

# **Handout 12-4.1:**

## **Kinerja—Local Governance Service**

### **Improvement Program Impact Evaluation**

#### **Case Study**

#### **Answer Key**

---

#### **CASE INTRODUCTION**

Over the course of the past decade, the Indonesian government has witnessed a largely successful democratic transition. However, Indonesia still faces some shortcomings and issues that could make it vulnerable to reversion, including weak rule of law, low levels of transparency and accountability, inadequate representation and corruption. In October of 2010, USAID awarded a 4.5 year, \$25 million contract to a consortium of 6 organizations to improve service delivery, and strengthen accountability and good governance among local governments. The project will work in 20 districts (5 districts each in 4 provinces) and focus on three sectors: health, education and business enabling environment.

Specific interventions are chosen through a consultative process, whereby the project team proposes 2 or 3 “packages” of pre-identified interventions for each sector, and the local governments choose the intervention(s) and sectors that best match their development needs/priorities. Because of the demand-driven nature of the selection process, the activities in any district may or may not cover all 3 sectors. Once a decision on interventions is made, the majority of implementation will be carried out through grants to local partners (CSOs, etc).

Kinerja’s M&E system calls for two Impact Evaluations.

#### **KEY CHALLENGES FOR AN IMPACT EVALUATION**

- **Governance Project, Sectoral Focus**

Though Kinerja will work with local governments to improve public service delivery in three sectors, the project will not directly affect services in these sectors. It is designed to *indirectly* influence outcomes by attempting to address governance issues presently inhibiting effective service delivery. Therefore, while many of the key measures of success for Kinerja are *sectoral* indicators, Kinerja activities are a step further removed from these indicators than the activities of a sectoral project would be.

- **Demand Driven Approach**

Kinerja is designed to tailor its approach based on local conditions and needs. While central to effective project implementation, this complicates the tracking of project results in two important ways. First, because project plans as they exist at any given time are potentially subject to significant change based on local conditions and demands, an M&(I)E system that lacks sufficient flexibility to adapt to those

changes may end up measuring performance in suboptimal ways. Second, because implementation at the district level is not standardized (i.e. there is not necessarily “homogeneity of treatment”) and districts select from a suite of interventions and sectors, specific program results may not be comparable across locations.

- **Evaluation Timing**

Also per the RFA, “impact evaluation results [should] accompany the Annual Reports at the end of the second year and the final year of the program.” Yet, the governance focus of the project means that there may be a lag between project activities and actual sectoral results. Furthermore, because of the time needed to collect and compile information relevant for the impact evaluation, evaluation results will reflect the situation not at the ends of the second and final years of Kinerja, but rather some months before.

- **Sample Size**

1. Kinerja will work in 20 districts and compare observed changes in those 20 districts to observed changes in 20 “Control” districts in order to identify project effects. However, with only 20 Treatment districts and 20 Control districts, it may be difficult to detect impact using statistical methods. The demand-driven approach further compounds the limited ability to detect impact using statistical methods: where Kinerja focuses on different issues in different districts, outcomes may also differ. Where similar outcomes exist only in a subset of districts, average effects across all 20 districts will be smaller, and thus even harder to detect.
2. The School-Based Management activity, which is one of the three ‘packages’ of activities offered, was selected by some of the districts and will be implemented in three districts in West Kalimantan province. Within each district, the project will work with 16 schools and track 16 control schools (randomly selected), leading to a total of 48 treatment and 48 control schools (16T/16C per district). For each school, data will be collected from: (1) one school principal, (2) one member of school committee, (3) one community leader, and (4) 10 randomly sampled parents of children enrolled in the school (grades 2-4).

## **PROPOSED M & (I)E DESIGN**

In light of these key challenges, The M&(I)E approach for Kinerja has been designed to comprise four discrete but integrated components:

1. Ongoing project monitoring focused on process, output, and outcome indicators.
2. A qualitative evaluation focusing on process and intermediate outcomes across a subset of treatment and control districts.
3. A quantitative, randomized evaluation of overall project effects across all 40 treatment and control districts.
4. A quantitative, randomized evaluation of the School-Based Management intervention that target randomly selected schools within districts that have selected support in this sector.

## TEAM TASKS

Discuss the following questions and be prepared to present to your colleagues:

1. Why are Components 1 & 2 of the M&IE design important? How would you expect them to complement the IE components?

**They help to explain and provide context for the IE designs. The monitoring system provides data for the IE, helps to monitor implementation, identifies issues in implementation, and helps to clarify the independent variable (if IE's are important for learning what works and what doesn't, we need to know what the 'what' is and monitoring helps us with that). The qualitative component is helpful for, among other reasons, providing context, understanding mechanisms of change, better understanding outcomes, exploring outliers, and identifying unintended effects.**

2. Regarding Component 3 (IE of overall program effects), which of the challenges are relevant to this design? Are there other challenges you anticipate?

**- Governance Project, Sectoral Focus: While the overall project is at heart a governance project, its primary outcomes are sectoral. It might not be reasonable to expect to see changes in sectoral indicators (for example, educational achievement or school attendance) at the district level for a program that works on school governance, especially when this is only one of three areas of support.**

**– Demand driven approach: Districts select which interventions and areas they are interested in. This means that only some areas will participate in the school governance component, while others will focus on the health component. Can we compare outcomes for districts which participate in different program components? It's certainly not fair to compare them on sector-based outcomes.**

**– Evaluation timing: For changes to appear at the district level, it might take some time. Expecting to measure changes after 1 year may be unrealistic, particularly if the focus is on governance rather than direct service delivery (which might theoretically demonstrate results more quickly in a given sector, although they may not be of the same magnitude or sustainability developed through a governance approach).**

**– Sample Size: A small sample size (20T and 20C) is exacerbated by the above issues and is a major threat to measuring statistically significant impacts of the project as a whole.**

3. Regarding Component 4 (IE of the School-Based Management component), which of the challenges are relevant to this design? Are there other challenges you anticipate?

**In this design, the first and third (governance focus and timing) still present potential threats. The demand driven approach is no longer an issue since we limit our focus to those districts implementing the school governance component. Also, sample size is less of an issue since within each district we**

have 16T and 16C schools leading to a total sample size of 96 schools (and many more families in the school catchment areas).

4. What are the main benefits and drawbacks of each design (Component 3 & 4)? Which of the two designs offer the highest learning potential?

**Component 3:** This allows us to understand the broader impacts of the program as a whole, but given the challenges listed above, our impact estimates may lack precision or certainty. Also, given the heterogeneity of treatment (i.e. different districts select different activities) we will be limited in our ability to learn about specific development hypotheses. This component is therefore better suited for accountability than learning.

**Component 4:** While this design allows us to more precisely and credibly estimate the impact of the school governance component, the big downside in this design is that it only measures the impact of one component of the project- the school governance component. We can't generalize these results to the areas where the health or business components were implemented. So this offers higher learning potential (in terms of testing and understanding a development hypothesis with precision and credibility), but is somewhat limited in terms of accountability since it does not seek to measure the impacts of the whole program (or whole set of activities being implemented).

5. What factors influenced your answer to Q4? What factors (of the project, context or evaluation designs) influence the likely success of the IE in this case?

The challenges listed above influence answer to Q4. Many of these challenges relate to the intersection of the project, context and evaluation design, and many could be avoided by early planning. For example, on the project side, because of the governance approach a longer time horizon for outcome measurement may allow for better measurement of change. Likewise, given the demand driven nature of the program and its effect on potential outcome measures and sample size, perhaps an IE of the overall program may not be recommended or perhaps during project planning there may have been opportunities for increasing sample size or scope of the project. From the evaluation side, it was clear that given the context and project design, the overall IE would not yield the type of clear, precise results desired (this was part of an IE pilot program which sought to test the suitability of IE designs for DG programs, so this was an important outcome), so an additional IE design was developed.

More generally, factors from each area (the project, context, and evaluation) may influence the success of the evaluation. Participants may suggest additional factors beyond the challenges listed above, including topics discussed in the slides, such as spillovers, attrition, non-compliance, etc.

#### **Bonus Questions:**

1. Why are these problems occurring? What are underlying causes of the challenges?

Principally, the issues are caused by trying to do an IE at 'too high' of a level. Generally, the broader the program, the more complicated the IE. Sample sizes tend to decrease (at least when the interventions are targeted at higher levels, such as districts rather than communities or households), homogeneity of treatment is difficult to ensure, and outcomes may vary, both in expectation and actuality, across locations or components. IEs of more focused programs (or components of programs) are typically better able to explore specific development hypothesis in a way that yields credible evidence.

2. What might be some threats to internal and external validity for the Kinerja IE design?

**Internal: spillovers, attrition, small sample size, non-compliance (treatment communities changing the type of services desired)**

**External: heterogeneity/lack of clarity of treatment, unclear causal pathway from governance to sector outcomes (for the broader IE), limited scope, both geographically and sectorally, only relevant for communities that choose school-based intervention so might not be generalizable (for narrower IE)**

3. How would you approach the political implications of randomizing a given intervention within districts?

**Transparency is critical, so ensuring that all parties are aware of their opportunity to participate helps gain buy-in. Doing public randomization also adds to transparency.**

4. Demand-driven project design can significantly complicate attempts at impact evaluation. Why?

**If multiple types of interventions (and sectors) are offered, we may expect different types of outcomes. If so, how do we compare impact of a schooling intervention to a health intervention? Moreover, If participants can select which type of intervention (including the sector) they participate in, there will likely be some selection bias and comparing outcomes (if we can agree on a set of outcomes to compare) for those that select the health intervention against those that select the business intervention may be biased (perhaps because those that select the health intervention may be starting from a generally lower level of development).**

5. What are some takeaways from the Kinerja case for your role as Evaluation Specialists helping to design and supervise USAID Impact Evaluations?

**The main takeaways include:**

1. **Importance of early planning, including identifying the objectives of the evaluation and then finding the right tools and design to achieve those evaluation objectives. There are usually**

**creative evaluation designs that allow us to evaluate in less than ideal circumstances, but they tend to yield less certain results or might not directly address the evaluation objectives.**

- 2. Focusing on broader projects can be difficult for IE. IEs tend to focus on smaller, narrower projects or components of projects.**
- 3. Understanding the strengths and weaknesses of IEs helps us know when they are most effectively used. They are like any tool in that they best address certain questions or objectives and shouldn't be seen as solutions to every question.**