



Impact Evaluation - Unraveling the opaque

Markus Olapade, 3ie

Arab Youth & Entrepreneurship
Conference

Doha, 16.2.2013

Why evaluate?

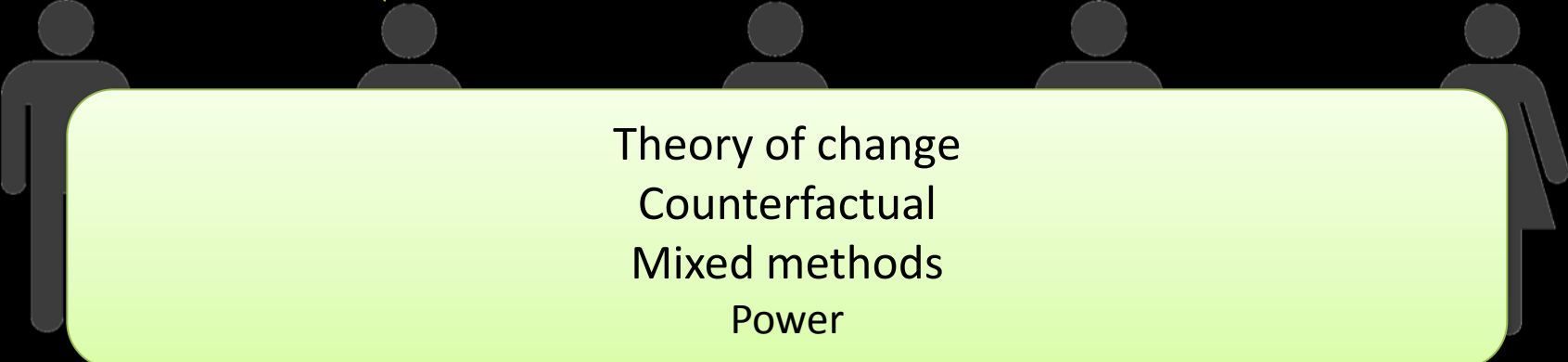
Efficacy: Does it work?

Would it have happened anyway?

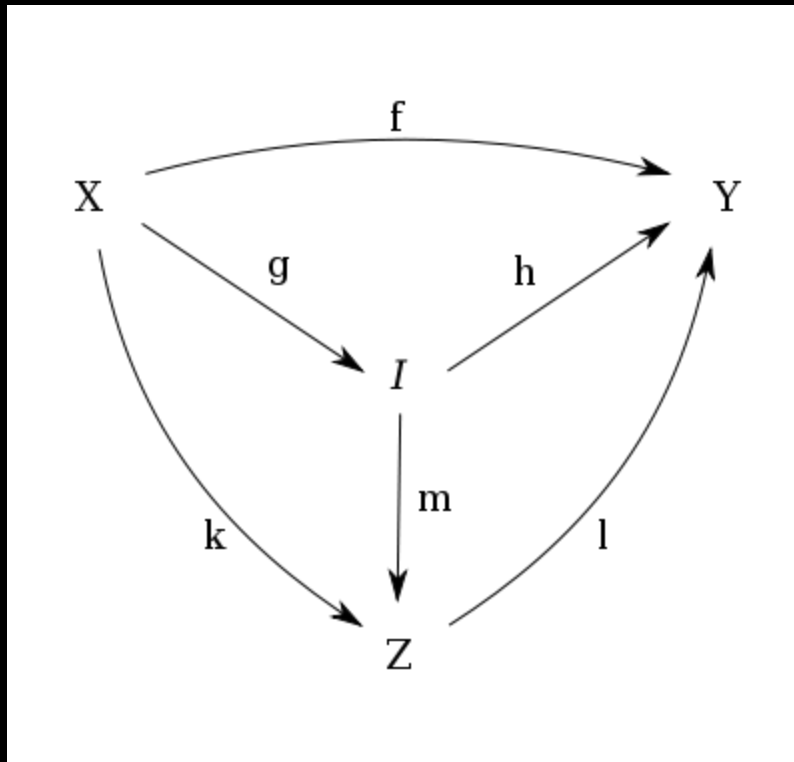
Are there other ways, that are *cheaper to get the same impact?*

Did the program *cause* the change?

If the program caused the effect, how *much* was the effect?

A row of five dark grey silhouettes of people sitting at a table, with a large light green rounded rectangle in front of them containing text.

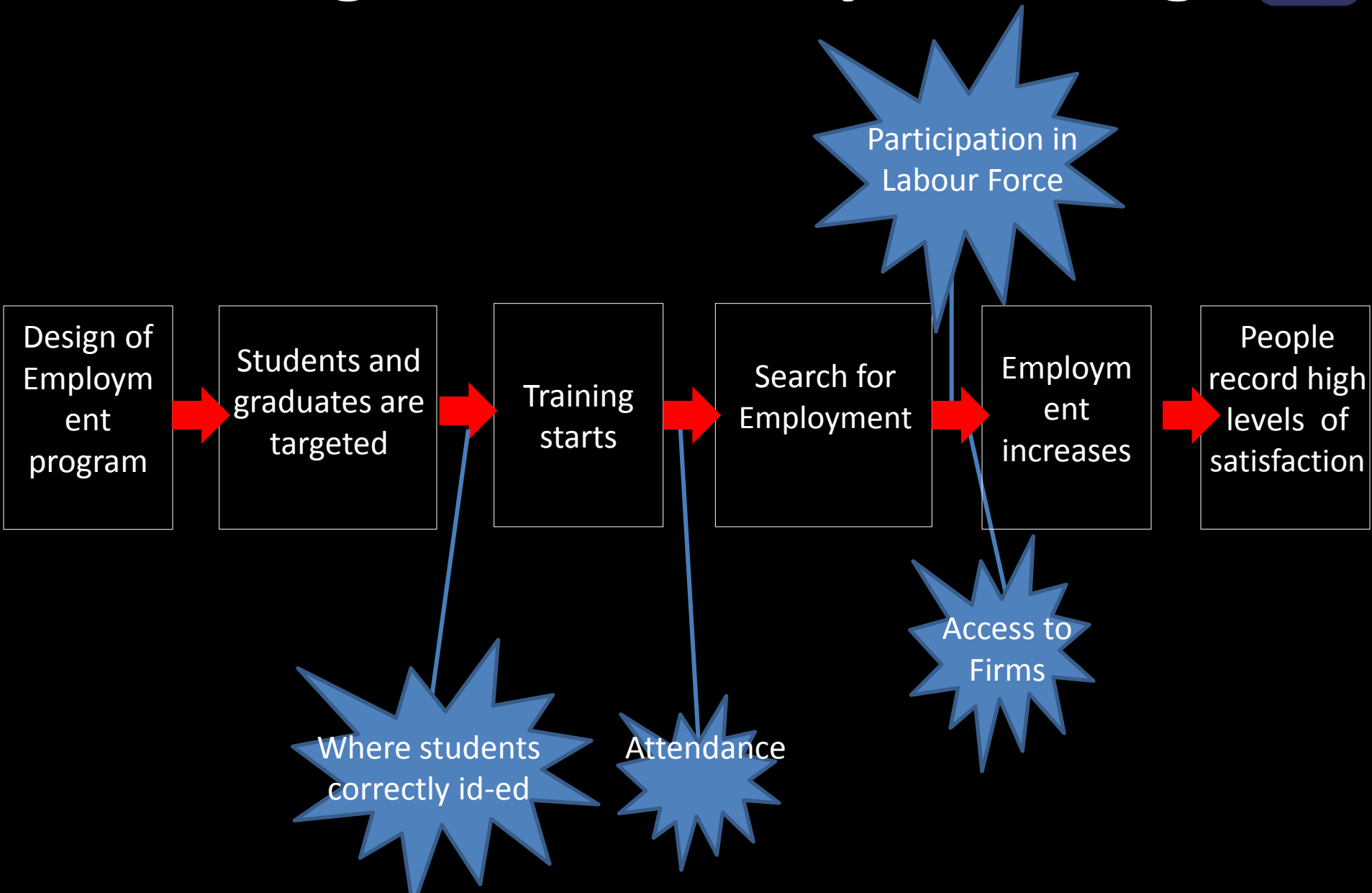
Theory of change
Counterfactual
Mixed methods
Power



**THE FIRST STEP TO EVALUATING
IMPACT: THE THEORY OF CHANGE...**

**A YOUTH EMPLOYMENT PROGRAM:
A THEORY OF CHANGE**

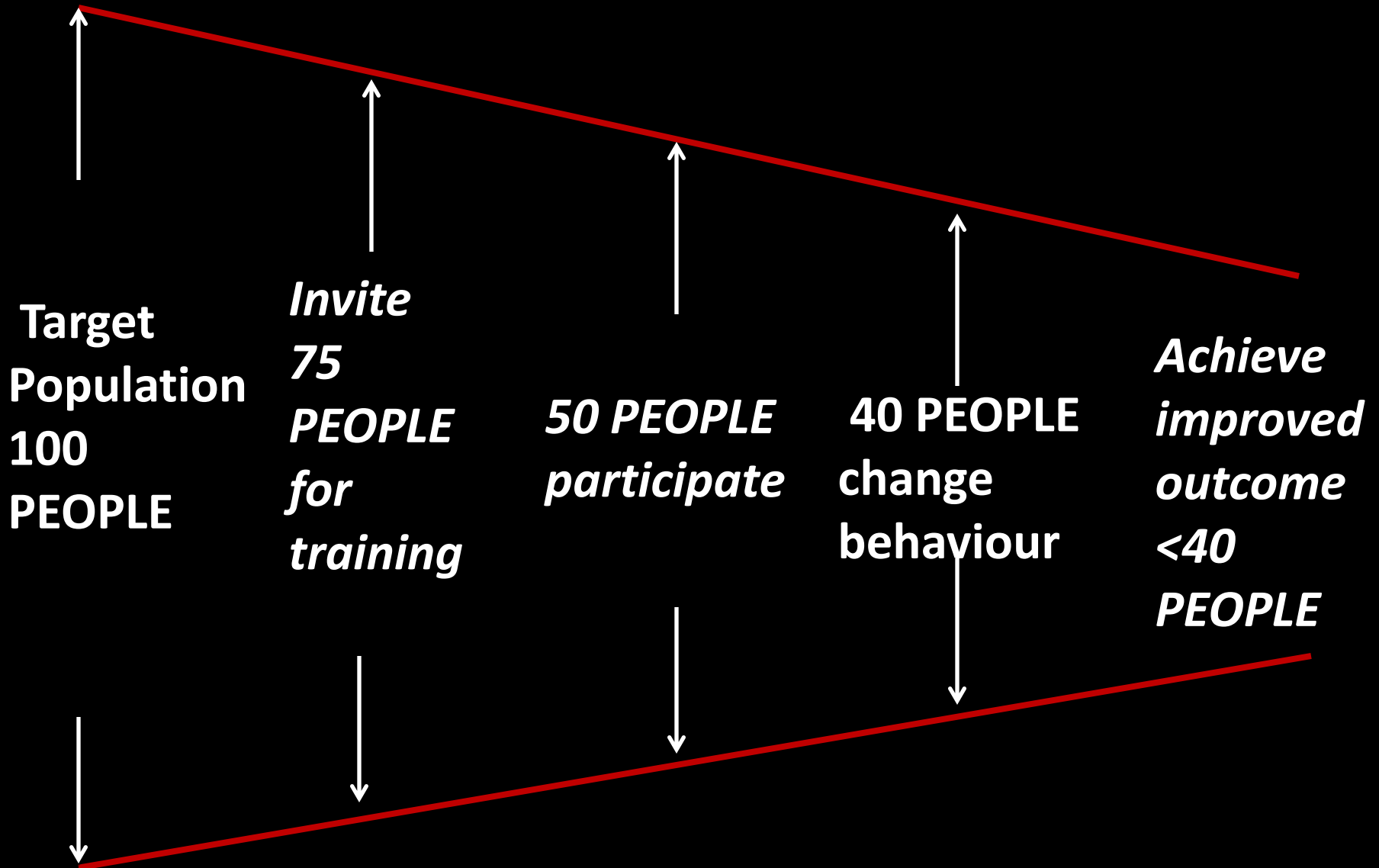
Starting with a Theory of Change



The theory of change

- Starting point: Outcome OR activity
- Specify linkage between the two
- What are other factors that might affect the linkage?
- What are the assumptions that you are implicitly making?

Funnel of Attrition



Monitoring and Evaluation

- Monitoring – Frequent, cheap (time and money), easy, transparent, systemic, helps identify implementation challenges;
- Evaluation – Mid-project (formative) or end of project (summative) or ex post (impact), expensive;



A Youth Employment program *caused* a 7% increase in Employment of the *included* population.

LETS PARSE THE STATEMENT!

THE HOLY GRAIL STATEMENT.

Causality? The attribution problem in projects.

What do we need to measure impact?



Employment of high school graduates six months after graduation.



| | Before | After |
|---------------------|--------|-------|
| Project (treatment) | | 92 |
| comparison | | |

The majority of “evaluations” have just this information ... which means we can say absolutely nothing about impact

Before versus after comparison

Before versus after = $92 - 40 = 52$

| | Before | After |
|---------------------|--------|-------|
| Project (treatment) | 40 | 92 |
| comparison | | |

Naïve statement: “the training has led to an increase in employment”

This ‘before versus after’ approach is outcome monitoring.

Outcome monitoring has its place, but it is not impact evaluation

Post-treatment comparison
Single difference = $92 - 84 = 8$

| | Before | After |
|---------------------|--------|-------|
| Project (treatment) | | 92 |
| comparison | | 84 |

But we don't know if they were similar before...
though there are ways of doing this (statistical
matching = quasi-experimental approaches)

$$\text{Double difference} = (92-40)-(84-26) = 52-58 = -6$$

| | Before | After |
|---------------------|--------|-------|
| Project (treatment) | 40 | 92 |
| comparison | 26 | 84 |

Conclusion: Longitudinal (panel) data, with a comparison group, allow for the strongest impact evaluation design (though still need matching).

SO WE NEED BASELINE DATA FROM PROJECT AND COMPARISON AREAS/INDIVIDUALS

Constructing the Counterfactual

Counterfactual is often constructed by selecting a group not affected by the program

Non-randomized:

Argue that a certain excluded group mimics the counterfactual.

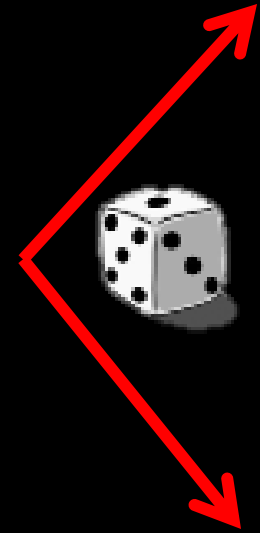
Randomized:

Use random assignment of the program to create a control group which mimics the counterfactual.

Step I
Eligible units

Step II
Evaluation sample

Step III
Random Assignment



Control



Treatment



External Validity



Internal Validity



Ineligible



Eligible

Validity

- A tool to assess credibility of a study
 - ***Internal validity***: can we attribute our impact estimates to the program and not to something else?
 - ***External validity***: can we generalize our impact estimates from this program to other populations, time periods, countries, etc?

In other words, there are *no systematic* differences between the two groups.

Critical things evaluation

- Theory of change
- Credible counterfactual
- Monitoring and process evaluations
- Large Sample
- Mixed methods
- Knowing the external validity of the methods.
- Accountability AND Learning



Thank you.

www.3ieimpact.org

molapade@3ieimpact.org