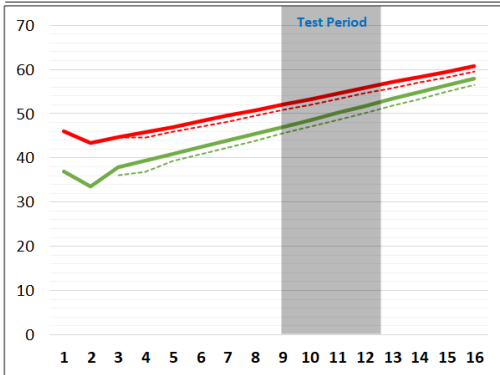
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



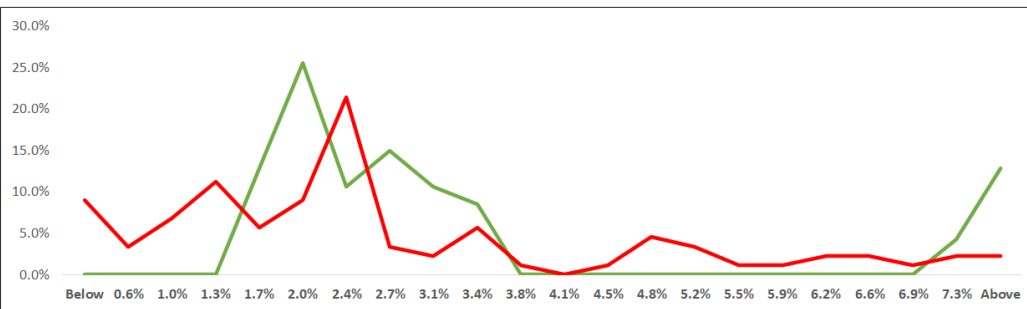
<b>Differences at t+4 years vs average prior to test period</b>	
Index Values (Median Average)	3.83
Actual Values (Median Average)	1.66
Actual Values (Mean Average)	3.04

<b>Test Statistic</b>	
Index Values (Median Average)	36.2%
Actual Values (Median Average)	20.8%
Actual Values (Mean Average)	6.9%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

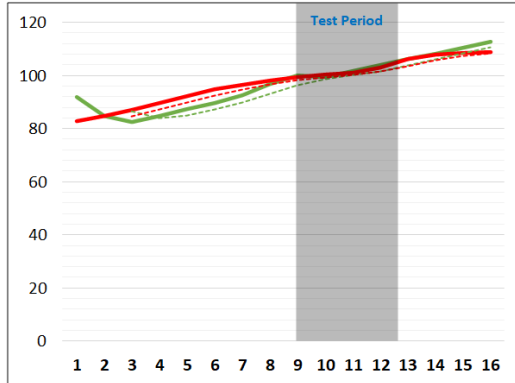


## People Using Safely Managed Drinking Water Services

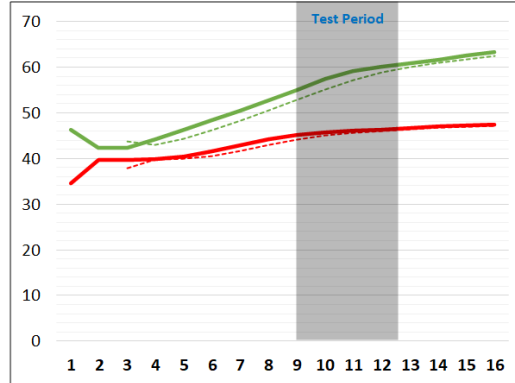
<b>Indicator</b>	Effectiveness of internal controls for non-salary expenditure
<b>Measure</b>	People using safely managed drinking water services (% of population) (WHO/UNICEF)

Numbers in sample <sup>1</sup>	
Improving	9
Not Improving	15
<b>TOTAL</b>	<b>24</b>

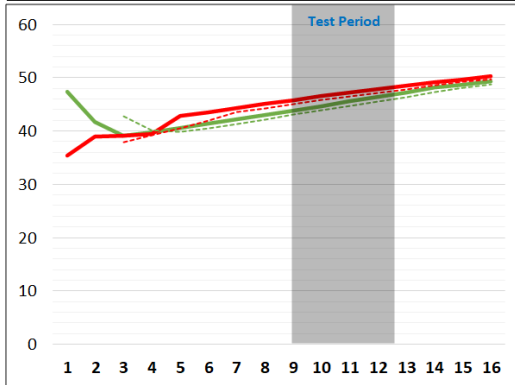
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



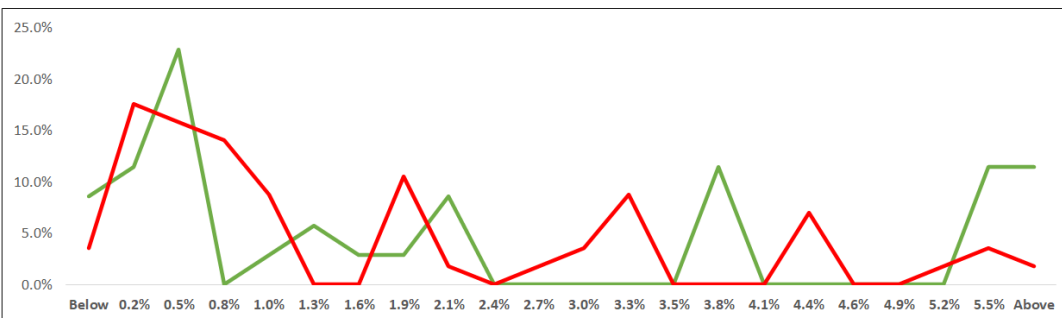
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	7.61
Actual Values (Median Average)	8.60
Actual Values (Mean Average)	1.26

Test Statistic	
Index Values (Median Average)	18.9%
Actual Values (Median Average)	0.7%
Actual Values (Mean Average)	35.1%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



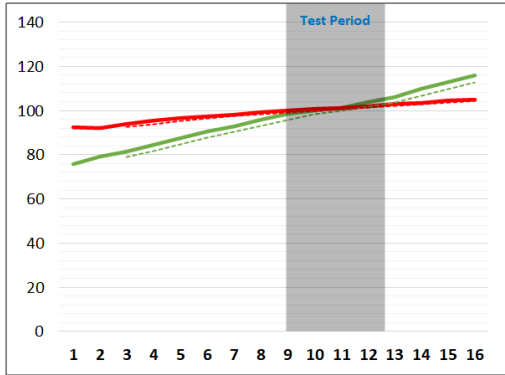
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## People Using At Least Basic Sanitation Services

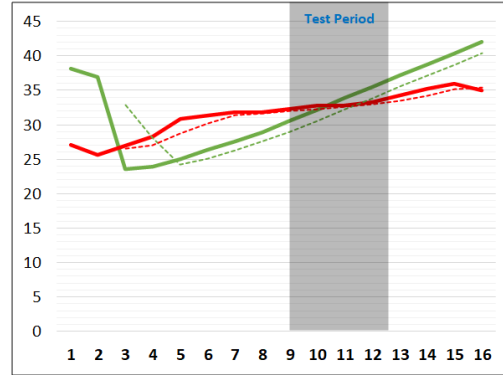
**Indicator** Effectiveness of internal controls for non-salary expenditure  
**Measure** People using at least basic sanitation services (% of population) (WHO/UNICEF)

Numbers in sample <sup>1</sup>	
Improving	24
Not Improving	40
<b>TOTAL</b>	<b>64</b>

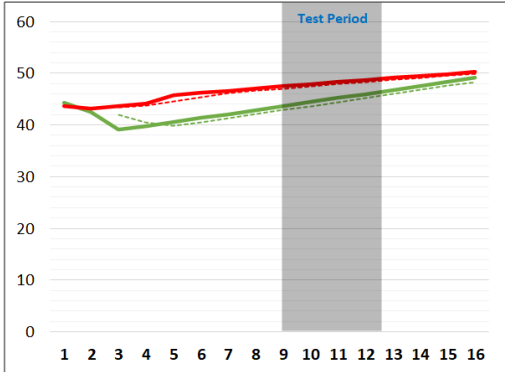
Index Values (Median Average)



Actual Values (Median Average)



Actual Values (Mean Average)



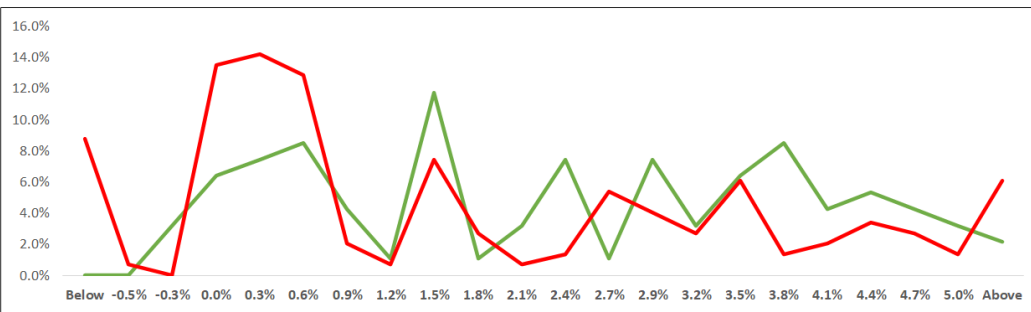
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	16.88
Actual Values (Median Average)	11.61
Actual Values (Mean Average)	3.62

Test Statistic	
Index Values (Median Average)	0.1%
Actual Values (Median Average)	0.0%
Actual Values (Mean Average)	1.3%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

Red - Not improving in indicator during test period  
Green - Improving in indicator during test period

Histogram of Percentage Changes in Actuals post Test-Period



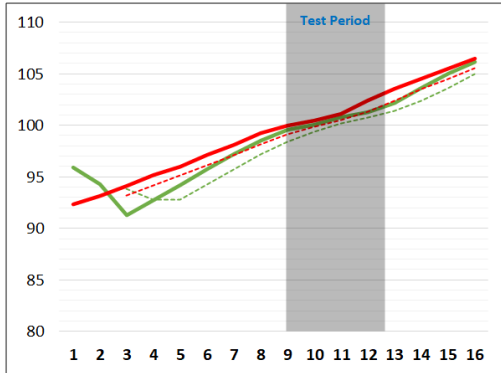
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## People Using At least Basic Drinking Water Services

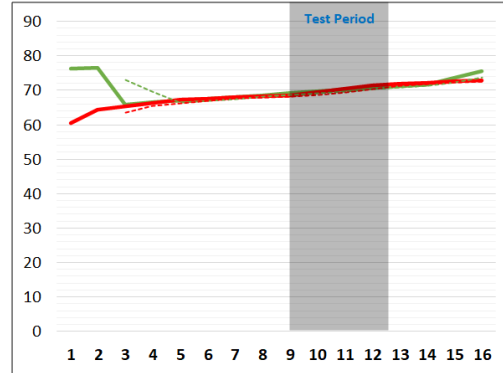
**Indicator** Effectiveness of internal controls for non-salary expenditure  
**Measure** People using at least basic drinking water services (% of population) (WHO/UNICEF)

Numbers in sample <sup>1</sup>	
Improving	24
Not Improving	40
<b>TOTAL</b>	<b>64</b>

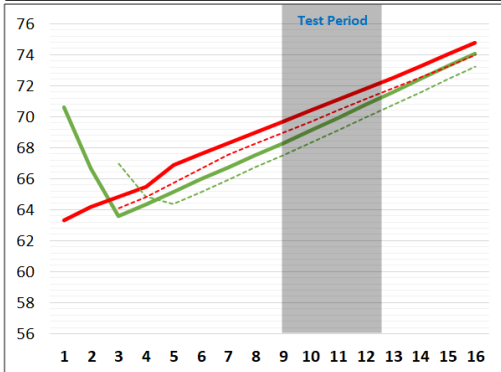
Index Values (Median Average)



Actual Values (Median Average)



Actual Values (Mean Average)



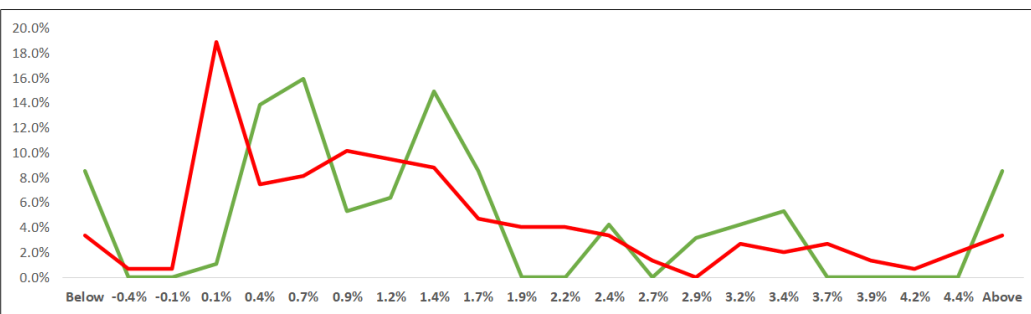
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	0.84
Actual Values (Median Average)	2.65
Actual Values (Mean Average)	0.89

Test Statistic	
Index Values (Median Average)	41.8%
Actual Values (Median Average)	7.3%
Actual Values (Mean Average)	31.1%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

Red - Not improving in indicator during test period  
Green - Improving in indicator during test period

Histogram of Percentage Changes in Actuals post Test-Period



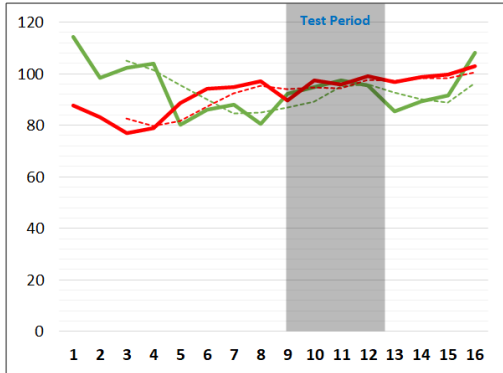
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## Estimated Public Gross Fixed Capital Formation

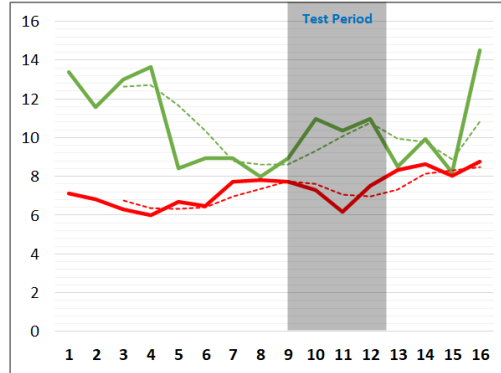
**Indicator** Effectiveness of internal controls for non-salary expenditure  
**Measure** Estimated Public Gross Fixed Capital Formation (% GDP) (Calculated from World Bank data)

Numbers in sample <sup>1</sup>	
Improving	23
Not Improving	36
<b>TOTAL</b>	<b>59</b>

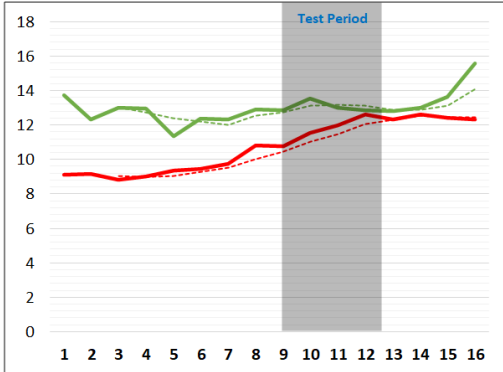
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



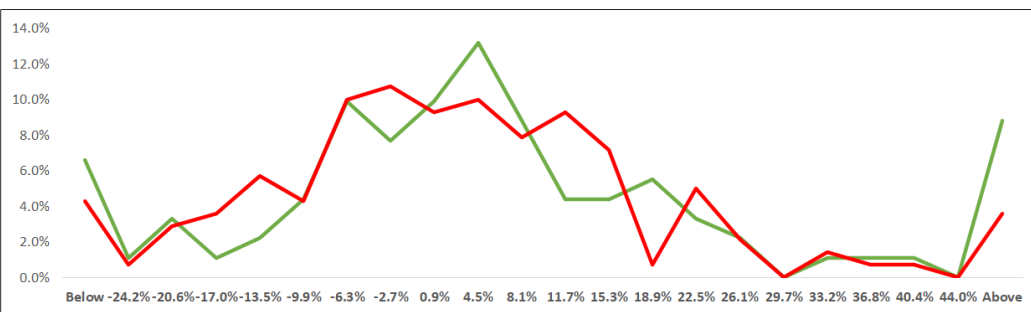
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	15.41
Actual Values (Median Average)	4.37
Actual Values (Mean Average)	0.86

Test Statistic	
Index Values (Median Average)	19.1%
Actual Values (Median Average)	3.7%
Actual Values (Mean Average)	36.2%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



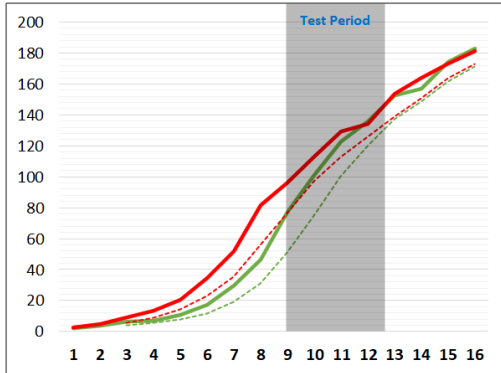
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## Mobile Cellular Subscriptions

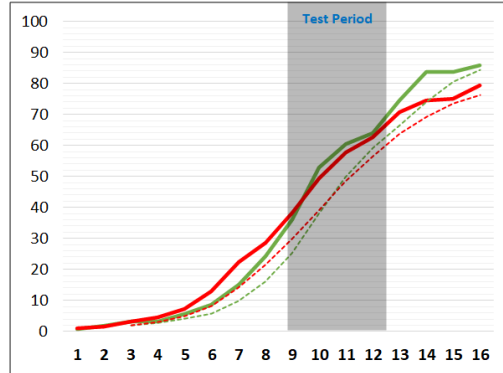
<b>Indicator</b>	Effectiveness of internal controls for non-salary expenditure
<b>Measure</b>	Mobile cellular subscriptions (per 100 people) (CEL.SETS) (World Bank)

Numbers in sample <sup>1</sup>	
Improving	26
Not Improving	42
<b>TOTAL</b>	<b>68</b>

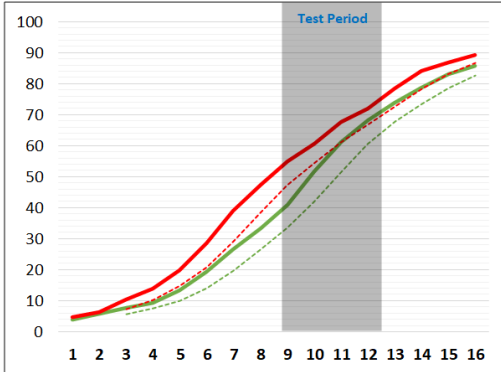
Index Values (Median Average)



Actual Values (Median Average)



Actual Values (Mean Average)



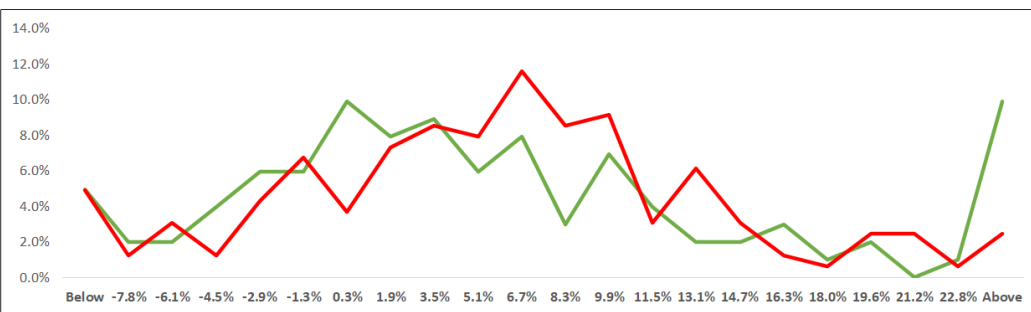
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	22.61
Actual Values (Median Average)	11.01
Actual Values (Mean Average)	7.08

Test Statistic	
Index Values (Median Average)	6.9%
Actual Values (Median Average)	3.7%
Actual Values (Mean Average)	12.4%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

Red - Not improving in indicator during test period  
Green - Improving in indicator during test period

Histogram of Percentage Changes in Actuals post Test-Period



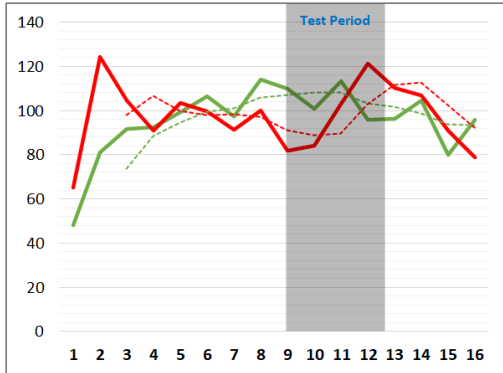
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## Net Acquisition of Non-Financial Assets

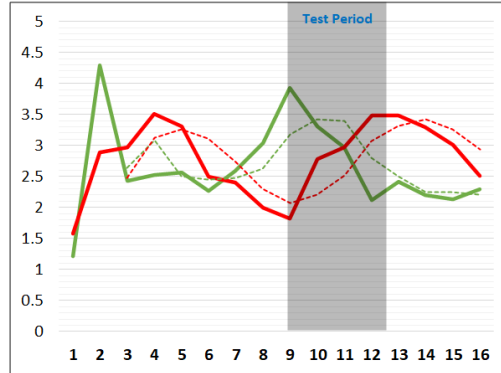
**Indicator** Effectiveness of internal controls for non-salary expenditure  
**Measure** Net acquisition of non financial assets (% of GDP) (GC.AST.TOTL.GD.ZS) (World Bank)

Numbers in sample <sup>1</sup>	
Improving	11
Not Improving	11
<b>TOTAL</b>	<b>22</b>

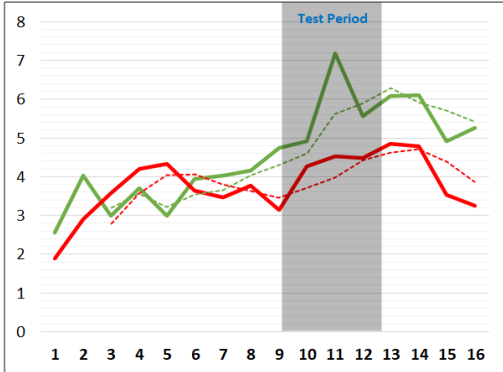
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



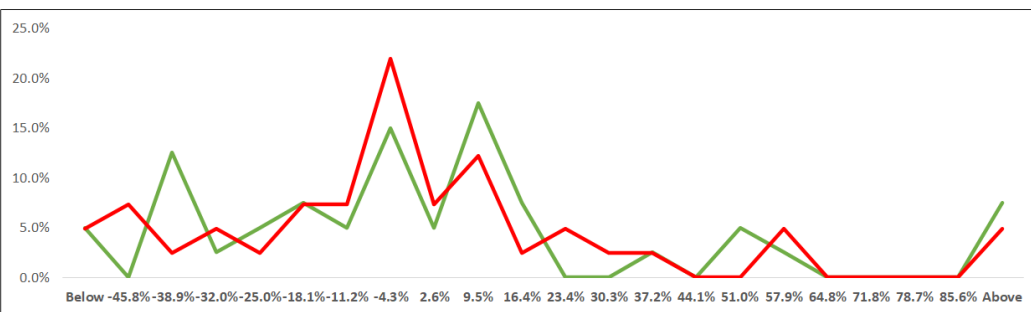
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	11.76
Actual Values (Median Average)	(0.28)
Actual Values (Mean Average)	2.03

Test Statistic	
Index Values (Median Average)	24.7%
Actual Values (Median Average)	-
Actual Values (Mean Average)	0.2%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

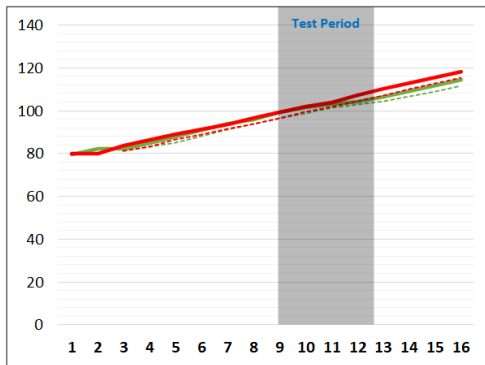
# Multiyear Perspective in Fiscal Planning, Expenditure Policy, and Budgeting

## Water and Sanitation Composite Index

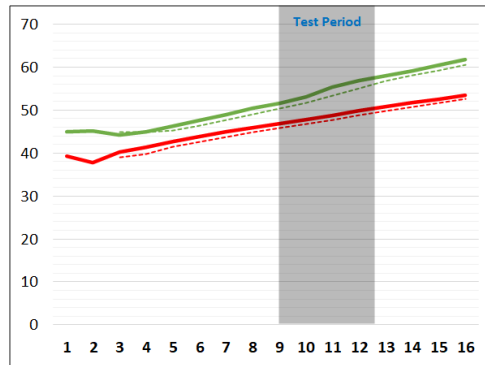
<b>Indicator</b>	Multi-year perspective in fiscal planning, expenditure policy and budgeting
<b>Measure</b>	Water and Sanitation Composite Index (WSS) (ADB, AIDI) (3 year offset)

Numbers in sample <sup>1</sup>	
Improving	17
Not Improving	19
<b>TOTAL</b>	<b>36</b>

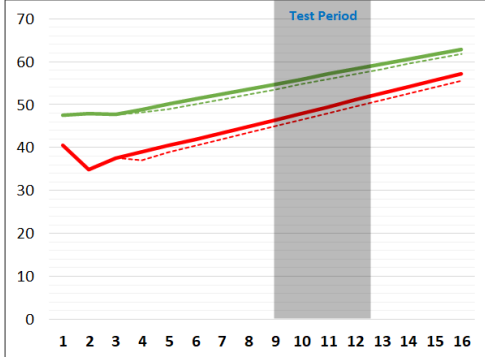
Index Values (Median Average)



Actual Values (Median Average)



Actual Values (Mean Average)



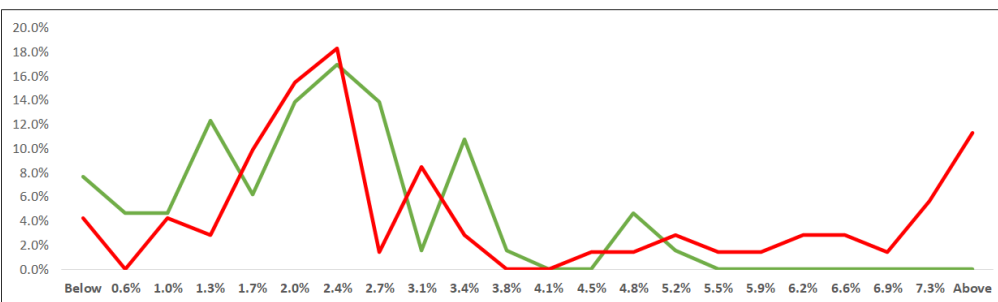
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	(3.51)
Actual Values (Median Average)	4.30
Actual Values (Mean Average)	(3.37)

Test Statistic	
Index Values (Median Average)	-
Actual Values (Median Average)	1.0%
Actual Values (Mean Average)	-

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

Red - Not improving in indicator during test period  
Green - Improving in indicator during test period

Histogram of Percentage Changes in Actuals post Test-Period



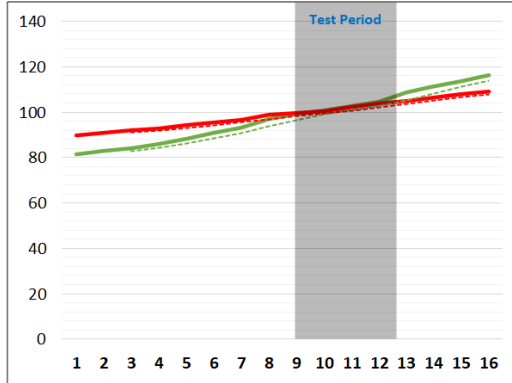
1/ Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## People Using Safely Managed Sanitation Services

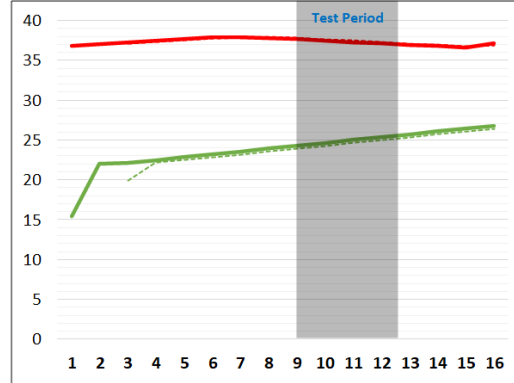
**Indicator** Multi-year perspective in fiscal planning, expenditure policy and budgeting  
**Measure** People using safely managed sanitation services (% of population) (WHO/UNICEF)

Numbers in sample <sup>1</sup>	
Improving	7
Not Improving	6
<b>TOTAL</b>	<b>13</b>

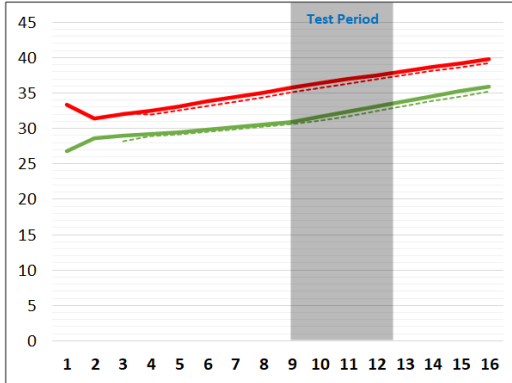
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



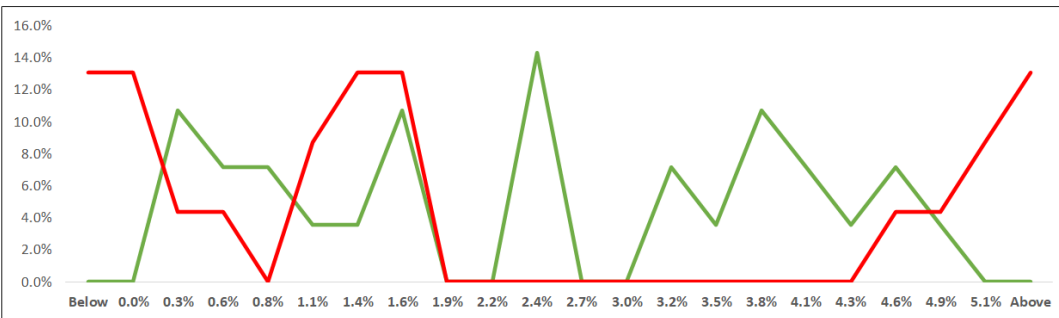
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	11.19
Actual Values (Median Average)	4.17
Actual Values (Mean Average)	0.12

Test Statistic	
Index Values (Median Average)	16.8%
Actual Values (Median Average)	8.9%
Actual Values (Mean Average)	48.3%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

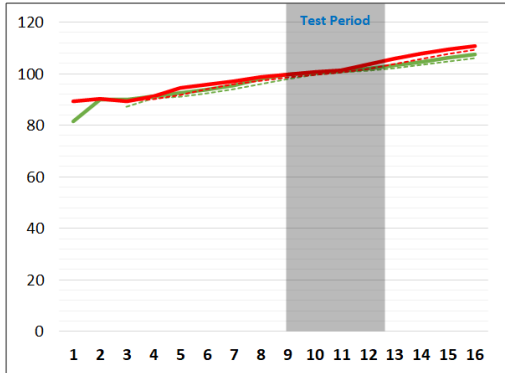


## People Using At Least Basic Sanitation Services

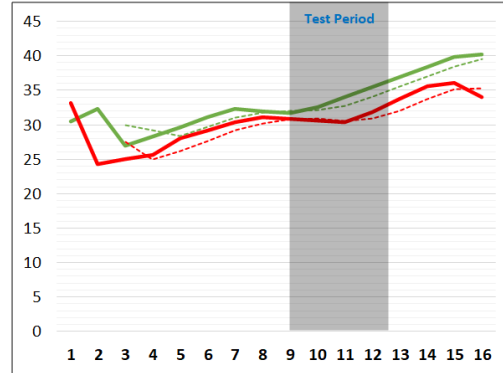
**Indicator** Multi-year perspective in fiscal planning, expenditure policy and budgeting  
**Measure** People using at least basic sanitation services (% of population) (WHO/UNICEF)

Numbers in sample <sup>1</sup>	
Improving	31
Not Improving	33
<b>TOTAL</b>	<b>64</b>

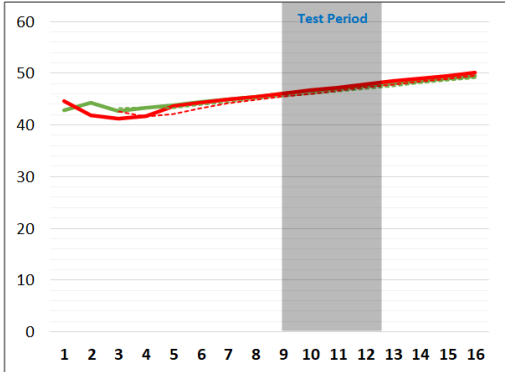
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



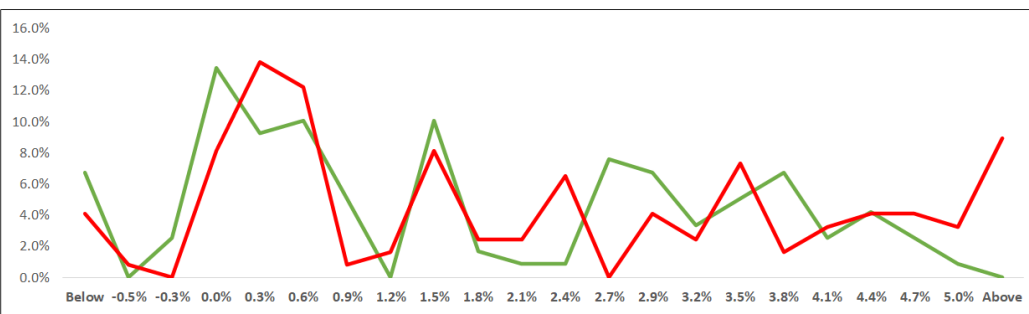
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	(1.81)
Actual Values (Median Average)	4.57
Actual Values (Mean Average)	(0.63)

Test Statistic	
Index Values (Median Average)	-
Actual Values (Median Average)	0.4%
Actual Values (Mean Average)	-

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

Red - Not improving in indicator during test period  
 Green - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



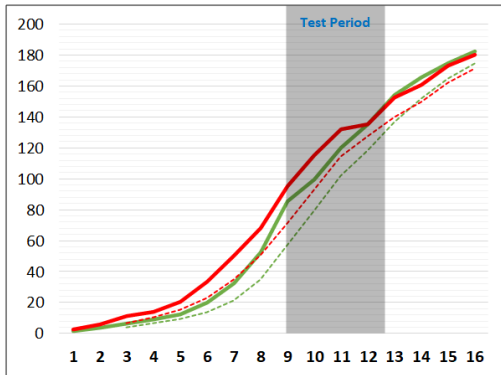
<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## Mobile Cellular Subscriptions

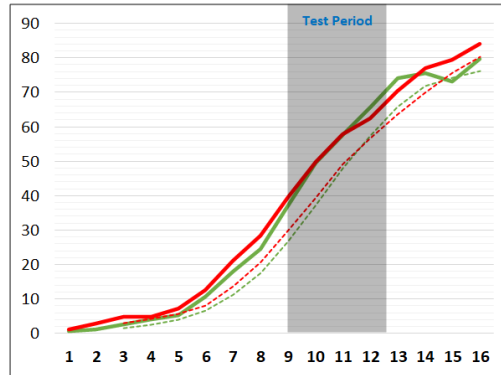
**Indicator** Multi-year perspective in fiscal planning, expenditure policy and budgeting  
**Measure** Mobile cellular subscriptions (per 100 people) (CEL.SETS) (World Bank)

Numbers in sample <sup>1</sup>	
Improving	33
Not Improving	35
<b>TOTAL</b>	<b>68</b>

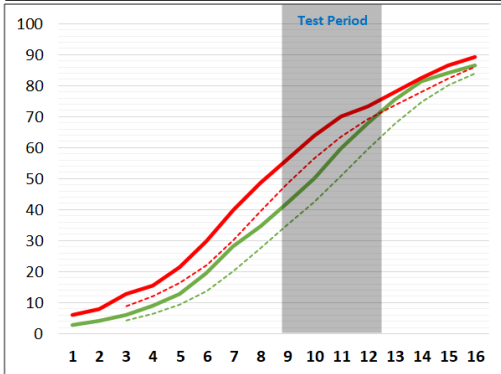
**Index Values (Median Average)**



**Actual Values (Median Average)**



**Actual Values (Mean Average)**



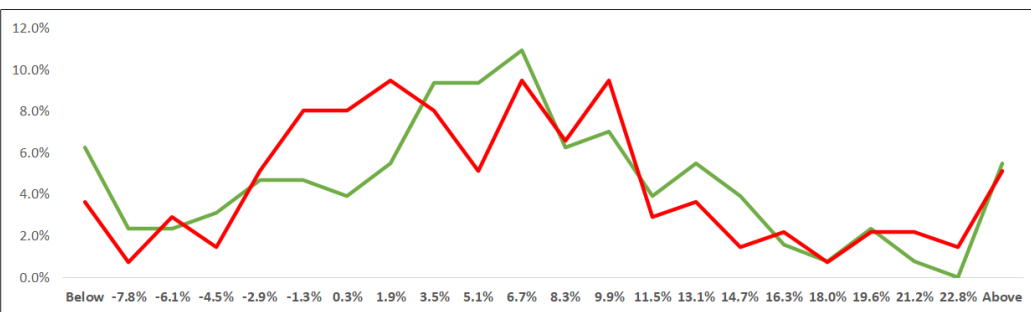
Differences at t+4 years vs average prior to test period	
Index Values (Median Average)	16.23
Actual Values (Median Average)	(1.58)
Actual Values (Mean Average)	8.63

Test Statistic	
Index Values (Median Average)	8.6%
Actual Values (Median Average)	-
Actual Values (Mean Average)	6.2%

The above table shows significance tests, a probability of less than 10% indicates limited likelihood that the values of the measure 4 years after the improvement in the indicator, are equal to each other. I.e. that there is a statistically significant difference between the set with improved PEFA scores, and the set without improved PEFA scores

**Red** - Not improving in indicator during test period  
**Green** - Improving in indicator during test period

**Histogram of Percentage Changes in Actuals post Test-Period**



<sup>1/</sup> Values must show an improvement of 0.5 in the PEFA score for that indicator, in addition to having sufficient numbers of years before and after the change period

## References

---

- Acevedo-Espinoza, Sebastián, Aliona Cebotari, Kevin Greenidge, and Geoffrey Keim. 2015. “External Devaluations: Are Small States Different?” IMF Working Paper 15/240, International Monetary Fund, Washington, DC.  
[https://www.researchgate.net/publication/298722908\\_External\\_Devaluations\\_Are\\_Small\\_States\\_Different](https://www.researchgate.net/publication/298722908_External_Devaluations_Are_Small_States_Different).
- Addison, Douglas. 2013. “The Quality of Budget Execution and Its Correlates.” Policy Research Working Paper 6657, World Bank, Washington, DC.  
<http://documents.worldbank.org/curated/en/259201468314673894/pdf/WPS6657.pdf>.
- Andrews, Matt. 2010. “How Far Have Public Financial Management Reforms Come in Africa?” Harvard Kennedy School, Cambridge, MA.  
[https://www.pempal.org/sites/pempal/files/attachments/RWP10-018\\_Andrews.pdf](https://www.pempal.org/sites/pempal/files/attachments/RWP10-018_Andrews.pdf).
- Andrews, Matt. 2011. “Which Organizational Attributes Are Amenable to External Reform? An Empirical Study of African Public Financial Management.” *International Public Management Journal* 14 (2): 131–56. doi: 10.1080/10967494.2011.588588.
- Andrews, Matt, Marco Cangiano, Neil Cole, Paolo de Renzio, Philipp Krause, and Renaud Seligmann. 2014. “This Is PFM.” CID Working Paper 285, Center for International Development, Harvard University, Cambridge, MA.  
[https://www.hks.harvard.edu/sites/default/files/centers/cid/files/publications/faculty-working-papers/285\\_Andrews\\_This+is+PFM.pdf](https://www.hks.harvard.edu/sites/default/files/centers/cid/files/publications/faculty-working-papers/285_Andrews_This+is+PFM.pdf).
- Asian Development Bank. 2017. *Public Expenditure and Financial Accountability: Assessing Public Financial Management Performance and Influencing Reform Processes: Experience from Asia and the Pacific*. Manila: Asian Development Bank.  
<http://dx.doi.org/10.22617/BRF179129>.
- Bandypadhyay, Sruti. 2016. “Linking PFM Reforms to Service Delivery Outcomes: A Spotlight on the Middle East and North Africa.” *Middle East Review of Public Administration (MERPA)* 2 (3). <https://app.scholasticahq.com/api/v1/attachments/1323/download>.
- Baudienville, Geraldine. 2012. “Public Financial Management Reforms in Fragile States: The Case of Democratic Republic of the Congo.” Overseas Development Institute, London.
- Cangiano, Marco, Tessa Curristine, and Michel Lazare. 2013a. “Managing Public Investment.” In *Public Financial Management and Its Emerging Architecture*, edited by Marco Cangiano, Tessa Curristine, and Michele Lazare, ch. 10. Washington, DC: International Monetary Fund. doi: <https://doi.org/10.5089/9781475531091.071>.
- Cangiano, Marco, Tessa Curristine, and Michel Lazare. 2013b. “Reflections on Two Decades of Public Financial Management Reforms.” In *Public Financial Management and Its Emerging Architecture*, edited by Marco Cangiano, Tessa Curristine, and Michele Lazare, ch. 1. Washington, DC: International Monetary Fund. doi: <https://doi.org/10.5089/9781475531091.071>.

- Chen, Siyan, Norman V. Loayza, and Marta Reynal-Querol. 2007. "The Aftermath of Civil War." Policy Research Working Paper 4190, World Bank, Washington, DC. <https://openknowledge.worldbank.org/handle/10986/7006>.
- Chen, Xing, Abdul Munasib, and Devesh Roy. 2012. "Financial Reforms and International Trade." IFPRI Discussion Paper 01182, International Food Policy Research Institute, Washington, DC. <http://cdm15738.contentdm.oclc.org/utis/getfile/collection/p15738coll2/id/126929/filename/127140.pdf>.
- Dabla-Norris, Era, Richard Allen, Luis-Felipe Zanna, Tej Prakash, Eteri Kvintradze, Victor Lledo, Irene Yackovlev, and Sophia Gollwitzer. 2010. "Budget Institutions and Fiscal Performance in Low-Income Countries." IMF Working Paper 10/80, International Monetary Fund, Washington, DC. <https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Budget-Institutions-and-Fiscal-Performance-in-Low-Income-Countries-23734>.
- de Renzio, Paolo. 2009. *Taking Stock: What Do PEFA Assessments Tell Us about PFM Systems across Countries?* London: Overseas Development Institute. <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/4359.pdf>.
- de Renzio, Paolo, Matthew Andrews, and Zac Mills. 2010. *Evaluation of Donor Support to Public Financial Management (PFM) Reform in Developing Countries: Analytical Study of Quantitative Cross-Country Evidence*. London: Overseas Development Institute. <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/6373.pdf>.
- de Renzio, Paolo, and Bill Dorintsky. 2007. "Tracking Progress in the Quality of PFM Systems in HIPCs: An Update on Past Assessments Using PEFA Data." PEFA Secretariat, Washington, DC. <https://www.pefa.org/resources/tracking-progress-quality-pfm-systems-hipcs>.
- Fama, Eugene F., Lawrence Fisher, Michael C. Jensen, and Richard Roll. 1969. "The Adjustment of Stock Prices to New Information." *International Economic Review* 10 (1): 1–21. <https://www.jstor.org/stable/2525569>.
- Ferry, Marin. 2015. "The Carry and Stick Approach to Debt Relief: Overcoming Moral Hazard." Working Paper DT/2015-14, Université Paris-Dauphine, Paris. [http://en-dial.ird.fr/content/download/226232/3434414/version/2/file/Ferry\\_DT2015-14.pdf](http://en-dial.ird.fr/content/download/226232/3434414/version/2/file/Ferry_DT2015-14.pdf).
- Fonchamnyo, Dobdinga C., and Molem C. Sama. 2016. "Determinants of Public Spending Efficiency in Education and Health: Evidence from Selected CEMAC Countries." *Journal of Economics and Finance* 40 (1): 199–210. <https://link.springer.com/article/10.1007/s12197-014-9310-6>.
- Fritz, Verena, Stephanie Sweet, and Marijn Verhoeven. 2014. *Strengthening Public Financial Management: Exploring Drivers and Effects*. Washington, DC: World Bank. <https://openknowledge.worldbank.org/handle/10986/20631>.

- Fritz, Verena, Marijn Verhoeven, and Ambra Avenia. 2017. *Political Economy of Public Financial Management Reforms: Experiences and Implications for Dialogue and Operational Engagement*. Washington, DC: World Bank.  
<https://openknowledge.worldbank.org/bitstream/handle/10986/28887/121436-15-11-2017-18-52-15-NTDofPFMReformsReportWeb.pdf?sequence=5&isAllowed=y>.
- Fukuda-Parr, Sakiko, Patrick Guyer, and Terra Lawson-Remer. 2011. “Does Budget Transparency Lead to Stronger Human Development Outcomes and Commitments to Economic and Social Rights?” IBP Working Paper 4, International Budget Partnership, Washington, DC. <https://www.internationalbudget.org/wp-content/uploads/IBP-Working-Paper-4-Budget-Transparency-and-Development-Outcomes-and-Rights.pdf>.
- Goryakin, Yevgeniy, Paul Reville, Andrew Mirelman, Rohan Sweeney, Jessica Ochalek, and Marc Suhrcke. 2017. *Public Financial Management and Health Service Delivery: A Literature Review*. London: Overseas Development Institute.  
<http://eprints.whiterose.ac.uk/116571/1/11463.pdf>.
- Hadley, Sierd, and Mark Miller. 2016. *PEFA: What Is It Good for? The Role of PEFA Assessments in Public Financial Management Reform*. London: Overseas Development Institute. <https://www.odi.org/sites/odi.org.uk/files/resource-documents/10484.pdf>.
- Haque, Tobias Akhtar, David Knight, and Dinuk Jayasuriya. 2012. “Capacity Constraints and Public Financial Management in Small Pacific Island Countries.” *Asia & The Pacific Policy Studies* 2(3): 609-622. DOI:10.1002/app5.79.
- Hedger, Edward, and Paolo de Renzio. 2010. “What Do Public Financial Management Assessments Tell Us about PFM Reform?” Background Note, Overseas Development Institute, London, July. <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/5904.pdf>.
- Hemming, Richard, Michael W. Kell, and Axel Schimmelpfennig. 2003. “Fiscal Vulnerability and Financial Crises in Emerging Market Economies.” IMF Occasional Paper 218, International Monetary Fund, Washington, DC.  
<https://www.semanticscholar.org/paper/Fiscal-vulnerability-and-financial-crises-in-market-Hemming-Kell/6197191c572d1058487591e8a119a50f0c63df7d>.
- IMF (International Monetary Fund). 2015. “Making Public Investment More Efficient.” IMF Policy Paper, International Monetary Fund, Washington, DC.  
<https://www.imf.org/external/np/pp/eng/2015/061115.pdf>.
- Jacobs, Davina Frederika. 2008. “A Review of Capital Budgeting Practices.” IMF Working Paper 08/160, 1–24, International Monetary Fund, Washington, DC.  
<https://ssrn.com/abstract=1160053>.
- Kristensen, Jens Kromann, Martin Bowen, Urška Zrinski, Cathal Long, and Shakira Mustapha. 2019a. “Introduction: What Is PFM and Why Is It Important?” In *PEFA, Public Financial Management, and Good Governance*, edited by Jens Kromann Kristensen, Martin Bowen, Urška Zrinski, Cathal Long, and Shakira Mustapha, 1–7. Washington, DC: World Bank. doi: <https://doi.org/10.1596/978-1-4648-1466-2>.

- Kristensen, Jens Kromann, Martin Bowen, Urška Zrinski, Cathal Long, and Shakira Mustapha. 2019b. “Measuring PFM Performance through PEFA: Approach and Methodology?” In *PEFA, Public Financial Management, and Good Governance*, edited by Jens Kromann Kristensen, Martin Bowen, Urška Zrinski, Cathal Long, and Shakira Mustapha, 9–36. Washington, DC: World Bank. doi: <https://doi.org/10.1596/978-1-4648-1466-2>.
- Lawson, Andrew. 2012. *Evaluation of Public Financial Management Reform: Burkina Faso, Ghana, and Malawi 2001–2010 Final Synthesis Report*. Stockholm: Swedish International Development Cooperation Agency. <https://www.oecd.org/derec/afdb/publicmanagementreform.pdf>.
- Lederman, Daniel, and Claudio J. Rojas. 2018. “Three Macroeconomic Trends around the Onset of Armed Conflict in Developing Economies.” Policy Research Working Paper 8647, World Bank, Washington, DC. <http://documents1.worldbank.org/curated/ru/763071542314874296/pdf/WPS8647.pdf>.
- Manova, Kalina. 2008. “Credit Constraints, Equity Market Liberalizations, and International Trade.” *Journal of International Economics* 76 (1): 33–47. <https://doi.org/10.1016/j.jinteco.2008.03.008>.
- Miller, Mark, and Tom Hart. 2017. *Strengthening Public Investment Management: Reviewing the Role of External Actors*. London: Overseas Development Institute.
- Mishra, Jaideep Kumar. 2014. “Study of Public Financial Management Systems for Two Asian Regional Groups: Comparisons and Key Features for SAARCH and ASEAN Regions.” *OECD Journal on Budgeting* 14 (3): 1–25. <https://dx.doi.org/10.1787/budget-14-5jrw9czdg326>.
- Mustapha, Shakira. 2019a. “Budget Credibility, Fiscal Outcomes, and PFM Performance in Fragile and Nonfragile Countries.” In *PEFA, Public Financial Management, and Good Governance*, edited by Jens Kromann Kristensen, Martin Bowen, Urška Zrinski, Cathal Long, and Shakira Mustapha, 61–92. Washington, DC: World Bank. <https://elibrary.worldbank.org/doi/book/10.1596/978-1-4648-1466-2>.
- Mustapha, Shakira. 2019b. “Political Institutions and PFM Performance.” In *PEFA, Public Financial Management, and Good Governance*, edited by Jens Kromann Kristensen, Martin Bowen, Urška Zrinski, Cathal Long, and Shakira Mustapha, 37–59. Washington, DC: World Bank. <https://elibrary.worldbank.org/doi/book/10.1596/978-1-4648-1466-2>.
- Omollo, Bernad. 2018. “Public Financial Management Reforms and Fiscal Policy: A Qualitative Content Analysis of the Impact of PFM Reforms on Country Budget Income in Sub-Saharan Africa: A Case for Kenya and Rwanda.” *OECD Journal on Budgeting* 18 (1): 167–201. [https://www.oecd-ilibrary.org/governance/public-financial-management-reforms-and-fiscal-policy-a-qualitative-content-analysis-of-the-impact-of-pfm-reforms-on-country-budget-income-in-sub-sahara-africa\\_budget-18-5j8l804fvj8v](https://www.oecd-ilibrary.org/governance/public-financial-management-reforms-and-fiscal-policy-a-qualitative-content-analysis-of-the-impact-of-pfm-reforms-on-country-budget-income-in-sub-sahara-africa_budget-18-5j8l804fvj8v).
- PEFA Secretariat. 2011. *Public Financial Management Performance Measurement Framework*. Washington, DC: PEFA Secretariat.



- [https://www.pefa.org/sites/pefa/files/resources/downloads/PMFEng-finalSZreprint04-12\\_1.pdf](https://www.pefa.org/sites/pefa/files/resources/downloads/PMFEng-finalSZreprint04-12_1.pdf).
- PEFA Secretariat. 2018. *Stocktake of PFM Diagnostic Tools 2016*. Washington, DC: PEFA Secretariat. <https://www.pefa.org/resources/stocktake-pfm-diagnostic-tools-2016>.
- Piatti-Funkfkirchen, Moritz, and Pia Schneider. 2018. “From Stumbling Block to Enabler: The Role of Public Financial Management in Health Service Delivery in Tanzania and Zambia.” *Health Systems and Reform* 4 (4): 336–45. <https://www.tandfonline.com/doi/full/10.1080/23288604.2018.1513266>.
- Piatti-Funkfkirchen, Moritz, and Lodewijk Smets. 2019. “Public Financial Management, Health Financing, and Under-Five Mortality: A Comparative Empirical Analysis.” IDB Working Paper IDB-WP-976, Inter-American Development Bank, Washington, DC. <https://www.econstor.eu/handle/10419/208166>.
- Prakash, Tej, and Ezequiel Cabezon. 2008. “Public Financial Management and Fiscal Outcomes in Sub-Saharan African Heavily-Indebted Poor Countries.” IMF Working Paper 08/217, International Monetary Fund, Washington, DC. <https://www.imf.org/external/pubs/ft/wp/2008/wp08217.pdf>.
- Pretorius, Carole, and Nico Pretorius. 2009. *A Review of Reform Literature*. London: Department for International Development. <https://europa.eu/capacity4dev/file/10170/download?token=vmzv2IKo>.
- Quak, Evert-Jan. 2020. “The Impact of Public Finance Management (PFM) Reforms on Education in Tanzania.” Institute of Development Studies, Brighton. [https://assets.publishing.service.gov.uk/media/5e3c2787e5274a08deb46b83/711\\_Public\\_Finance\\_Management\\_Reforms\\_Impact\\_Education\\_Tanzania.pdf](https://assets.publishing.service.gov.uk/media/5e3c2787e5274a08deb46b83/711_Public_Finance_Management_Reforms_Impact_Education_Tanzania.pdf).
- Rao, Sumedh. 2013. “Measuring the Impact of PFM Reforms on Service Delivery.” University of Birmingham. <http://www.gsdrc.org/docs/open/hdq946.pdf>.
- Salinas, Gonzalo, and Ataman Aksoy. 2006. “Growth before and after Trade Liberalization.” Policy Research Working Paper 4062, World Bank, Washington, DC. <http://documents1.worldbank.org/curated/en/622251468141266954/pdf/wps4062.pdf>.
- Sarr, Babacar. 2015. “Credibility and Reliability of Government Budgets: Does Fiscal Transparency Matter?” IBP Working Paper, International Budget Partnership, Washington, DC. <http://internationalbudget.org/wp-content/uploads/Credibility-and-Reliability-of-Government-Budgets-Working-Papers.pdf>.
- Vlaicu, Razvan, Merijn Verhoeven, Francesco Grigoli, and Zachary Mills. 2014. “Multiyear Budgets and Fiscal Performance: Panel Data Evidence.” *Journal of Public Economics* 111 (March): 79–95. <https://www.aeaweb.org/conference/2014/retrieve.php?pdfid=326>.
- Welham, Bryn, Tom Hard, Shakira Mustapha, and Sierd Hadley. 2017. *Public Financial Management and Health Service Delivery: Necessary, but Not Sufficient?* London: Overseas Development Institute. <https://www.odi.org/sites/odi.org.uk/files/resource-documents/11462.pdf>.

World Bank. 2010. *Public Financial Management Reform in the Middle East and North Africa: An Overview of Regional Experience*. Washington, DC: World Bank.

<https://www.pefa.org/resources/public-financial-management-reform-middle-east-and-north-africa-overview-regional>.

World Bank. 2012. *Public Financial Management Reforms in Post-Conflict Countries: Synthesis Report*. Washington, DC: World Bank.

<http://documents.worldbank.org/curated/en/945231468340162289/pdf/699640WPOP1206070023B0PFM0Web0Final.pdf>.

World Bank. 2013. *Beyond the Annual Budget: Global Experience with Medium-Term Expenditure Frameworks*. Washington, DC: World Bank.

<http://documents.worldbank.org/curated/en/354601468330959258/pdf/Beyond-the-annual-budget-global-experience-with-medium-term-expenditure-frameworks.pdf>.

World Bank. 2016. *Strengthening Public Financial Management Reform in Pacific Island Countries*. Washington, DC: World Bank.

<http://documents.worldbank.org/curated/en/303711471955858894/pdf/Output-080416.pdf>.