Education Public Expenditure Reviews for Eastern & Southern Africa:
The Good, The Bad & The Future

Sue E. Berryman and Fadila Caillaud
The authors want to thank three groups who contributed to this exercise. A Macroeconomics and Fiscal Management (MFM) and Governance Global Practice (GGP) team assessed trends in the coverage and analytic quality of PERs across multiple sectors and time. They presented the methodology that they had developed for their review and its conclusions in a Brown Bag Lunch on December 16, 2015: Public Expenditure Reviews: New database and stock-take. Where should we go from here? Their approach to this exercise helped us refine our approach, and we were able to use their findings to create context for what we found in our review of recent education PERs for East/South Africa. Lev Freinkman, Marijn Verhoeven, Nicola Smithers, and Nazmul Chaudhury were particularly helpful, and Zeljko Bogetic shared papers relevant to the exercise.

The authors also want to thank those involved in the Brown Bag Lunch on April 11, 2016, where the initial findings of this report were presented and discussed: Education PERs, The Good, The Bad, and The Future--Experiences from Eastern and Southern Africa. Sajitha Bashir, Practice Manager, Education Global Practice; Amit Dar, Director, Strategy and Operations, HD Practice Group; Hiroshi Saeki, Senior Economist, Education Global Practice; and Samer Al-Samarrai, Senior Economist, Education Global Practice added to the authors’ presentations and engaged the participants in the Brown Bag in a lively and thoughtful discussion; Celia A Dos Santos Falas, Program Assistant, Education Global Practice, and Ma. Lorelei L. Lacdao, Program Assistant, Education Global Practice organized the logistics for the Brown Bag.

Finally, we would like to thank the two peer reviewers of the draft report: Samer Al-Samarrai, Senior Economist, Education Global Practice, and Lars Søndergaard, Program Leader, Human Development and Poverty, South East Asia Country Management Unit. Their comments were constructive and fruitful and led to an improved final report.

ACKNOWLEDGEMENTS

The authors want to thank three groups who contributed to this exercise. A Macroeconomics and Fiscal Management (MFM) and Governance Global Practice (GGP) team assessed trends in the coverage and analytic quality of PERs across multiple sectors and time. They presented the methodology that they had developed for their review and its conclusions in a Brown Bag Lunch on December 16, 2015: Public Expenditure Reviews: New database and stock-take. Where should we go from here? Their approach to this exercise helped us refine our approach, and we were able to use their findings to create context for what we found in our review of recent education PERs for East/South Africa. Lev Freinkman, Marijn Verhoeven, Nicola Smithers, and Nazmul Chaudhury were particularly helpful, and Zeljko Bogetic shared papers relevant to the exercise.

The authors also want to thank those involved in the Brown Bag Lunch on April 11, 2016, where the initial findings of this report were presented and discussed: Education PERs, The Good, The Bad, and The Future--Experiences from Eastern and Southern Africa. Sajitha Bashir, Practice Manager, Education Global Practice; Amit Dar, Director, Strategy and Operations, HD Practice Group; Hiroshi Saeki, Senior Economist, Education Global Practice; and Samer Al-Samarrai, Senior Economist, Education Global Practice added to the authors’ presentations and engaged the participants in the Brown Bag in a lively and thoughtful discussion; Celia A Dos Santos Falas, Program Assistant, Education Global Practice, and Ma. Lorelei L. Lacdao, Program Assistant, Education Global Practice organized the logistics for the Brown Bag.

Finally, we would like to thank the two peer reviewers of the draft report: Samer Al-Samarrai, Senior Economist, Education Global Practice, and Lars Søndergaard, Program Leader, Human Development and Poverty, South East Asia Country Management Unit. Their comments were constructive and fruitful and led to an improved final report.
This review covers PERs done in the Eastern and Southern Africa region, for which I was Practice Manager at the time. I am gratified that this effort produced significant findings relevant to the Education Global Practice’s SABER School Finance Team, especially, and the new Education Finance Community of Practice that they have helped to create.

TheApparently variable coverage and quality of education PERs being conducted for East/South Africa stimulated my request for this report. I asked its authors to assess several questions about recently completed education PERs, PETS, and QSDS for East/South Africa.

- What topics did the PERs address?
- Could a comparative, regional database be created for the variables reviewed?
- Were the data analyses appropriate, given the issues identified and the quality of the data?
- What did these analyses find?
- Which were especially strong PERs and why?
- What did the assessment of these PERs imply about standards for good PERs that can guide practitioners?

The question also arose about whether the Bank was extracting the maximum value out of completed PERs. Three issues were particularly salient.

- If PERs measured core indicators, could a database be constructed that showed the distribution of countries on important public expenditure variables? If not, which variables should be selected as core, with their measurement required for all PERs?
- Are the findings of PERs being used in the policy dialogue with Governments?
- Are the Bank’s task teams using PER findings to shape the preparation of education projects?

The last two of these questions could not be adequately addressed for methodological reasons. However, the robust findings for the first question have stimulated discussions within the Education Global Practice’s SABER School Finance Team about the practicality of and limits to core indicators.

Sajitha Bashir
Education Practice Manager, Eastern Africa
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PER</td>
<td>Public expenditure review</td>
</tr>
<tr>
<td>PETS</td>
<td>Public expenditure tracking survey</td>
</tr>
<tr>
<td>QSDS</td>
<td>Quantitative service delivery survey</td>
</tr>
<tr>
<td>MFM</td>
<td>Macroeconomics and Fiscal Management Global Practice</td>
</tr>
<tr>
<td>GGP</td>
<td>Governance Global Practice</td>
</tr>
<tr>
<td>SABER</td>
<td>Systems Approach for Better Education Results</td>
</tr>
<tr>
<td>TVET</td>
<td>Technical vocational education and training</td>
</tr>
<tr>
<td>NGO</td>
<td>Non-governmental organization</td>
</tr>
<tr>
<td>UNESCO</td>
<td>United Nations Educational, Scientific and Cultural Organization</td>
</tr>
<tr>
<td>UIS</td>
<td>UNESCO Institute of Statistics</td>
</tr>
<tr>
<td>EMIS</td>
<td>Education management information system</td>
</tr>
<tr>
<td>OECD</td>
<td>Organization for Economic Co-operation and Development</td>
</tr>
<tr>
<td>MoF</td>
<td>Ministry of Finance</td>
</tr>
<tr>
<td>EGP</td>
<td>Education Global Practice</td>
</tr>
<tr>
<td>SACMEQ</td>
<td>Southern and Eastern Africa Consortium for Monitoring Educational Quality</td>
</tr>
<tr>
<td>PDO</td>
<td>Project development objective</td>
</tr>
<tr>
<td>TTL</td>
<td>Task team leader</td>
</tr>
<tr>
<td>PAD</td>
<td>Project appraisal document</td>
</tr>
<tr>
<td>KCSE</td>
<td>Kenya Certificate of Secondary Education</td>
</tr>
<tr>
<td>IGCE</td>
<td>International General Certificate of Secondary Education</td>
</tr>
<tr>
<td>PISA</td>
<td>Program for International Student Assessment</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>O&amp;M</td>
<td>Operations and maintenance</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross domestic product</td>
</tr>
<tr>
<td>ECD</td>
<td>Early child development</td>
</tr>
<tr>
<td>DPL</td>
<td>Development policy lending</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

Acknowledgements ............................................................................................................................... i  
Preface .................................................................................................................................................... ii  
Acronyms .............................................................................................................................................. iii  
Figures ................................................................................................................................................... iv  
Tables ................................................................................................................................................... v  
Box ....................................................................................................................................................... v  
Executive Summary .................................................................................................................................... vi  

I. Introduction ........................................................................................................................................... 1  
A. Why was this analysis conducted? ........................................................................................................ 1  
B. What theoretical approach and methodology were used? ...................................................................... 2  
  1. Conceptual approach ............................................................................................................................ 2  
  2. Methodology ....................................................................................................................................... 2  
C. What were the limitations to this study? .................................................................................................... 5  
D. Organization of the Report ................................................................................................................... 5  

II. Did the Content Coverage of PERs Vary? ............................................................................................ 6  
A. PERs tended to focus on a limited number of sub-sectors ...................................................................... 7  
B. PERs’ content coverage was highly variable ........................................................................................ 8  
C. PERs varied in how they measured the “same” variable ........................................................................ 9  
D. PERs varied significantly in the depth of their content coverage by country ...................................... 10  
E. The depth of coverage varied significantly across domain areas ....................................................... 11  
F. The PERs omitted a few issues completely ........................................................................................... 12  
  1. Role of the state .................................................................................................................................... 12  
  2. Financial sustainability ....................................................................................................................... 13  
G. Conclusions ......................................................................................................................................... 14  

III. Results for Data and Analytic Quality ............................................................................................... 15  
A. PERs used five types of data .................................................................................................................. 15  
B. The quality of different data sources varied ........................................................................................ 17  
C. PERs tended to conduct relatively simple statistical analyses ............................................................... 19  
D. The quality of the analyses was very uneven ....................................................................................... 19  
  1. Deepening the analysis ....................................................................................................................... 19  
  2. Determining the etiology of a problem revealed by the PER .............................................................. 19  
E. Some domains were particularly under-analyzed ............................................................................... 22  
F. Summary and conclusions ................................................................................................................... 22  

IV. Doing Better ....................................................................................................................................... 24  
A. Should PERs be expected to measure selected core variables? ............................................................. 24  
B. What questions should PERs answer? .................................................................................................... 24  
C. Be alert to the opportunities and limitations of different data sources ........................................... 25  
D. Conclusion .......................................................................................................................................... 26
V. Challenges

A. Data can be irretrievably terrible

B. The budget and/or time frame for conducting a PER can make it impossible to produce a good quality product

C. Using a PER to guide project preparation requires that the new project follow the PER close in time

D. The PER team may lack access to Government officials who can act on PER recommendations

VI. Recommendations

VII. Annexes

A. Annex 1. Documents reviewed and coded

B. Annex 2. Content Analysis Coding Sheet: Conceptual Domains and Detailed Specification of Variables within each Domain

C. Annex 3. Variation in How the "Same" Variable is Measured: Two Examples by Country

Figures

Figure 1: Percent of same variables measured by number of countries

Figure 2: Depth of coverage by country

Figure 3: Depth of coverage by domain area

Figure 4: Frequency of use of data source types across the East/South Africa sample of PERS

Figure 5: Variation by country in number of data sources used

Tables

Table 1: Sub-sectors assessed by PERs by country

Table 2: Countries for which domain areas were not addressed

Table 3: Funding available for each East/South Africa PER reviewed

Table 4: Universe of PERs, QSDS, and PETS reviewed and coded

Box

Box 1: Examples of how PERs improved data quality
EXECUTIVE SUMMARY

Why was this analysis conducted?

A sufficient number of education public expenditure reviews, quantitative service delivery surveys, and public expenditure tracking surveys had recently been completed for East and South African countries to explore several questions.

- What topics did the PERs address?
- Could a comparative, regional database be created for the variables reviewed?
- Were the data analyses appropriate, given the issues identified and the quality of the data?
- What did these analyses find?
- Which were especially strong PERs and why?
- What did the assessment of these PERs imply about standards for good PERs that can guide practitioners?
- Were the findings of PERs used in policy dialogue with Governments?
- Are the Bank’s task teams using PER findings to shape the preparation of education projects?
What theoretical approach and methodology were used? What were their limitations?

**Approach**

The conceptual framework for assessing the content coverage and analytic quality of PERs, QSDS, and PETS was based on the theoretical frameworks that underlie PERs: welfare economics and public economics. Specifics of the framework reflected the Bank’s literature on PERs, the 2004 PER Guidelines and the concept note for a World Public Expenditure Review Stocktaking undertaken by the Macroeconomics and Fiscal Management (MFM) and Governance (GGP) Global Practice teams.

**Methodology**

The sample of PERs, PETS, and QSDS evaluated consisted of those recently completed for the education sectors of Ethiopia, Kenya, Madagascar, Malawi, Mauritius, Seychelles, Sudan, Zambia, and Zimbabwe. All were published between 2013 and 2016.

Methods were developed to assess two basic questions: the document’s content coverage and the quality of its data analysis. The methods used by the MFM and GGP PER stocktaking team provided some guidance.

Content analysis of each document was used to assess its content coverage, with the content analysis coding sheet being developed inductively from an analysis of a small sample of PERs and modified as the coding proceeded. The final sheet had 11 domains, such as allocative and technical efficiency or equity of financing. PERs addressed multiple aspects of most domains, resulting in a total of 54 variables. Since the coding sheets were developed inductively, they could not show which domains were not covered by any of the PERs for any of the countries. The 2004 PER Guidelines for education and other Bank documents were used to identify omitted domains or variables.

Judgments about quality were based on assessments of the credibility of the data sources used in each PER’s analyses and on the analyses themselves. The quality of the data used in each PER was judged on the basis of the type of data source and any assessments of the data provided by the PER, including information on recoding.

In judging the data analyses conducted, the statistical method for each analysis was first coded—e.g., incidence table, cross-tabulation, regression analysis. Two dimensions of the quality of data analyses were then assessed. One was whether the analytic method fit the nature and quality of the data. The second dimension was whether an issue shown to be important was pursued analytically—or, at the least, was flagged as a priority for future analyses.

This exercise conducted only a narrow assessment of whether PER findings were used in project preparation and did not measure the use of findings in the policy dialogue.
Did the content coverage of PERs vary?

PERs were not expected to cover all of the same topics. Countries differ in their financing issues for the sector and sub-sectors. The availability and quality of data, the PER budget, and the time frame for the task all affect what a PER can realistically cover. For example, smaller budgets tend to require a focus on fewer sub-sectors, which affect the variables assessed.

The intent was to map the topics that PERs actually covered in order to determine two things:

1. Whether topics fundamental to a PER—e.g., the equity of financing—were omitted or under-addressed.

2. Whether the PER’s choices explicitly signaled an understanding of the theoretical context for PERs.

The content coverage of the documents was evaluated in five ways:

1. Did the PERs assess all or only a limited set of sub-sectors?

2. Did PERs all measure any core variables in the same way so that a comparative database could be created?

3. What was the depth of coverage by country? This reveals the comprehensiveness and depth of coverage by country.

4. What was the depth of coverage by domain? This reveals comprehensive versus skimpy coverage by domain.

5. What variables are not assessed or are under-assessed?
The analysis of content coverage yielded these findings.

- The overall results for the analysis of the content of the sample of education PERs for East/South Africa were strikingly similar to those that the MFM/GGP team found in their analysis of 76 PERs for multiple sectors across the six regions of the World Bank. The variability in coverage was high and often unexplained. Analyses showed that PERs' coverage of the 54 variables coded was highly variable. Thirteen of the 54 variables (24 percent) were assessed by only one country. About 30 of the variables (56 percent) were assessed by no more than three countries. At the other extreme, five of the 54 variables (nine percent) were assessed by all 9 countries.

PERs were expected to differ in the variables that they assessed because of variable country priorities, the specifics of financing problems, data availability, and the budget and time frame for the PER. However, the extent of the content variation between PERs was still surprising. More important, they did not reflect a shared framework in the sense of explaining why topics normally considered part of this core diagnostic had been excluded.

- PERs often conducted only a partial assessment of the education sector because (often for budget reasons) they had to focus on limited sub-sectors. Most focused on primary and secondary education.

Pre-school was the sub-sector most frequently omitted, sometimes because Government had just committed public funding to this sub-sector (Zambia); because pre-school was primarily privately funded by NGOs or households, with little Government data; or because preschool and primary financing could not be disaggregated.

Not only is an increasing share of Bank funding being targeted on these sub-sectors, but both pose important expenditure questions to answer about the financing role of the state, returns to education, and financing equity. African countries are tending to allocate an increasing share of the public education budget to higher education while they tend to under-fund TVET.

- Several domains were not well covered: equity of financing, economic returns to education, progression in school, instructional time, human resources, and infrastructure. This list includes several domains theoretically important for PERs, such as returns to education, equity of financing, and human resources. Given PERs' theoretical foundations that stress the impact of financing on the poor, the most important under-covered domain was equity of financing. Although all PERs addressed at least one of the six variables under this domain, about half of the PERs assessed only one of the six equity variables, with a benefits incidence analysis being performed for only four of the nine countries.

- The most flagrant omissions were the lack of attention to the role of the state versus the private sector and financial sustainability, given the sector's policy priorities and the country's demographic and macroeconomic projections.
How good were PERs' data sources and analyses?

PERs used five types of data:

- Data generated by country systems
- Data generated by national surveys
- Data generated by cross-national data systems
- Donor and international research reports
- New data collection

Appropriately, all PERs relied heavily on data generated by country systems and, to a somewhat lesser extent, on data from national surveys such as household surveys and census data. Three PERs (Madagascar, Sudan, and Zambia) included new data collection that substantially increased their analytic depth.

Information about data sources was disturbingly hit-or-miss, leaving readers with limited ability to assess the trustworthiness of findings. The data category with the most documented problems was data generated by country systems. Data generated by national surveys such as household surveys was judged to be relatively trustworthy, but that generated by cross-national data systems was judged to be mixed. OECD data, as from *Education at a Glance*, were judged to be of high quality because of the exceptional processes in place that produce them, but UNESCO Institute of Statistics (UIS) data were judged to be of quite uneven quality because they depend on country-specific EMIS data. The fifth data source, new data collection, was well documented and of good quality for the three countries with major new data collection.

The primary analytic problem with the PERs was not the lack of sophisticated statistical techniques, although such techniques were not much used. The bigger problem was the lack of analyses designed to pursue and interpret important descriptive findings. Better PERs were ones that deepened the analysis and ones that "drilled down" to interpret and understand the etiology of a problem revealed by the descriptive analysis. When PERs went deeper, the PER team was much more likely to have an empirical basis for proposing recommendations that could plausibly ameliorate the problem surfaced by the PER. The quality of a PER’s recommendations should affect the value of the PER to the policy dialogue with Government and to design choices for subsequent projects.

Although there were notable exceptions in each case, PERs did not adequately analyze three domains: sources, channels, and uses of funds; governance and financing arrangements; and the efficiency of inputs and internal efficiency.
How can we do better?

All of the PERs reviewed had strengths, and a few were strong across the board. However, the high levels of variation between them in content and analytic quality implied a lack of agreement in the Bank and in the education sector about core content and analytic standards for this core diagnostic.

This report was not intended to and does not presume to offer consensually-based guidelines on conducting education PERs. However, the review of the East/South Africa PERs surfaced points that can contribute to the task of the Education Finance Community of Practice: systematically updating and revising the guidelines and standards for conducting education PERs.

Should all PERs populate specified core data tables? In the interests of being able to compare countries on key variables, it would be theoretically desirable if the sector could agree on a limited set of core data tables that all PERs should populate. However, this report found that comparability across countries may be elusive for more than a very few variables such as public financing of education as a share of GDP or of the government budget.

What questions should PERs answer? The bigger problem is having a coherent framework to guide the conduct of PERs. What basic domains should be covered? What constitute standards for good analysis? There are a limited number of basic questions that PERs should address or whose exclusion should be explicitly explained. Questions being considered by the Education Finance Community are these:

1. Who finances education, how is it channeled, and what does it buy?
2. How much does the government spend and on what?
3. Is there an adequate public financial management system in place?¹
4. How much is enough (adequate)? In cases of decentralized financing, are there vertical and/or horizontal imbalances? Are the revenue-mobilizing capacities of sub-national units sufficient to cover their financing responsibilities?
5. What can be afforded in the medium and long term, given the country’s macroeconomic trajectory, the sector’s policy objectives, and its projected demographics (sustainability)?
6. Are public resources being used efficiently and effectively?
7. Does public spending protect equity? Does it reflect an appropriate education financing role for the state?

¹ If the education PER is part of a comprehensive PER, the MFM team will probably analyze the PFM system in depth. The education PER team may just need to flag issues of particular import for the education sector.
**PERs should assess the quality of all of its data sources.** PERs will have to rely on the five data sources identified in the East/South Africa sample: data generated by country systems, national surveys, data generated by cross-national data systems, donor and international research reports, and new data collection. Information about data sources in the East/South Africa sample was unacceptably hit-or-miss. A PER should include a concise methodological section that describes all primary data sources used; how the team assesses the quality of each data source and why; and actions that the team took—or was unable to take—to compensate for key missing data or messy data.

The review of the East/South Asia PERs surfaced lessons about each type of data source. For example, BOOST is most useful if the raw data from the Ministry of Finance (MoF) is sufficiently disaggregated. If planning to build a BOOST, PER teams should be aware that this is a major—and budget-draining—exercise. National surveys can be a goldmine, especially household consumption or income surveys for analyzing equity issues. New data collection is indicated when the issue is critical and the data required to assess it are unavailable or unusable.

---

**What challenges do PER teams confront?**

Frequently PER teams confront certain challenges. One is irretrievably terrible data. Another is a laughably inadequate budget for the task. A third is the poorly aligned timing between a PER and preparation of the next education project, even though a good PER can significantly contribute to the design of a new project. A fourth is a lack of access to Ministry of Finance government officials who can act on financing and efficiency recommendations of the PER. The report suggests ways of ameliorating each of these challenges.
# Recommendations

1. The Education Global Practice (EGP) should develop a consensus within the Practice about the content, analytic, and budget standards for PERs in the education sector and issue guidelines for these standards to all education practice managers with oversight responsibility for PERs and to all PER teams. Content standards do not mean that a PER must address each issue. It does mean that exclusions of key topics should be explained.

2. The guidelines should:

   a. Specify questions that the PER should evaluate, with advice on what tend to be better data sources for answering them.
   
   b. Identify a minimum budget floor for a PER.
   
   c. Identify the likely cost range for needed recoding and cleaning of data, especially of Ministry of Finance data.
   
   d. Identify the likely cost range for different types of new data collection that may be needed (e.g., QSDS/PETS; case studies).
   
   e. Give guidance that education practice managers can use in their dialogue with the country unit on the optimal timing of a PER—e.g., completion in time to influence the preparation of national budgets, or the design of the next education project.
   
   f. Identify the likely calendar time required to complete a PER under different conditions, such as: i) reasonable data and no new data collection; ii) no new data collection but considerable data recoding; and iii) new data collection.

3. The guidelines should include examples of good analyses for specific questions. For example, they should include good examples of analyses that "drill down" to locate the drivers of expenditure inefficiencies, such as those for teachers, infrastructure, textbooks, instructional time, progression in school, or budget execution rates.

4. The guidelines should alert practice managers to several issues that these managers should discuss with the relevant CMU prior to launching any PER, including:

   a. Expected date for a new education project in order to maximize the value of a PER for project design.
   
   b. Budget adequacy for the PER.
   
   c. Adequacy of the calendar time for conducting the PER.
   
   d. Help in involving the MoF during the PER and at the point of disseminating the PER’s final results and recommendations.
A number of education public expenditure reviews and a few quantitative service delivery surveys and public expenditure tracking surveys have recently been completed for East and South African countries. They provided a sufficient body of work to explore several questions.

What topics did the PERs address?

Could a comparative, regional database be created for the variables reviewed?

Were the data analyses appropriate, given the issues identified and the quality of the data?

What did these analyses find?

Which were especially strong PERs and why?

What did the assessment of these PERs imply about standards for good PERs that can guide practitioners?

Were the findings of PERs used in policy dialogue with Governments?

Are the Bank’s task teams using PER findings to shape the preparation of education projects?
B. What theoretical approach and methodology were used?

CONCEPTUAL APPROACH

The conceptual framework for assessing the content coverage and analytic quality of PERs, QSDS, and PETS was based on the theoretical frameworks that underlie PERs: welfare economics and public economics. Specifics of the framework reflected additional sources: the Bank’s literature on PERs; the 2004 Guidance for Preparing PERs in the Human Development Sector, especially the core guidance and the guidance for the education sector; and the January 2013 concept note for a World Public Expenditure Review Stocktaking proposed by the Macroeconomics and Fiscal Management (MFM) and Governance (GGP) Global Practice teams.

METHODOLOGY

The sample of education PERs, PETS, and QSDS evaluated consisted of those published between 2013-2016 for Ethiopia, Kenya, Madagascar, Malawi, Mauritius, Seychelles, Sudan, Zambia, and Zimbabwe. The 17 documents reviewed for these nine countries and their coding status are listed in annex 1. In two cases (Ethiopia and Kenya) the education PER was conducted in the context of a comprehensive PER. The Sudan case was a comprehensive PER, but without a stand-alone education PER. It focused on Sudan’s financing of its decentralized functions, which included education.

Methods were developed to assess each document’s content coverage and the quality of its data sources and data analysis, with the methods used by the MFM and GGP PER stocktaking team providing useful guidance. Two of the original questions--use of PER findings in policy dialogue with Governments and in project design--were best assessed by surveys that the budget for this task could not accommodate. Even had surveys been financially possible, their value was questionable for reasons discussed below.

Content coverage. Content analysis of each document was used to assess its content coverage. The content analysis coding sheet was developed inductively from an analysis of a small sample of PERs and modified as the coding proceeded. The final coding sheet (see annex 2) consisted of 11 broad content categories, or domains, and specific variables within each domain. Each document was coded for whether it addressed variables within the domains.

---


3 The Mauritius and Kenya education PERs were published in 2013; the Kenya comprehensive PER and the Seychelles and Sudan PERs, in 2014. The Ethiopia, Madagascar, and Malawi PERs were published in 2015; and the Zambia QSDS and PER and the Zimbabwe PER, in 2016.

4 The sample of comprehensive PERs was much too small and idiosyncratic to reveal any interesting patterns between stand-alone education PERs and ones conducted in the context of a comprehensive PER.
The financing categories consisted of financing (Where does the money come from? Where does the money go? What does it buy?); allocative and technical efficiency; and equity of financing. The education-specific categories consisted of learning outcomes; wage and employment returns to education; enrollment in school; progression in school; textbooks; instructional time; human resources; and infrastructure.

PERs addressed multiple aspects of most domains, resulting in a total of 54 variables. (See annex 2.) Textbooks and instructional time were the only two domains that had no sub-variables, and learning outcomes had only two sub-variables. However, human resources and financing both had nine sub-variables.5

Since the coding sheets were developed inductively, they could not show us which domains were not covered by any of the PERs for any of the countries. The Bank’s PER literature, especially the 2004 Guidance for Preparing PERs in the Human Development Sector, was used to identify domains omitted by all PERs reviewed.

Judgments about quality were based on assessments of the credibility of the data sources used in each PER’s analyses and on the analyses themselves. The quality of the data used in each PER was judged on the basis of the source and any data issues flagged by the PER. PERs used five types of data: those generated by country systems, national surveys, data generated by cross-national data systems, donor and international research reports, and new data collection. When the quality of data generated by any of these sources could not be assessed because of a lack of information, it was coded as “unknown”.

5 For example, for human resources the PER might assess one or more of 9 variables: student/teacher ratios; teachers’ attendance and effort on the job; teachers’ attrition rates; intra-system transfers of teachers (rates and destinations); incentives offered teachers; types of teachers (e.g., contract; community; regular); teachers’ subject matter knowledge; teachers’ instructional practices; and in-service training of teachers.
This included Ministry of Finance accounts, Education Management Information Systems (EMIS), national learning assessments, and sector-specific databases, such as school mapping and teacher databases. Judging the quality of data from these sources had to depend on comments and analyses in the PER itself that rated the consistency, coherence, and relative accuracy of the data. This information sometimes appeared in an annex to the PER. Even if the PER did not comment on data quality, several PERs used BOOST to build finance data bases of more consistent quality. It was assumed that when the PER used BOOST, the team had confronted and resolved as many data problems with the country's financing data as possible.

A third data source is cross-national data sources or data-generating systems, such as the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), the UNESCO Institute for Statistics, or OECD. For these data sources, the international literature on the likely data quality by source or by system was used.

Donor and international research reports, including the World Bank, reports by other donor partners, NGO-conducted studies, and the international research literature—constituted the fourth source. It was difficult to judge the credibility of data from this source. Agencies such as the World Bank and academic journals have standards and processes such as peer review requirements that at least create a floor on quality. However, small donor groups, such as NGOs, may lack such processes. Studies conducted by these groups can vary widely in quality, depending on the individual doing the study.

The final source was new data collection. New data collection was often in the form of case studies, but occasionally in the form of a PETS/QSDS whose findings fed into the PER. Since the PER team designed and at least managed the conduct of such studies, these PERs described their study designs and methods of data collection. This information could be used to judge quality.

National surveys

National surveys usually conducted by the statistics unit of a country, such as population censuses and household surveys, were a second source. The quality of such surveys tends to be reasonably adequate. Data are collected under the same protocol. The staffs of units conducting these surveys benefit from international experience with designing, administrating, and analyzing such surveys. Donors also often fund technical assistance to help such units professionalize the conduct of the surveys under their jurisdiction.

Cross-national data sources or data-generating systems

A third data source is cross-national data sources or data-generating systems, such as the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), the UNESCO Institute for Statistics, or OECD. For these data sources, the international literature on the likely data quality by source or by system was used.

Donor and international research reports

Donor and international research reports, including the World Bank, reports by other donor partners, NGO-conducted studies, and the international research literature—constituted the fourth source. It was difficult to judge the credibility of data from this source. Agencies such as the World Bank and academic journals have standards and processes such as peer review requirements that at least create a floor on quality. However, small donor groups, such as NGOs, may lack such processes. Studies conducted by these groups can vary widely in quality, depending on the individual doing the study.

New data collection

The final source was new data collection. New data collection was often in the form of case studies, but occasionally in the form of a PETS/QSDS whose findings fed into the PER. Since the PER team designed and at least managed the conduct of such studies, these PERs described their study designs and methods of data collection. This information could be used to judge quality.
In judging the data analyses conducted, the statistical method for each analysis was first coded--e.g., incidence table, cross-tabulation, regression analysis. Two dimensions of the quality of data analyses were then assessed. One was whether the analytic method fit the nature and quality of the data. Poor quality data are not improved by entering them into a regression analysis. Similarly, opportunities are missed if rich data are not exploited by relatively sophisticated analytic means.

The second dimension was whether an issue shown to be important was pursued analytically—or, at the least, was flagged as a priority for future analyses. Did the PER pursue the analytic trail to the point where practical recommendations to mitigate the problem could be made? For example, showing that teacher salaries are consuming almost the entire recurrent budget is not sufficient. Why is this happening? What can be done about it? If the PER lacked the data, budget, or time to pursue the issue analytically, did it underscore that collecting/analyzing the needed data was an urgent future priority?

C. What were the limitations to this study?

This exercise conducted only a narrow assessment of whether PER findings were used in project preparation and did not measure the use of findings in the policy dialogue.

The influence of a PER on project preparation should be greatest if a project starts preparation very soon after the completion of a PER. The longer the elapsed time, all else equal, the lower the probable relevance of the PER (or any analysis) and the less influence we should expect the PER to have on project design. We could determine if a project had been or is being prepared soon after the completion of a PER. If so, we could look for evidence that the PER had influenced preparation. Was the PER referred in the PAD? Did the PER data seem to be used, especially in the analysis of the sectoral and institutional context and in selection of the PDO? Since the quality of a PER’s recommendations affects the extent to which a project design can benefit from the PER, we could also assess the quality of recommendations in the PER.

However, systematic interviews with the relevant players are preferable for determining how a PER has influenced project preparation. These players are primarily in the Bank, such as the PER TTL, the TTL for a project prepared subsequent to the PER, and the education sector manager. Assessing how the PER has influenced the policy dialogue requires interviews. These players are occupants of positions in the Bank, in Government, and possibly among the development partners.

The budget for this task could not be stretched to cover the costs of designing and conducting interviews. The value of such interviews was also questionable for the effects on the policy dialogue. Indeed, only recently completed PERs were assessed. Thus, only their short-term influence on the policy dialogue could have been evaluated, not their longer term effects.

D. Organization of the Report

Chapters II and III present the main findings of the review of the East/South Africa PERs. Chapter II assesses coverage--commonality, depth, omitted variables, and under-covered variables. Chapter III assesses data sources, data quality, the statistical methods used by the PERs, and the quality of their analyses. Chapter IV focuses on the lessons learned from this review for improving the quality of education PERs. Chapter V highlights challenges that PER teams often face. Chapter VI concludes with recommendations.

6 In the Ethiopia case the PER was conducted in the context of project design. In the Zambia case a high quality QSDS/PETS and PER had just been completed prior to preparation of a new project.
II. DID THE CONTENT COVERAGE OF PERs VARY?

The content coverage of the documents is evaluated in five ways.\(^7\)

- **Did the PERs assess all or only a limited set of sub-sectors?**
- **Were all variables assessed by all PERs?** Specifically, what percent of the total variables identified were assessed by what number of the nine countries? This analysis reveals the possibilities for creating a comparative, regional database for the measured variables.
- **What was the depth of coverage by country?** This reveals the comprehensiveness and depth of coverage by country.
- **What was the depth of coverage by broad content categories?** This reveals comprehensive versus skimpy coverage by category.
- **What variables are not assessed or are under-assessed?**

---

\(^7\) In some cases, measures of the content coverage for a country are based on a single document, such as Malawi or Zimbabwe; in others, they are based on multiple documents, such as Kenya, Zambia, or Ethiopia.
A. PERs tended to focus on a limited number of sub-sectors

As Table 1 shows, the Ethiopia PER alone addressed all sub-sectors in some detail, with Zambia coming close except for the pre-school sub-sector. Although the Kenya and Seychelles PERs addressed all sub-sectors, their analyses were very sketchy.

Most PERs focused on primary and secondary education, with Malawi focusing on primary education. Pre-school was the sub-sector most frequently omitted. The Sudan PER focused on whether decentralized financing was working relative to decentralized functions, including education, and pursued this issue for primary and secondary education only. If one PER had reasonable statistics on pre-school, the other PERs only skimpily treated the sub-sector—in some cases, pre-school was primarily privately funded by NGOs or households, with little Government data, or preschool and primary financing could not be disaggregated.

Only two PERs adequately assessed TVET and higher education. Four PERs had virtually no assessment of TVET, and five PERs had only a sketchy assessment of higher education. Not only is an increasing share of Bank funding being targeted on these sub-sectors, but both pose important expenditure questions to answer about the financing role of the state, returns to education, and financing equity. African countries are tending to allocate an increasing and share of the public education budget to higher education while they tend to under-fund TVET.

II. Did the content coverage of PERs vary?
### Table 1: Sub-sectors assessed by PERs by country

<table>
<thead>
<tr>
<th>Country</th>
<th>Preschool</th>
<th>Primary</th>
<th>Secondary</th>
<th>Tivet</th>
<th>Tertiary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethiopia</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Kenya</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Madagascar</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Malawi</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Mauritius</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Seychelles</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Sudan</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Zambia PETS/QSDS</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Zambia PER</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
</tbody>
</table>

**KEY**

- ● Some analytic attention
- ○ Sketchy attention
- ● Virtually no attention

### B. The content coverage of the PERs was highly variable

Analyses show that these PERs’ coverage of content was highly variable. Although variability in coverage was expected, its extent was surprising. More disturbing was the lack of any broadly shared framework within which analytic choices were explicitly made.

As indicated in annex 2, the content analysis of the documents identified 54 separate variables. For each variable, the number of PERs assessing that variable for their countries was calculated. Figure 1 shows that almost 13 of the 54 variables (24 percent) were assessed by only one country. At the other extreme, five of the 54 variables (nine percent) were assessed by all nine countries. About 30 of the variables (56 percent) were assessed by no more than three countries. About 11 of the 54 variables (21 percent) were assessed by 4-6 countries, and about 12 of the 54 variables (23 percent) were assessed for 7-9 countries.
C. PERs varied in how they measured the "same" variable

The analyses also found that PERs varied in how they measured what, at one level, is the "same" variable. As a result, creating a comparative database across these PERs for any set of core variables was not possible.

The PERs for at least seven countries evaluated 23 percent of the same variables. These included: learning outcomes, number of children in school, gross enrollment rates, student/teaching ratios, budget execution rates, unit costs, cross-sectoral functional allocations, intra-sectoral functional allocations, economic allocations, the efficiency of expenditures, and variations in resources by level, location, family income, or other qualifiers.

However, PERs differed in how they measured these “shared” variables. PERs used different measurement instruments (e.g., national and thus non-comparable systems for assessing learning outcomes); different levels of education for which the measurement occurred (e.g., primary, but not secondary); different types of expenditure (e.g., recurrent expenditures but not capital expenditures); different measures of the variation in the allocation of resources (e.g., by province, by level, by family income) and in the nature of the resource itself (e.g., textbooks; school grants; teachers).

Annex 3 details how two variables were measured at the country level: budget execution rates that were assessed for all 9 countries and intra-sectoral allocations by function that all countries except Sudan assessed. These variables were selected because it was thought that they might be relatively comparably measured.

Even for these variables there were differences between countries at the detailed level of measurement.

In the case of budget execution rates...

- **1st PER** reports these rates for capital budgets only
- **2nd PER** reports them for recurrent budgets only
- **3rd PER** reports execution rates for primary and secondary education only.
For intra-sectoral allocations, one case reports the share going to primary education only; in another, to primary and secondary education only. The differences are not huge, and those that occur do not mean that the data provided are not useful. They are. However, they imply a much more fragmented database than had been hoped.

### D. PERs varied significantly in the depth of their content coverage by country

Figure 2 displays the depth of coverage by country for each main domain. It is assumed that country A had a greater depth of coverage than country B if: a) a domain was addressed for country A but not for country B; or b) a higher percent of the variables within a domain was measured for country A than for country B. Each bar of color for a country measures the percent of variables of a main variable addressed for that country.

This figure reveals a huge spread across the PERs in depth of coverage, both in terms of the domains addressed and the depth of coverage within a domain. Three PERs stand out as having the highest depth of coverage across the domains and variables within domains: Zambia, Madagascar, and Ethiopia. Sudan looks as though it performs poorly, but was in fact an excellent PER. It focused, not on education, but the governance and financing framework for public functions, including education.

**For intra-sectoral allocations, one case reports the share going to primary education only; in another, to primary and secondary education only. The differences are not huge, and those that occur do not mean that the data provided are not useful. They are. However, they imply a much more fragmented database than had been hoped.**

**Figure 2: Depth of coverage by country**

- **Zimbabwe**
- **Zambia**
- **Sudan**
- **Seychelles**
- **Mauritius**
- **Malawi**
- **Madagascar**
- **Kenya**
- **Ethiopia**

**Share of variables assessed by domain, cumulated across domains by country**

- **Learning outcomes**
- **Economic returns to education**
- **Enrollment in school**
- **Progress in school**
- **Textbooks**
- **Instructional time**
- **Human resources**
- **Infrastructure**
- **Financing**
- **Allocative & technical efficiency**
- **Equity of financing**
Table 2 shows the percent of countries for which a domain was not addressed at all, regardless of the number of variables within the domain. PERs for all nine countries addressed at least one dimension of four domains: human resources, financing, allocative and technical efficiency, and equity of financing. As expected, since it was not an education-focused PER, Sudan shows up as not addressing most of the education variables. However, a surprising number of education-focused PERs did not address these domains at all: economic returns to education, progression in school, textbooks, instructional time, or infrastructure.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Zero Attention</th>
<th>% with no attention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning outcomes</td>
<td>Sudan</td>
<td>11</td>
</tr>
<tr>
<td>Economic returns to education</td>
<td>Kenya, Madagascar, Malawi, Seychelles, Sudan, Zimbabwe</td>
<td>67</td>
</tr>
<tr>
<td>Enrollment in school</td>
<td>Sudan</td>
<td>11</td>
</tr>
<tr>
<td>Progression in school</td>
<td>Seychelles, Sudan, Zimbabwe</td>
<td>33</td>
</tr>
<tr>
<td>Textbooks</td>
<td>Ethiopia, Mauritius, Sudan, Zimbabwe</td>
<td>44</td>
</tr>
<tr>
<td>Instructional time</td>
<td>Ethiopia, Kenya, Malawi, Mauritius, Seychelles, Sudan</td>
<td>67</td>
</tr>
<tr>
<td>Human resources</td>
<td>All address at least one variable for this domain</td>
<td>0</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>Mauritius, Seychelles, Sudan</td>
<td>33</td>
</tr>
<tr>
<td>Financing</td>
<td>All address at least one variable for this domain</td>
<td>0</td>
</tr>
<tr>
<td>Allocative and technical efficiency</td>
<td>All address at least one variable for this domain</td>
<td>0</td>
</tr>
<tr>
<td>Equity of financing</td>
<td>All address at least one variable for this domain</td>
<td>0</td>
</tr>
</tbody>
</table>

E. The depth of coverage varied significantly across domain areas

Figure 3 calculates the extent to which the variables within a domain are covered across the nine countries. It thus shows domains that were well-covered versus those only skimpily covered for the countries reviewed. For example, returns to education had four specific measures of the domain. Each of two countries measured one of the four variables, and one country measured three of the four variables. However, six countries measured none of the four variables. Averaging the row percent for each of the four variables across the nine countries reviewed yields a coverage rate of 13.8 percent.

II. Did the content coverage of PERs vary?
This figure shows that, at over 90 percent, allocative and technical efficiency was the domain best covered. There were five sub-variables within this domain. Six of the nine countries addressed all five variables, and three of the nine countries addressed four.

Several domains, however, were not well covered: equity of financing, economic returns to education, progression in school, instructional time, human resources, and infrastructure. Some of these domains are theoretically important for PERs, such as equity of financing, returns to education, and human resources. Given that the theoretical foundations of PERs stress the impact of financing on the poor, the most important under-covered domain was equity of financing. Although all PERs addressed at least one of the six variables under this domain, about half of the PERs assessed only one of the six equity variables. For example, a benefits incidence analysis was performed for only four of the nine countries.

Some other issues were less under-covered than poorly covered, and these cases are discussed under the assessment of the analytic quality of PERs.

**F. The PERs omitted a few issues completely**

If a broad content domain or a variable within a domain emerged in the review of a PER, it was added to the coding sheets. As noted in the Introduction, this inductive strategy meant that issues that education PERs should have addressed but did not were not flagged by the content analysis sheets. The most flagrant omissions were the lack of attention to the role of the state versus the private sector and financial sustainability.

1. **Role of the State**

PERs are situated in the broader paradigms of welfare economics and public economics. These paradigms assume that aggregate public spending should be allocated to programs within and across sectors to maximize...
social welfare, including the impact on the poor and relative to the contribution that the private sector can make. Does government intervention in general and public expenditures in particular enhance efficiency and/or equity relative to the private sector?  

PERs reflected their theoretical foundations only partially, as evidenced by the fact that several PERs did not confront basic questions about the provision and the financing role of the state versus that of the private sector by level. This puzzling gap may be partly attributable to the fact that by focusing on primary and secondary education, most PERs conducted only a partial assessment of the education sub-sectors and thus did not focus on questions about allocative efficiency. For only two countries were TVET and higher education adequately assessed, with two having no assessment and about half receiving only a sketchy assessment.

The Ethiopia and Zambia PERs addressed the issue well. A few other PERs had data on public vs. private enrollments and/or number of schools. A few assembled data that, looked at together, raised major questions about financing roles. However, these PERs failed to go the next step—confronting the question of trade-offs between public and private provision and financing. Consistent with this coverage gap was the scant attention paid to economic payoffs to education.

2. Financial Sustainability

PERs tended to be static. They looked at the current status of the sector, not usually at any freight trains that might be bearing down on it. None examined demographic trends by age group to assess what these trends implied for the sector’s budget. Changes in birth and survival rates can imply the need to plan for a costly expansion of the system—or possible financing dividends that proactive management of the sector could reap from declining population growth rates. With some notable exceptions, PERs did not check the financing implications of planned Government policies or the potential budgetary implications of the country’s macroeconomic trajectory.

---

9 Public expenditures in many developing countries are still providing and financing the provision of private goods and services in the education sector which can be provided in the private market, while simultaneously under-providing or under-funding public goods with large externalities and benefits to the poor.

10 These included: a) the Zambia PER that estimated the budget implications of Government’s fully funding its policy of free secondary education; b) the Madagascar PER that estimated the budget implications of Government’s recent decision to progressively integrate community teachers into the civil service.
The overall results for the analysis of the content of the sample of education PERs for East/South Africa were strikingly similar to those that the MFM/GGP team found in their analysis of 76 PERs for multiple sectors across the six regions of the World Bank. The variability in coverage was high and often unexplained. PERs were expected to differ in the variables that they assessed because of country priorities, the specifics of financing problems, data availability, and the budget and time frame for the PER. However, the extent of the content variation between PERs was still surprising. More important, they did not reflect a shared framework in the sense of explaining why topics normally considered part of the PER core diagnostic had been excluded.

1. PERs often conducted a reasoned but nonetheless partial assessment of the education sub-sectors. Most focused on primary and secondary education, with only a third assessing the tertiary sub-sector in detail; only about a fifth, the TVET sub-sector. This restricted scope may partly account for the fact that several PERs did not confront basic questions about the provision and the financing role of the state versus that of the private sector by level and under-analyzed financing equity and economic returns to education. An increasing share of Bank funding is being targeted on TVET and higher education. The design of any TVET or higher education project should be informed by PER analyses of the financing role of the state, as indicated by Government’s allocations among the sub-sectors, returns to education, and financing equity.

2. There is a huge spread across the PERs in the depth of coverage, as measured by the domains evaluated and the depth of coverage within a domain, as measured by the percent of variables measured within a domain.

3. PERs also vary in how they measure what, at one level, is the "same" variable. As a result, creating a comparative database across PERs for at least core variables is not now possible. PERs had different objectives, and countries differed in how they defined the variables in question.

4. Several domain areas were not well covered: equity of financing, economic returns to education, progression in school, instructional time, human resources, and infrastructure. This list includes several domains theoretically important for PERs, such as returns to education, equity of financing, and human resources. Given PERs’ theoretical foundations that stress the impact of financing on the poor, the most important under-covered domain was equity of financing. Although all PERs addressed at least one of the six variables under this domain, about half of the PERs assessed only one of the six equity variables, with a benefits incidence analysis being performed for only four of the nine countries.

5. The most flagrant omissions were the lack of attention to the role of the state versus the private sector and financial sustainability.
III. RESULTS FOR DATA AND ANALYTIC QUALITY

The use by PER teams of different data sources used by PER teams, assessments of the quality of those sources, and judgments about the quality of their analyses are reported in this chapter.

A. PERs used five types of data

The introductory chapter indicated that PERs used five types of data: data generated by country systems, data generated by national surveys, data generated by cross-national data systems, donor and international research reports, and new data collection. Figure 4 shows the relative use of different data sources across the five types.

Appropriately, all PERs relied heavily on data generated by country systems and, to a somewhat lesser extent, on data from national surveys such as household surveys and census data.

Three PERs included new data collection that substantially increased their analytic depth:

MADAGASCAR  SUDAN  ZAMBIA
Figure 5 shows variation by country in the types of data sources used. There is no right or wrong in terms of the data sources used. Ethiopia, Madagascar, and Zambia were all strong PERs. However, Zambia and Madagascar used all five sources, while Ethiopia relied on only two data sources: data generated by country systems (EMIS and BOOST) and on a rich array of national surveys (2011 Ethiopian Welfare Monitoring Survey; 2013 Young Lives household survey; 2011 Ethiopia Household Income, Consumption and Expenditure Survey).

The PERs used donors’ research reports, but not always critically. In some cases, what could have been correlational relationships were interpreted as causal. PERs did not make much use of the international evaluation literature, such as meta-analyses of rigorous evaluations of the effects of inputs on students’ participation and learning outcomes. These studies provide important data on the likely effectiveness of different investments and the conditions under which effectiveness occurs. As such, they are useful for PERs’ analyses of the effectiveness and efficiency of inputs.

---

B. The quality of different data sources varied

Regardless of the data type, information about data sources was disturbingly hit-or-miss, leaving readers with limited ability to assess the trustworthiness of findings. Some PERs assessed some, but not all, of the main data sources used in their analyses. Other PERs were silent about the quality of their data sources—or had only oblique and casual comments that left the reader unclear about the extent and depth of data problems.

The data category with the most documented problems was the category of data generated by country systems. This category included Ministry of Finance budgets and expenditure data; Ministry of Education budgets and expenditure data; data generated by education management information systems (EMIS); and sector-specific databases, such as national learning assessments, school mapping and teacher databases; or databases specific to a sub-sector, such as TVET or higher education. EMIS and financing data, including BOOST data, were the most frequently singled out as problematic. PERs that dealt with sub-national data (e.g., Sudan and Ethiopia) noted quality problems with these data that seemed to exceed those for national data. Box 1 has good examples of measures that PERs took to improve the quality of data as much as possible.

Box 1: Examples of how PERs improved data quality

The Ethiopia PER noted that codes for BOOST data were not consistently applied, especially at the regional and woreda sub-national levels. Data were recoded as accurately as possible, but the team had to make some assumptions to do so. Donor funding funneled through MoFED (often to the Regional Bureaus of Finance and Economic Development, or BoFEDs) or to the line ministries such as the MoE was not consistently reflected in BOOST, and the team had to go directly to donor funding accounts to reconstruct this source of funding.

In terms of Government financing of education as a percent of GDP and of total public expenditures, the Madagascar PER identified and reconciled major inconsistencies across data sources in the share of total budget allocated to education. The PER team found variations between sources of up to 10 percentage points, with the direction of change not consistent from one source to the next, especially for the post-crisis period. It identified problems with estimating allocations by level of education—e.g., changes across time in program classification; salaries not broken down by program; inability to allocate administrative costs by program. It corrected the data required for these estimates as much as possible. It corrected labor costs to construct trends from 2006-13 in expenditures for labor, capital, and other recurrent costs.

FIRST CATEGORY OF DATA

In terms of learning assessment data, in some cases the data came from national assessments with no international documentation. For example, one PER reported learning results from SACMEQ, Uwezo and KCSE (Kenya Certificate of Secondary Education). Two of these constitute international data collection regimes with some known properties. However, the KCSE, which seems specific to Kenya, has unknown technical properties.

Footnote: SACMEQ (Southern and Eastern Africa Consortium for Monitoring Educational Quality) has been administered for years in multiple countries. Thus, design and administration problems should have surfaced and been ameliorated, given multi-country exposure and multiple administrations. Uwezo, which is Kiswahili for “capability”, is less known. Since 2009, Uwezo has implemented large-scale nationally representative household surveys to assess the basic literary and numeracy competencies of school-aged children across Kenya, Tanzania and Uganda.
SECOND CATEGORY OF DATA

The second category of data, data generated by national surveys such as household surveys, was judged to be relatively trustworthy. This was not because PERs provided any evidence that allowed a judgment call, but because of general knowledge about the nature of these surveys and the units administering them. Those collecting survey data work under uniform administrative protocols and training. Although the data collected will have variable quality, the variations are rarely as large as those associated with data provided by schools or districts. These surveys are also usually conducted by the national statistics unit of a country which has often benefited from international technical assistance in designing, administrating, and analyzing such surveys.

THIRD CATEGORY OF DATA

The third category of data, data generated by cross-national data systems, was judged to be mixed. International learning assessments such as SACMEQ, the International General Certificate of Secondary Education (IGCSE), or OECD’s Program for International Student Assessment (PISA), have known properties and are of adequate—in some cases, excellent—quality.

However, international statistics were more mixed. OECD data from Education at a Glance were judged to be of high quality because of the exceptional processes in place that produce them. However, UNESCO Institute of Statistics (UIS) data were judged to be quite uneven in quality. UIS collects data via one or more questionnaires sent annually or biennially to over 200 countries and territories around the world. UIS data are considerably more trustworthy than UNESCO’s education statistics prior to the establishment of the Institute of Statistics in 1999. Despite UIS’s persistent work with countries to improve the quality of their education data, UIS data remain very uneven because they ultimately stem from the same EMIS databases that the PERs often flagged as problematic. This is especially the case for data from low or lower middle income countries that were the appropriate comparators for some of the PER countries. Similarly, World Development Indicator data were judged to somewhat uneven, depending on the sources of the data for specific indicators.

FOURTH CATEGORY OF DATA

Donor and international research reports, the fourth data source, were conducted by the World Bank (e.g., previous World Bank PERs), other donor partners, NGOs, and the international research community. It was not always easy to judge the credibility of data from this source. Agencies such as the World Bank and academic journals have standards and processes such as peer review requirements that at least create a floor on quality. However, small donor groups, such as NGOs, may lack such processes. Studies conducted by these groups can vary widely in quality, depending on the individual doing the study. The only way to judge data from these studies is to read the original studies, which was not possible for this analysis.

FIFTH CATEGORY OF DATA

The fifth data source, new data collection, was well documented for the three countries that had major new data collection: Madagascar, Sudan, and Zambia. The quality of these data collection initiatives was good.

---

10 Estimating sources of financing and the equity of financing require expenditure data from household surveys. Three PERs did not use household surveys at all, although one of these three used EMIS data to estimate household expenditures by level of education. Of those PERs that did use household surveys, some PERs failed to exploit the data to conduct in-depth analyses of financing equity.
C. PERs tended to conduct relatively simple statistical analyses

The PERs conducted four types of analyses: univariate or bivariate incidence tables or charts; correlations or cross-tabulations; linear regressions of different types (estimates of econometric functions; ANOVA calculations; standard regressions); and simulations. For the nine countries about a dozen linear regressions of some type and two simulations were conducted. Almost all analyses were univariate or bivariate incidence estimates. Although not statically complex, these estimates displayed often hard-won data of significant interest.

D. The quality of the analyses was uneven

The primary analytic problem with these PERs was not the lack of sophisticated statistical techniques, although, as indicated, such techniques were not much used. The bigger problem was the lack of analyses designed to pursue and interpret important descriptive findings. The MFM/GGP team found the same issue in its large PER assessment, describing many PERs as “Very non-critical, descriptive”.

1

Deepening the analysis

In some cases the issue was whether the PER deepened the analysis. For example, the Ethiopia PER did not stop with just describing economic allocations between capital, wage recurrent, and non-wage recurrent categories. It went on to assess potential input mix imbalances by estimating the “r-coefficient” or the recurrent O&M expenditures that are required to sustain the public sector services per unit of investment in a public sector services facility. R-coefficients should be steady or increasing. However, the PER found that r-coefficients for the Ethiopia education sector were declining over time from 0.27 to 0.22, a finding that signals that added new capital investments are likely to face growing problems with O&M funding.

The Madagascar PER’s analysis of budget execution rates reflected a sophisticated examination of factors that can affect budget execution rates for the education sector and for Government as a whole. These included executed expenditures recorded in the government accounts that did not fully account for all realized expenditures, leading to over- or underestimated execution rates, and blockages such as quarterly regulation rates for executing budgets.

2

Determining the etiology of a problem revealed by the PER

In other cases, the issue was whether the PER pursued “the next question”. When analyses find disturbing findings, the PER has no solid empirical basis for proposing recommendations that can plausibly ameliorate them unless the PER has “drilled down” to interpret and understand the etiology of these findings. Obviously, budget, time, and/or available data may preclude

11 The r-coefficient should measure the labor and other operating costs (supplies, utilities, etc.), routine maintenance and repairs, and replacement or rehabilitation investments to sustain the productivity of the investment.
the work required to figure out why something problematic is happening. In this case, the PER must flag that such analyses have to be done in order to generate plausible solutions to the problem.

The PERs had excellent examples of determining what was driving problems highlighted by the analyses. For example, the Malawi PER created a time series of repetition rates by grade and gender. The series showed stubbornly high or in some cases worsening rates across time. From a public policy perspective, pupil-years that do not result in promotion constitute a waste of resources since the inputs consumed do not result in the desired output.

The PER first estimated the internal efficiency or output efficiency of each primary school, distinguishing between pupil-years that resulted in promotion and those that resulted in repetition or dropout. It found that the coefficient of output efficiency for all primary schools in Malawi had improved from 65 to 73 percent over the course of the past decade, attributable to substantial declines in dropout rates for all grades, but not to declines in repetition rates.

The PER then sought to determine drivers of repetition. Since schools substantially varied in their output efficiency, the PER could test for relationships between inputs and repetition rates. It compared schools with repetition rates of less than 5 percent with all other schools in the sample, demonstrating that schools with low rates of repetition had significantly better resource endowments. Specifically, it conducted regression analyses using the QSD Survey data and EMIS data for 170 schools in the sample, finding that at grade 1 the availability of classrooms and the amount of funds available for non-salary recurrent expenditures both had a statistically significant impact on the promotion rate.

The Zambia PETS/QSDS documented dismal student/textbook ratios for three core subjects for urban and rural schools at the primary and secondary levels. There was little difference in ratios between urban and rural areas for primary schools, regardless of subject. The ratios in urban primary schools were worse than for urban secondary schools, ranging from 5 to 6.5 students per textbook, depending on subject. However, the ratios for rural secondary schools were generally worse than for rural primary schools, with 7 students per math textbook being the worst.

The team determined that multiple factors accounted for the poor ratios. First, the current textbook budget could not fully cover procurement of all the textbooks needed for students. Second, the timing of curriculum development and the timing for procuring textbooks were misaligned. Textbook delivery was significantly delayed in 2013 because of new curriculum development and a lack of procurement capacity in the decentralized unit. As a result, the textbooks were procured centrally, but only after the revised curriculum had been published in the middle of the 2013 academic year. Third, the District Education Board Secretaries (DEBS) are expected to distribute the textbooks to primary schools, but they have no textbook delivery fund.

Good examples aside, PERs too often uncovered problems and then dropped the issue, leaving the reader asking why what had been uncovered was happening and leaving the PER unable to identify recommendations that could specifically target the problem.

Good examples aside, PERs too often uncovered problems and then dropped the issue, leaving the reader asking why what had been uncovered was happening and leaving the PER unable to identify recommendations that could specifically target the problem.

<table>
<thead>
<tr>
<th>Case 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>The education sector accounted for the lion’s share of the total spending at 5.8 percent of GDP and a quarter of the total public budget for the last five years, with a nominal increase in spending of 46 percent between 2010-11 and 2013-14. Neither the reasons for the increase nor its sustainability are explored.</td>
</tr>
</tbody>
</table>
The PER flagged a serious lack of fiscal space for non-personnel costs that should be financed by the public budget, with over 99 percent of the total budget for the Ministry of Primary and Secondary Education going to finance employment costs. Given only 1 percent of the budget for non-wage recurrent and capital costs, households had to contribute significant funds through a combination of levies and fees to ensure adequate operating spending and occasional capital investment by schools. As a result, schools serving communities with greater ability to mobilize private resources typically offered a superior quality of learning environment than those whose ability to pay was lower.

Given the already generous budget envelope for primary and secondary (6.5 percent of GDP and 20 percent of total public expenditure), the solution has to be getting the wage bill under control. The PER reports that the average student-teacher ratio was 35:1 in early child development (ECD) classes, 36:1 in primary school classes, and 23:1 in secondary schools. Acceptable ratios at the ECD and primary levels for a low income country can be debated. However, the PER correctly notes that the secondary school ratios, which are lower than those for high-income members of the OECD such as Chile, constitute a luxury that a low income country can ill afford. That said, the PER is unable to recommend specific actions to reduce the wage bill. It does not dig further to understand the drivers of these ratios (their policy/political basis? deployment issues?) or of the wage bill itself (high salaries? benefits?)

The PER found significant variation in pupil/teacher ratios between the lower and higher grades and between schools with pupil/teacher ratios averaged across the grades. In grades 7 and 8 the average PTR was 40 students per teacher. However, it was over 100:1 in grade 1 and 80:1 in grade 2. The average PTR across all grades of primary schools was lower than 40:1 in about 25 percent of the schools surveyed, in the range of 40:1 to 70:1 in about 40 percent of the schools, and above 70:1 in the remaining 35 percent of schools surveyed. However, the sources of this highly variable distribution are unknown.

Learning outcomes were reported as being weak for virtually all PER countries in the sample. If an objective of PERs is to assess whether aggregate public spending is allocated to programs that maximize social welfare, one might expect some effort by PERs to model causes of better learning, perhaps borrowing models from sector colleagues. The PER can then ask whether the sector’s financing reflects factors that seem to enhance learning. Some PERs had partial or implicit causal models of learning (e.g., Ethiopia Education PER), but without a clear rationale for the assumptions in these implicit models. For example, pupil/teacher ratios and pupil/classroom ratios were often assumed to affect learning outcomes, but without explicit tests whether variations in these variables were even associated with variations in learning outcomes, less alone causal of them.

---

12 The international evidence on the economic effects of more learning is overwhelming. Analyses of growth rate differences between 1960 and 2000 for eight regions of the world found: 1) that these differences were completely described by differences in the population’s acquisition of cognitive skills; and 2) that school attainment was unrelated to differences in growth rates. In other words, it is the cognitive skills that students acquire in school, not “seat time” in school, that powerfully affect individual earnings, the distribution of income, and economic growth (, Hanushek and Woessmann, 2007 and 2009).
E. Some domains were particularly under-analyzed

Three domains could have been better analyzed.

Domain 1: Sources, Channels, and Uses of Funding

Most PERs assembled data relevant to this issue. However, these data tended to be scattered across the PER, leaving the reader with no clear sense of who funded (Government, development partners, private households, other sources); the revenue sources for public education budgets; how the funding flowed—who got the money initially and secondarily (pass-through arrangements); or what the money bought. Assembling data such as these is challenging, but once in place, they become the backbone of the PER’s analysis of the sector’s financing.

Domain 2: The Governance and Financing Arrangements

In some cases the governance arrangements were described, but not always. When the PER was silent about sub-national responsibilities for providing and financing educational services, it was not known whether the country was not using decentralization arrangements, was in the process of decentralizing, or had decentralized arrangements in place. When decentralization was in place, PERs—with the notable exception of the Sudan PER—also often did not ask questions raised by decentralized financing arrangements. These include questions about vertical and horizontal fiscal imbalances and about the revenue-raising authorities and capacities of sub-national units.

Domain 3: Efficiency of Questions

Again with notable exceptions, PERs did not adequately analyze a range of efficiency questions. These include the distributional efficiency of inputs across schools, districts, or provinces, especially teachers, infrastructure (e.g., school location), and textbooks. Only one PER calculated the internal efficiency of one level of education (primary education).

F. Summary and conclusions

Appropriately, all PERs relied heavily on data generated by country systems and, to a somewhat lesser extent, on data from national surveys such as household surveys and census data. Three PERs (Madagascar, Sudan, and Zambia) included new data collection that substantially increased their analytic depth. PERs under-used international evaluation reports.

Information about data sources was disturbingly hit-or-miss, leaving readers with limited ability to assess the trustworthiness of findings. The data category with the most documented problems was the category of data generated by country systems. Data generated by national surveys such as household surveys was judged to be relatively trustworthy, but that generated by cross-national data systems was judged to be mixed. OECD data from Education at a Glance were judged to be of high quality because of the exceptional processes in place that produce them, but UNESCO Institute of Statistics (UIS) data were judged to be of quite uneven quality because they depend on country-specific EMIS data. Donor and international research reports were considered of reasonable quality if the source had standards and peer review processes that created a floor on quality. Otherwise quality depended on the standards of individual authors. The fifth data source, new data collection, was well documented and of good quality for the three countries with major new data collection.
PERs used five types of data:

- Data generated by country systems
- Data generated by national surveys
- Data generated by cross-national data systems
- Donor and international research reports
- New data collection

The primary analytic problem with the PERs was not the lack of sophisticated statistical techniques, although such techniques were not much used. The bigger problem was the lack of analyses designed to pursue and interpret important descriptive findings. Better PERs were ones that deepened the analysis and ones that drilled down to understand the etiology of a problem revealed by the PER. When PERs went deeper, the PER team was much more likely to have an empirical basis for proposing recommendations that could plausibly ameliorate the problem surfaced by the PER.

Although there were notable exceptions in each case, PERs did not adequately analyze three domains: sources, channels, and uses of funds; governance and financing arrangements; and the efficiency of inputs and internal efficiency.
All of the PERs reviewed had strengths, and a few were strong across the board. However, the high levels of variation between them in content and analytic quality implied a lack of agreement in the Bank and in the education sector about the content and analytic standards for this core diagnostic.

This report was not intended to and does not presume to offer consensually-based guidelines on conducting education PERs. However, the review of the East/South Africa PERs surfaced points that can contribute to the task of the Education Finance Community of Practice: systematically updating and revising the guidelines and standards for conducting education PERs.

A. Should PERs be expected to measure selected core variables?

In the interests of being able to compare countries on key variables, it would be theoretically desirable if the sector could agree on a limited set of core data tables that all PERs should populate. However, this report found that comparability across countries may be elusive for more than a very few variables such as public financing of education as a share of GDP or of the government budget.

B. What questions should PERs answer?

The bigger problem is having a coherent framework to guide the conduct of PERs. What basic domains should be covered? What constitute standards for good analysis? There are a limited number of basic questions that PERs should address or whose exclusion should be explicitly explained.13

13 For any topic, the PER team may adduce data to show why that the topic is not an issue for the country in question or for the agreed-upon scope of the PER. Or the team may find that the data needed to assess the topic do not exist and cannot be collected within the budget and time constraints for the PER. In this case the team should flag the topic for future analysis.
Questions being considered by the Education Finance Community are these:

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who finances education, how is it channeled, and what does it buy?</td>
<td></td>
</tr>
<tr>
<td>How much does the government spend and on what?</td>
<td></td>
</tr>
<tr>
<td>Is there an adequate public financial management system in place?</td>
<td></td>
</tr>
<tr>
<td>How much is enough (adequate)? In cases of decentralized financing, are there vertical and/or horizontal imbalances? Are the revenue-mobilizing capacities of sub-national units sufficient to cover their financing responsibilities?</td>
<td></td>
</tr>
<tr>
<td>What can be afforded in the medium and long term, given the country's macroeconomic trajectory, the sector's policy objectives, and its projected demographics (sustainability)?</td>
<td></td>
</tr>
<tr>
<td>Are public resources being used efficiently and effectively?</td>
<td></td>
</tr>
<tr>
<td>Does public spending protect equity? Does it reflect an appropriate education financing role for the state?</td>
<td></td>
</tr>
</tbody>
</table>

C. Be alert to the opportunities and limitations of different data sources

PERs will generally have to rely on the five data sources identified in the East/South Africa sample: data generated by country systems, national surveys, data generated by cross-national data systems, donor and international research reports, and new data collection. The review of the East/South Asia PERs surfaced lessons about each type of data source.

- **Explicitly discuss the quality of all of the main data sources used for the PER.** As noted, information about data sources in the East/South Africa sample was unacceptably hit-or-miss. A PER should include a concise methodological section that describes all primary data sources used; how the team assesses the quality of each data source and why; and actions that the team took—or was unable to take—to compensate for key missing data or messy data.

- **Data generated by country systems tend to be the most problematic.** For financing data, BOOST may help teams structure and clean up MoF data prior to analysis. Experience indicates that a BOOST is most useful if the raw data from MoF is sufficiently disaggregated. Teams need to be aware that building a BOOST is a major—and budget-draining—exercise. If the education PER is part of a comprehensive PER, the MFM team is most likely to create a BOOST that the education team can then use.

- The other major country system on which education PERs have to rely is the EMIS, special purpose data bases, such as school mapping databases, and national learning assessments. EMISs are problematic, especially in low capacity countries. The team needs to be alert to incentives in the system that encourage...
misreporting of the numbers. For example, per capita financing creates incentives to inflate enrollment numbers. In such cases, the PER team will have to triangulate between data sources to estimate values of school-specific variables, enrollment rates or teacher absenteeism rates.

- Designing a good learning assessment is quite technical, and administrating it properly benefits from experience with carefully specified processes. Although it is argued below that it is preferable to use well-established international learning assessments, national learning assessments can meet technical design standards. Relative to cross-national learning assessments, they can also be better aligned with the country’s curriculum. However, the technical quality of these assessments should be scrutinized.

- **National surveys can be a goldmine.** Although the team needs to determine quirks and data quality problems with any survey, national surveys tend to measure variables important for PERs. For example, household consumption surveys are essential to creating poverty maps and to measuring household payments for education. Census data are essential for estimating changes in the size of school age cohorts that the education system will have to accommodate. The two main types of **cross-national data systems** are cross-national learning assessments and statistical databases which can be used to show where a country sits within the range of practice.

- Where possible, use cross-national learning assessments to evaluate the sector’s learning outcomes. Such assessments provide comparators. They also tend to meet higher design standards and better tested administration and data cleaning procedures than national learning assessments.

- **Choose comparators for countries close to the PER country in GDP per capita.** Comparing a low income country to the practice and performance of upper middle income/upper income countries makes little sense unless the objective is to provide an upper bound on practice.

- **Before starting the PER, be sure to scan for research studies germane to the PER and PER country.** Even to conclude that little is available, it is important to check for relevant and recent research conducted by the World Bank, donor partners, NGOs, or the international research community.

- Strongly consider new data collection when the issue is critical and the data required to assess it are unavailable or unusable. In the Zambia case the last education project to go to the Board was in 2001. In anticipation of a new project, a QSDS/PETS and PER were prepared. The teams for both documents collected important new data that were used to identify bottlenecks and problems in the sector and to set priorities for a new project. When new data collection is virtually essential, the team should bargain on the Bank budget for the PER or seek substantial trust funding in order to cover its costs.

---

### D. Conclusion

The high levels of variation between the East/South Africa PERs in content and analytic quality indicate the lack of agreement in the Bank and in the education sector about content and analytic standards for education PERs. If it wants to raise the quality and reduce the variability in quality between PERs, the sector needs to update a theoretically and empirically well-grounded framework for conducting a PER that can guide the work of PER teams.
V. CHALLENGES

PER teams may encounter conditions that make it difficult, if not virtually impossible, to deliver a high quality PER and/or to maximize payoffs from a good PER.

A. Data can be irretrievably terrible

Occasionally the financing data required for a PER are of such terrible quality that no amount of recoding and data cleaning can render them worth the analytic effort. This is most likely to be the case in countries with decentralized financing where most of the financing action is at sub-national levels. In these cases the only solution is new data collection of some variety that sheds light on what resources are reaching the point of service delivery.

Designing and fielding a QSDS is one option, where the resources and their sources can be established at the level of the school.

This solution was used in Nigeria, where the survey found disturbing levels of ghost students, ghost teachers, and ghost school construction in the samples of schools relative to sub-national "official" records.

Head teachers in each sampled school were also asked who funded different inputs to the school. For each input, they could name the two primary funders. Thus, sources that paid only small amounts for a particular input were not selected by head teachers for that input. The results bordered on financing chaos across schools within a state. The data showed that in neither state were key inputs funded from a limited and coherent set of sources. Schools were patching together funding where they could find it.
Another good option was that pursued for the Sudan PER. The team conducted case studies of financing and sub-national resource mobilization for four states, building up financing data and information on taxation in each case.

In most cases the PER team just confronts limited missing or poor quality data. In these cases the team can engage in limited new data collection or flag the problem for future attention.

B. The budget and/or time frame for conducting a PER can make it impossible to produce a good quality product

Time and budget are always constrained, but sometimes they are so constrained that it is virtually impossible for the team to construct a PER of good quality. In all cases PERs—or Policy Notes—should be clear about their TORs, time frames, and budgets in order to determine if each should be held to the standards for a PER as a core diagnostic.

Within the limits of the Bank’s records, table 3 shows the budget for each PER evaluated in ascending order of total budget. Funding varied substantially among the nine PERs, with the low outliers being Mauritius and Seychelles. These budgets were one-tenth the budgets for PERs with budgets in the middle range.

Given their low budgets, the Mauritius and Seychelles cases had to be defined as Policy Notes only.

Such Notes can certainly be helpful in the policy dialogue, may include some variables standardly addressed by PERs, but they cannot substitute for proper PERs. The teams for these two cases could not possibly ensure proper coverage and analytic depth, given such budgets. As it was, the unit had to cross-subsidize these activities from other activities.

The high outliers were Sudan and the Zambia PER + PETS/QSDS. Both Sudan and Zambia involved considerable new data collection. However, Madagascar, which also involved extensive recoding of financing data and new data collection, is at the low end of the cost distribution.
C. Using a PER to guide project preparation requires that the new project follow the PER close in time

The team for this report was asked if the findings of PERs were being used in the policy dialogue with Governments and if the Bank’s task teams were using PER findings to shape the preparation of education projects. This exercise conducted only a narrow assessment of the second question. Assessing the first question—and the second question more fully—required interviews that were beyond the scope of the budget for this exercise.

The influence of a PER on project preparation should be greatest if a project starts preparation very soon after the completion of a PER. The longer the elapsed time, all else equal, the lower the probable relevance of the PER (or any analysis) and the less influence we should expect the PER to have on project design. We could determine if a project had been or is being prepared soon after the completion of a PER. If so, we could look for evidence that the PER had influenced preparation. Was the PER referred in the PAD? Did the PER data seem to be used, especially in the analysis of the sectoral and institutional context and in selection of the PDO?

Three PERs influenced/are now influencing the preparation of four projects: Ethiopia, two small projects in Sudan, and Zambia. The Malawi PER may have influenced the Malawi 2016 education project.14 The other

---

14 The Malawi Education Sector Improvement Project went to Board at the end of August, 2016. No PAD has been archived, but the PID uses data from the PER.
six countries either had no education project subsequent to the PER/policy note or documents showed no evidence of influence. PER schedules and project preparation schedules often seem to be on separate tracks without consideration of how a good PER can facilitate project preparation. For example, the Madagascar PER was started in 2013, and a new project will be approved in 2017 or 2018. Mauritius has no education-specific operation, and the most recent DPLs do not include education triggers.

Since the quality of a PER’s recommendations affects the extent to which a project design can benefit from the PER, it should be noted that the quality of recommendations varied within the same PER and certainly across PERs. Some recommendations were feasible and specific, with the best flagging political economy and other implementation considerations. Others bordered on the vacuous.

D. The PER team may lack access to Government officials who can act on PER recommendations

If a PER is high quality, it can develop concrete and feasible recommendations that can be expected to mitigate financing or efficiency problems in the sector. The content of a recommendation determines under whose jurisdiction it falls, but most recommendations in education PERs will either fall under the jurisdictions of the MoF or of the MoE or both. The team will have access to the country’s education policymakers. However, especially if the education PER is part of a comprehensive PER, the sector-specific PER teams often have no access to a main interlocutor for PERs—the Minister of Finance.

The Madagascar PER team found a way to ensure that the Ministry of Finance was thoroughly aware of expenditure issues, the PER’s results, and its recommendations. It worked with Government to create a steering committee for the work conducted for the PER. The Ministry of Finance chaired the steering committee.
VI. RECOMMENDATIONS

This review confirmed the results of an earlier analysis of 76 PERs for multiple sectors across various regions done by the Macroeconomics and Governance Practices: high and unexplained variability in content and analytic quality. These results indicate a system problem, not a problem with a specific PER team or an education-specific problem.

The Education Global Practice (EGP) should develop a consensus within the Practice about the content, analytic, and budget standards for PERs in the education sector and issue guidelines for these standards to all education practice managers with oversight responsibility for PERs and to all PER teams. Content standards do not mean that a PER must address each issue. It does mean that exclusions of key topics should be explained.

The guidelines should:

1. Specify questions that the PER should evaluate, with advice on what tend to be better data sources for answering them.
2. Identify a minimum budget floor for a PER.
3. Identify the likely cost range for needed recoding and cleaning of data, especially of Ministry of Finance data.
4. Identify the likely cost range for different types of new data collection that may be needed (e.g., QSDS/PETS, case studies).
5. Give guidance on the optimal timing of a PER—e.g., completion in time to influence the design of the next education project.
Identify the likely calendar time required to complete a PER under different conditions, such as: i) reasonable data and no new data collection; ii) no new data collection but considerable data recoding; and iii) new data collection.

The guidelines should include examples of good analyses for specific questions. For example, they should include good examples of analyses that "drill down" to locate the drivers of expenditure inefficiencies, such as those for teachers, infrastructure, textbooks, instructional time, progression in school, or budget execution rates.

The guidelines should include the names of a small number of experts in the sector available to help PER team leaders plan and execute PERs.

The guidelines should alert practice managers to several issues that these managers should discuss with the relevant CMU prior to launching any PER.

- **Expected date for a new education project in order to maximize the value of a PER for project design**
- **Budget adequacy for the PER**
- **Adequacy of the calendar time for conducting the PER**
- **Help in involving the MoF during the PER and at the point of disseminating the PER's final results and recommendations.**
Annex table 1 documents the universe of PERs, QSDS, and PETS reviewed and how they were treated in the coding. The table shows that all 17 documents were reviewed, but only 13 were coded. Three of the 4 documents not coded (documents 6, 7, and 13) were summaries of material presented in more detail in another document that was coded. Sudan’s Synthesis Report (document 13) had a more cogent statement of findings than the main document, and document 13 was the source of findings for this country. The fourth document was a health PER (document 9), which, although informative when read alongside the companion education PER, was not coded.

### Table 4: Universe of PERs, QSDS, and PETS reviewed and coded

<table>
<thead>
<tr>
<th>Coding #</th>
<th>Document</th>
<th>Coding Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ethiopia Public Expenditure Review (published, November 2015)</td>
<td>Coded chapter 1 that analyzes Ethiopia’s fiscal position. This chapter had data germane to education for 2 variables. Chapters 2-5 that summarize findings for sector-specific PER background papers on health, social protection, water and sanitation, and education were not coded.</td>
</tr>
<tr>
<td>2</td>
<td>Ethiopia Education Financing Review (October 2015)</td>
<td>Coded</td>
</tr>
<tr>
<td>Coding #</td>
<td>Document</td>
<td>Coding Status</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>3</td>
<td>Kenya Comprehensive Public Expenditure Review: Eye on Budget (2013). The main education PER findings are reported here.</td>
<td>Coded</td>
</tr>
<tr>
<td>6</td>
<td>Madagascar 2014 Public Expenditure Review Education and Health (policy note in power point)</td>
<td>Reviewed, but not coded.</td>
</tr>
<tr>
<td>7</td>
<td>Madagascar 2015 Review of Public Expenditures in Social Sectors (Executive Summary)</td>
<td>Reviewed, but not coded.</td>
</tr>
<tr>
<td>8</td>
<td>Madagascar 2015 PER in Education (June, 2015)</td>
<td>Coded</td>
</tr>
<tr>
<td>9</td>
<td>Madagascar 2014 PER in Health (September 2015 version)</td>
<td>Reviewed, but not coded.</td>
</tr>
<tr>
<td>10</td>
<td>Malawi Primary Education PER (published, March 2015)</td>
<td>Coded</td>
</tr>
<tr>
<td>11</td>
<td>Mauritius Policy Note: Building Analytical Capacity to Raise Public Sector Efficiency (published, June 2013)</td>
<td>Coded</td>
</tr>
<tr>
<td>12</td>
<td>Seychelles: Programmatic Public Expenditure Policy Note for Health, Education, and Investment Management (March 2014)</td>
<td>Coded</td>
</tr>
<tr>
<td>13</td>
<td>Sudan State-Level Public Expenditure Review. Vol. I. Synthesis Report (published, May, 2014)</td>
<td>This document summarizes background papers in document 14. It was reviewed, but only its findings for the detailed background papers in document 14 were coded.</td>
</tr>
</tbody>
</table>
### B. Annex 2. Content analysis coding sheet: conceptual domains and detailed specification of variables within each domain

<table>
<thead>
<tr>
<th>Conceptual domain</th>
<th>Detailed specification of variables within the domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning outcomes</td>
<td>(Trends in) Learning outcomes (by level, subject, other qualifiers, determinants)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Determinants of learning outcomes</td>
</tr>
<tr>
<td>Wage and employment returns to education</td>
<td>(Trends in) Wage returns to education (by level, other qualifiers such as economic sector)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Employment effects of education</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Responses of students to labor market wage and employment signals (by level, field, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Skill shortages</td>
</tr>
<tr>
<td>Enrollment in school</td>
<td>(Trends in) Number of children in school (by level/type, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Number of children out of school (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Gross enrollment rate (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Net enrollment rate (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Net intake rate for primary education (by qualifiers)</td>
</tr>
</tbody>
</table>

---

15 Qualifiers include gender, residence (e.g., rural versus urban, province, district), family economic status, public versus private school.
<table>
<thead>
<tr>
<th>Conceptual domain</th>
<th>Detailed specification of variables within the domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progression in School</td>
<td>(Trends in) Overage entry into school</td>
</tr>
<tr>
<td></td>
<td>School participation status of school-aged cohort (by level, age, other qualifiers)</td>
</tr>
<tr>
<td>Textbooks</td>
<td>(Trends in) Internal efficiency</td>
</tr>
<tr>
<td>Instructional time</td>
<td>(Trends in) Attendance rate by those enrolled (by level, other qualifiers) (^\text{17})</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Dropout rate (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Repetition rates (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Uninterrupted progression through grades (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Shows joint effect of dropping out and grade repetition)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Transition or promotion rates (between grades, levels, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Completion rate (by level and other qualifiers)</td>
</tr>
<tr>
<td>Human resources</td>
<td>(Trends in) Student/textbook ratios (by level, other qualifiers such as subject)</td>
</tr>
<tr>
<td>Instructional time</td>
<td>(Trends in) Instructional hours (by level, by shift, by day/week/year, other qualifiers such as policy versus actual)</td>
</tr>
<tr>
<td>Human resources</td>
<td>(Trends in) Student/teacher ratios (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Teacher attrition (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Intra-system transfers (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Types of teachers(^\text{18}) (by level, other qualifiers such as credentials)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Participation in in-service training (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Teacher subject knowledge</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Teacher effort (attendance; time worked when attending; other qualifiers such as multi-grade teaching and multiple shift scheduling)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Teacher instructional practices and behavior</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Teachers' incentives and motivation</td>
</tr>
</tbody>
</table>

\(^{16}\) Attendance is defined as the average number of total school days that those enrolled in school actually show up in school.

\(^{17}\) Differences in level and source of financing, benefits, who hires, and qualifications define variations within the teaching corps.
<table>
<thead>
<tr>
<th>Conceptual domain</th>
<th>Detailed specification of variables within the domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infrastructure</td>
<td>(Trends in) Number of schools (by grade, by level, other qualifiers such as % running double or triple shifts)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Student/classroom or student/school ratios (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) % of classrooms that are multi-grade (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Average distance between households and nearest school (by level, by other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Physical characteristics of schools (by level, other qualifiers)</td>
</tr>
<tr>
<td>Financing</td>
<td>(Trends in) Sources, channels, and use of funds</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Private household funding of education (by level, inputs financed)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Shares of public versus household financing of education and school use of parental contributions</td>
</tr>
<tr>
<td></td>
<td>Trends in) Decentralization or deconcentration of functions and financing in education sector</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Public financial management</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Budget formulation process (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Budget credibility, comprehensiveness, and transparency</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Budget execution rates for education sector (by level, other qualifiers such as requirements for release of funds)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Unit costs (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Estimated sustainability of publicly financed inputs (by level, other qualifiers)</td>
</tr>
<tr>
<td>Allocative and technical efficiency</td>
<td>(Trends in) Cross-sectoral functional allocations: government financing of sector as % of GDP, % of total public financing (nominal versus real; other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Intra-sectoral functional allocations (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Economic allocations (by level, expenditure types, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Teacher salaries as share of total (or recurrent) education budget (by level, other qualifiers)</td>
</tr>
<tr>
<td></td>
<td>(Trends in) Efficiency of expenditures (by level, by inputs, by inputs relative to outcomes)</td>
</tr>
<tr>
<td>Conceptual domain</td>
<td>Detailed specification of variables within the domain</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------------------------------</td>
</tr>
<tr>
<td>Equity of financing</td>
<td><em>(Trends in)</em> Allocations relative to population shares of sub-national units (by total financing, specific inputs such as labor, capital, textbooks)</td>
</tr>
<tr>
<td></td>
<td><em>(Trends in)</em> Allocations to or by geographic units relative to their poverty rates</td>
</tr>
<tr>
<td></td>
<td><em>(Trends in)</em> Variation in resources (financing or inputs) by level, geographic unit, type of inputs, family income, other qualifiers</td>
</tr>
<tr>
<td></td>
<td><em>(Trends in)</em> Private (household) financing of education relative to income quintile (by level, other qualifiers such as estimates of opportunity costs by income quintile)</td>
</tr>
<tr>
<td></td>
<td><em>(Trends in)</em> Benefits incidence analysis (by family income, other qualifiers such as gender)</td>
</tr>
<tr>
<td></td>
<td><em>(Trends in)</em> Public costs of providing pro-poor free education (by level, other qualifiers)</td>
</tr>
</tbody>
</table>

C. Annex 3. Variation in how the "same" variable is measured: Two examples by country

<table>
<thead>
<tr>
<th>Country</th>
<th>Budget rates</th>
<th>Intra-sectoral allocations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethiopia</td>
<td>Capital budget execution rate (2009-2012). No hard data on budget execution are presented for salaries or non-wage recurrent expenditures, other than to say that both are &quot;well protected&quot;. It is not known whether &quot;well protected&quot; refers to the execution rates, to the approved budgets, or both.</td>
<td>Estimates shares of total, recurrent, and capital expenditures by level, possibly for 2004-2012. (The years in figures 3.4-3.6 are all mislabeled as 1996-2004).</td>
</tr>
<tr>
<td>Kenya</td>
<td>Shows execution rates for education in total. It reports execution rates for recurrent vs. development (capital) budgets for the &quot;human resource development&quot; sector. This seems to include education, social protection, and health, which means that the execution rates by types of expenditures cannot be isolated for education.</td>
<td>Reports trends from 2003/04 to 2010/11 for total spending in nominal and real terms and per student public spending in real terms by education function (sub-sector; education vs. administration).</td>
</tr>
<tr>
<td>Country</td>
<td>Budget rates</td>
<td>Intra-sectoral allocations</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Madagascar</td>
<td>Trends from 2006-2013 in rates for education (Current operations: regular wages; Other current operations; Internally financed Investment program; Externally financed Investment program) relative to overall government budget</td>
<td>Allocation of recurrent education expenditure by level and function, 2002-2013; Allocation of Current Non-Wage Expenditures by Level, 2002-2013. (Weakness in investment data precludes functional allocation analyses.)</td>
</tr>
<tr>
<td>Malawi</td>
<td>Execution of (revised) education budget for 2011/12 to 2013/14 (recurrent and investment budgets)</td>
<td>The PER decomposes recurrent expenditures within primary education from 2006/07 to 2013/14. No analysis of the shares going to other levels of education. Development or capital expenditures on education are not classified by level of education in Malawi’s financial accounts. Thus, without reconstituting the data, which this PER did not do, it cannot estimate the total share of allocations going to primary education relative to other levels of education.</td>
</tr>
<tr>
<td>Mauritius</td>
<td>Shows budget execution rates for 2010 and 2011 relative to original allocations and the budget as revised during the year. Differences in execution rates for recurrent vs. capital budgets discussed only verbally.</td>
<td>Reports from BOOST actual expenditures by function.</td>
</tr>
<tr>
<td>Seychelles</td>
<td>Calculates % executed for allocated recurrent budget items. No data on execution of capital budgets.</td>
<td>Figure 38 is supposed to show 2011 functional allocations by sub-sector. The figure is missing values, but the text reports approximate data.</td>
</tr>
<tr>
<td>Sudan</td>
<td>Sudan is a federalist system. Calculates state expenditure out-turns as % of original approved budget for 4 case study states by type of expenditure (wages/ salaries; goods and services; development). For two states the trend line is 2005-2012; for one state, 2006-2012; for the 4th state, 2006-2011.</td>
<td>Good estimates of inter-sectoral allocations, but none for intra-sectoral allocations by level of education and function.</td>
</tr>
<tr>
<td>Country</td>
<td>Budget rates</td>
<td>Intra-sectoral allocations</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Zambia</td>
<td>For 2013 calculates by level of education and administration, authorized provision, funded budget, expenditures, funded budget rates, and budget execution rates, finding a considerable gap between authorized provision and funded budget but little gap between funded budgets and execution rates. Also calculates the execution rates of primary school grants that are funneled through District Education Board Secretaries (DEBS) and of execution rates of secondary school grants.</td>
<td>Assembles trend data for 2006-2015 in government funding of education by function in nominal and constant terms and as shares of total by function by year.</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>Rates reported for 2014 for primary and secondary education. Displays budget execution rates for non-employment expense categories (non-salary recurrent and capital expenditures); text reports execution rates for salary items.</td>
<td>Reports shares going to primary (which seems to include two years of preschool) and secondary education only, not for TVET, post-secondary education, or administration.</td>
</tr>
</tbody>
</table>