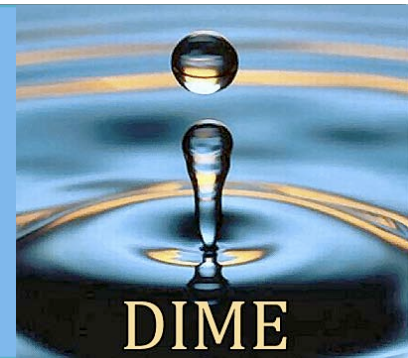


Arianna Legovini

Head, Development Impact Evaluation Initiative (DIME)
World Bank

Impact Evaluation for the Success of Policies

***“Job Training Programs for Youth”
Impact Evaluation Day
Latin America Region &
Human Development Network
May 11th and 12th, Washington, D.C.***



Do we know...

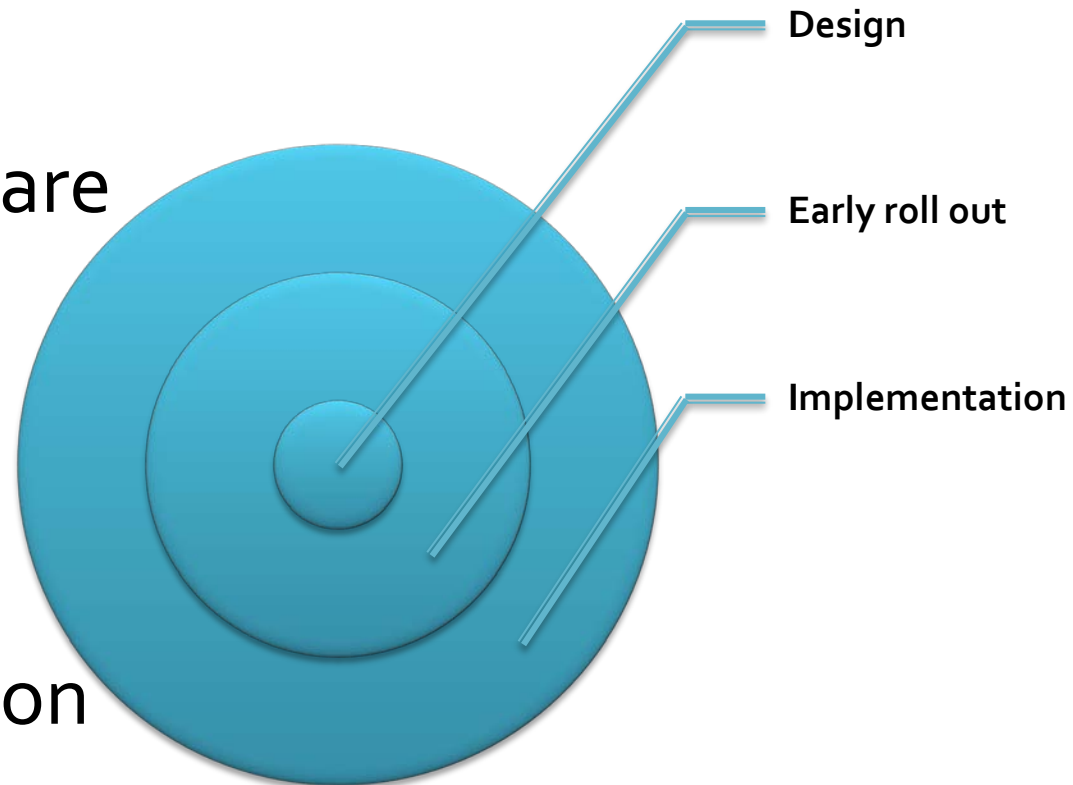


- How to get Pedro to transition from a dynamic of failure to one of success?
- How to make him think that he can do it and that is what he indeed wants? How to teach him leadership, responsibility and self-confidence?
- We know he needs some actual skills to convince an employer to take him, but which will be the ones that will land him a job tomorrow? And which will keep him employed for years to come?
- Are technical skills what he lacks or is the ability to present himself & project the image of competence and reliability? And how do we teach that?

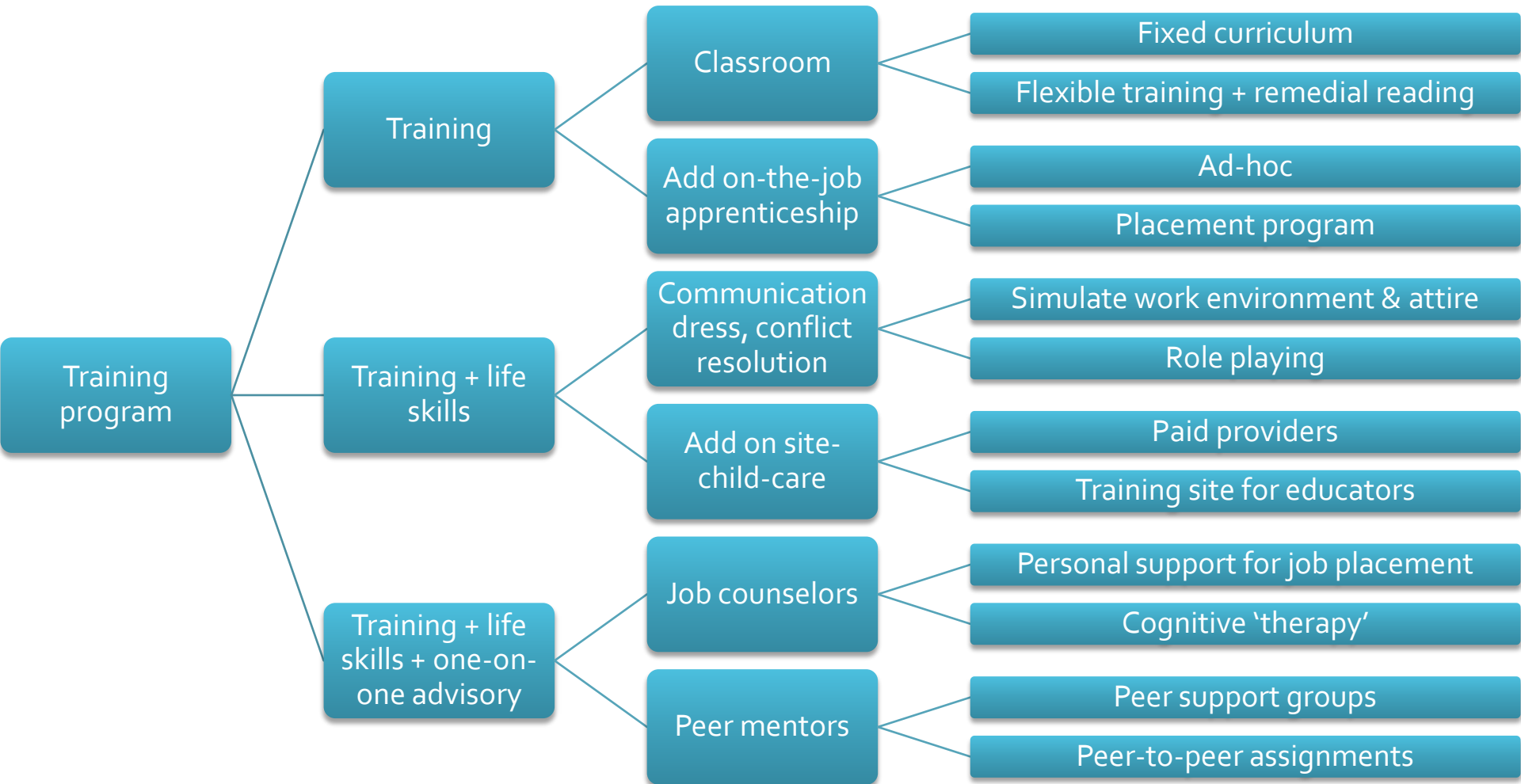
The decision process is complex



- A few big decisions are taken during design but many more decisions are taken during roll out & implementation



Developing a decision tree for a training program...



How to select between plausible alternatives?



- Establish which decisions might be critical and should be tested during roll-out
- Scientifically test critical nodes: measure the relative impact of alternatives
- Pick better
- Iterate
- *Cannot learn everything at once*
- *Select carefully what you want to test by involving all relevant partners*

Impact evaluation: Apply scientific method to understand human behavior



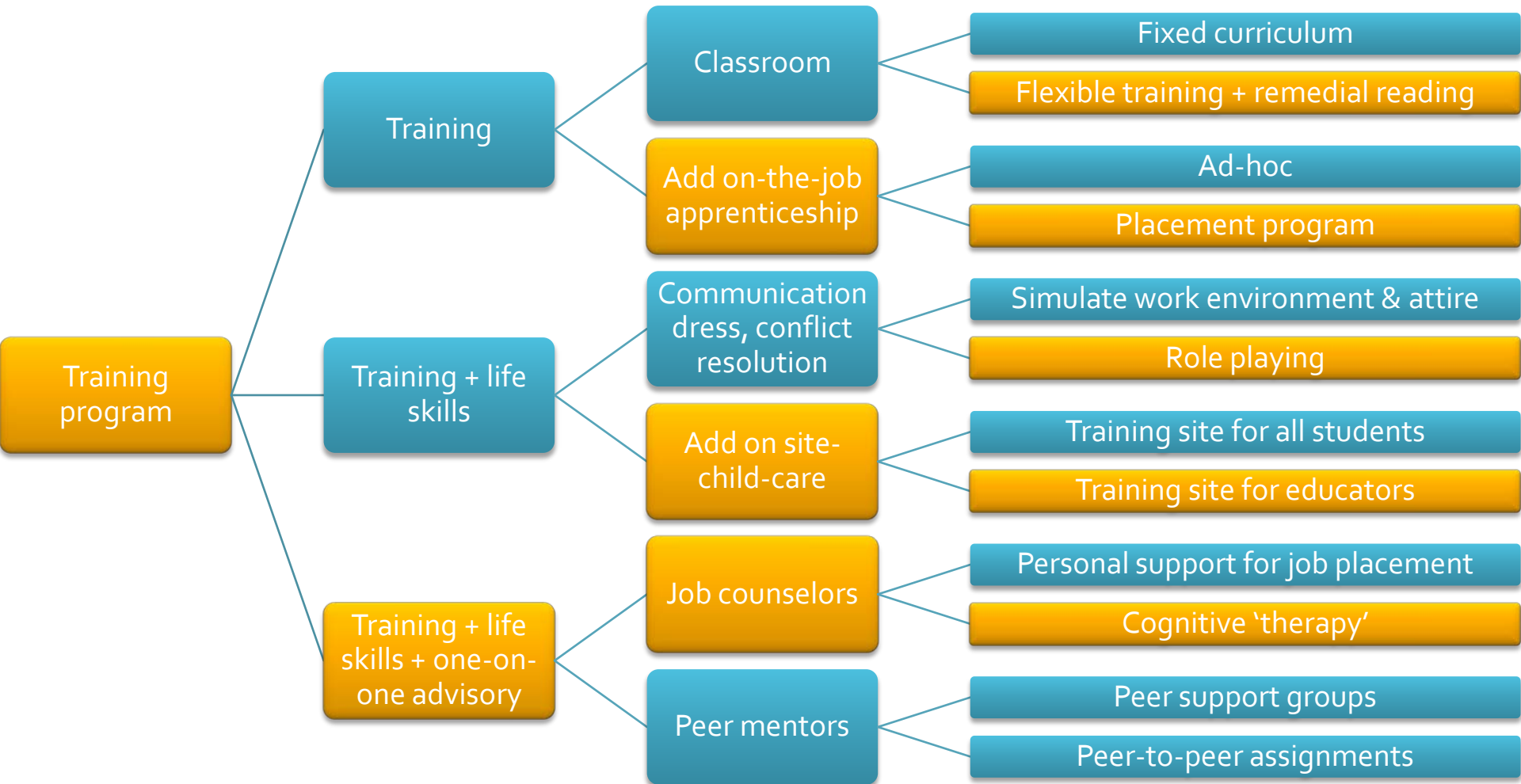
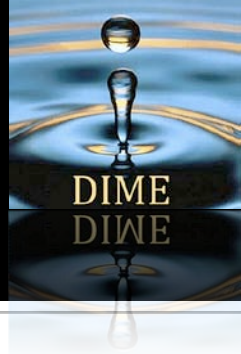
- Hypothesis
 - If we add an on-site child care center then the program will be able to attract and retain young women
 - If students work in the CC center then their class performance will improve
- Testing
 - Randomly assign on-site child care (CC) to training centers
 - Randomly assign students in CC training centers to devote few hours a week to CC
- Observations
 - Female enrolment rates in CC training centers increase—substitution?
 - Student absent days decline in CC training centers and completion improves
 - CC worker class performance unchanged
- Conclusion
 - CC improves attendance and completion; working in a CC does not enhance performance

Impact evaluation: Apply scientific method to understand human behavior



- Hypothesis
 - If the program includes remedial reading then students will feel better about themselves and more confident in their learning and labor potential
- Testing
 - Randomly assign low-literacy students to remedial reading
- Observations
 - Remedial reading students improve attendance, achievement, visits to counselors etc
- Conclusion
 - Remedial reading improves confidence and potential

Walk along the decision tree for your program to get more results



What and Why Evaluate?



- Impact evaluation can be used to measure program effectiveness
- Can also be used to improve quality of programs
 - Separate institutional performance from quality of intervention
 - Test alternatives and inform design in real time
 - Increase program effectiveness
- Plan for flexible programs that integrate a learning agenda
 - Plan for implementation of options not solutions
 - Find out what alternatives work best
 - Adopt better way of doing business and taking decisions

Conceive of impact evaluation as a package



- One umbrella evaluation to know how much the program delivers and several experiments within to know how to make the program delivers more



- Provide a series of useful (actionable) information at each step of the implementation of your program



- This sustains interest and makes sure the evaluation pays for itself



Shifting Program Paradigm



From:

- Program is a set of activities designed to deliver expected results

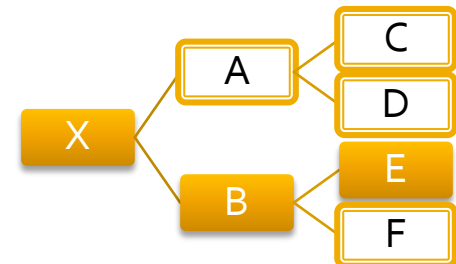
➤ Program will either deliver or not



To:

- Flexible program with a menu of alternatives and a learning strategy to find out which work best

➤ Change programs overtime to deliver more results



Shifting Evaluation Paradigm



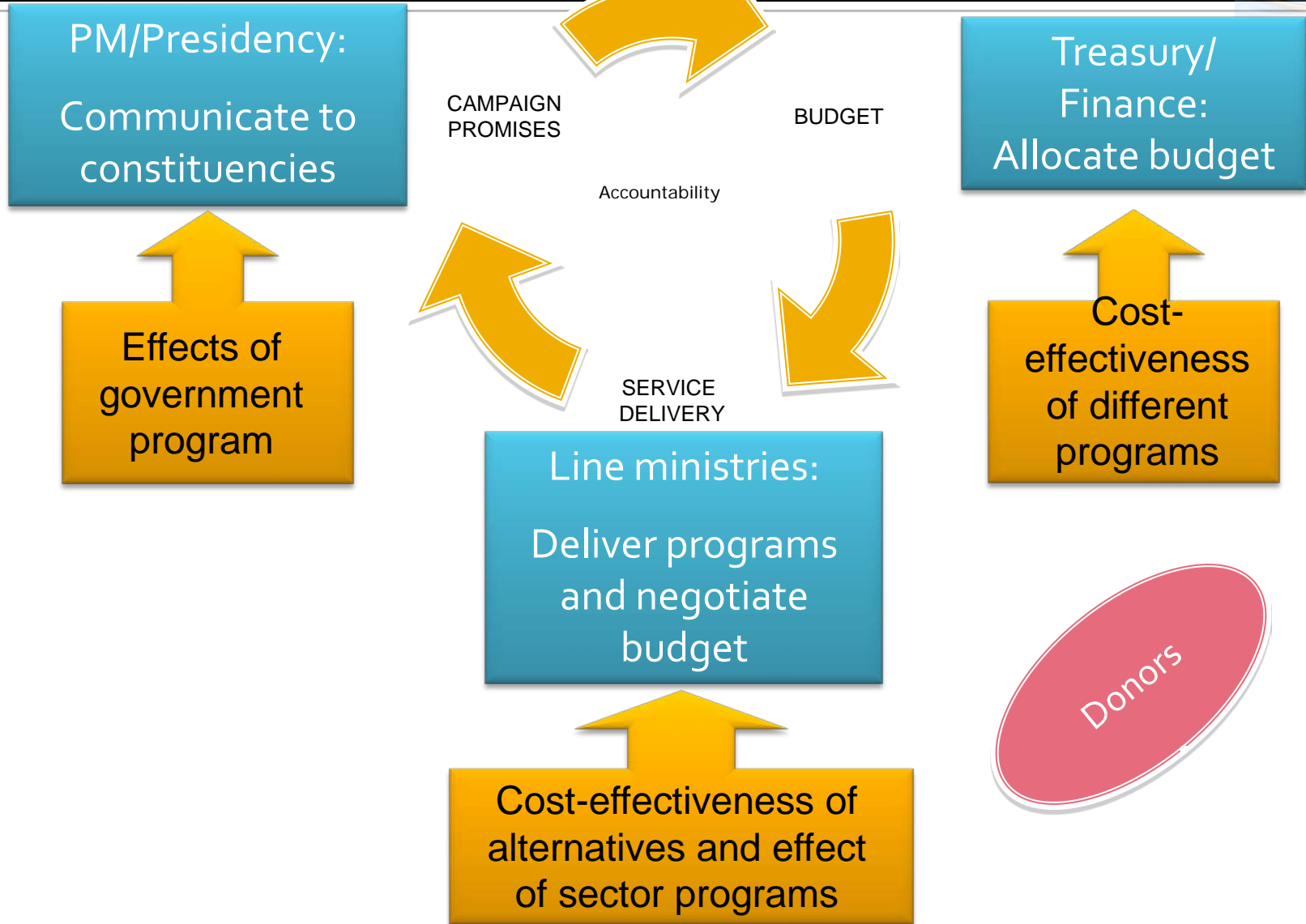
- From retrospective, external, independent evaluation
 - Top down
 - Determine whether program worked or not
- To prospective, internal, and operationally driven impact evaluation /externally validated
 - Just-in-time advice to improve effectiveness of program over time

Get everyone on board



- Make sure all relevant parties are at the table
- Identify key interventions and priorities for learning
- Create a team from the program that will be responsible for the evaluation
- Design basic evaluation “package”—this will evolve over time
- Discuss with senior management to validate and align their expectations

Who?

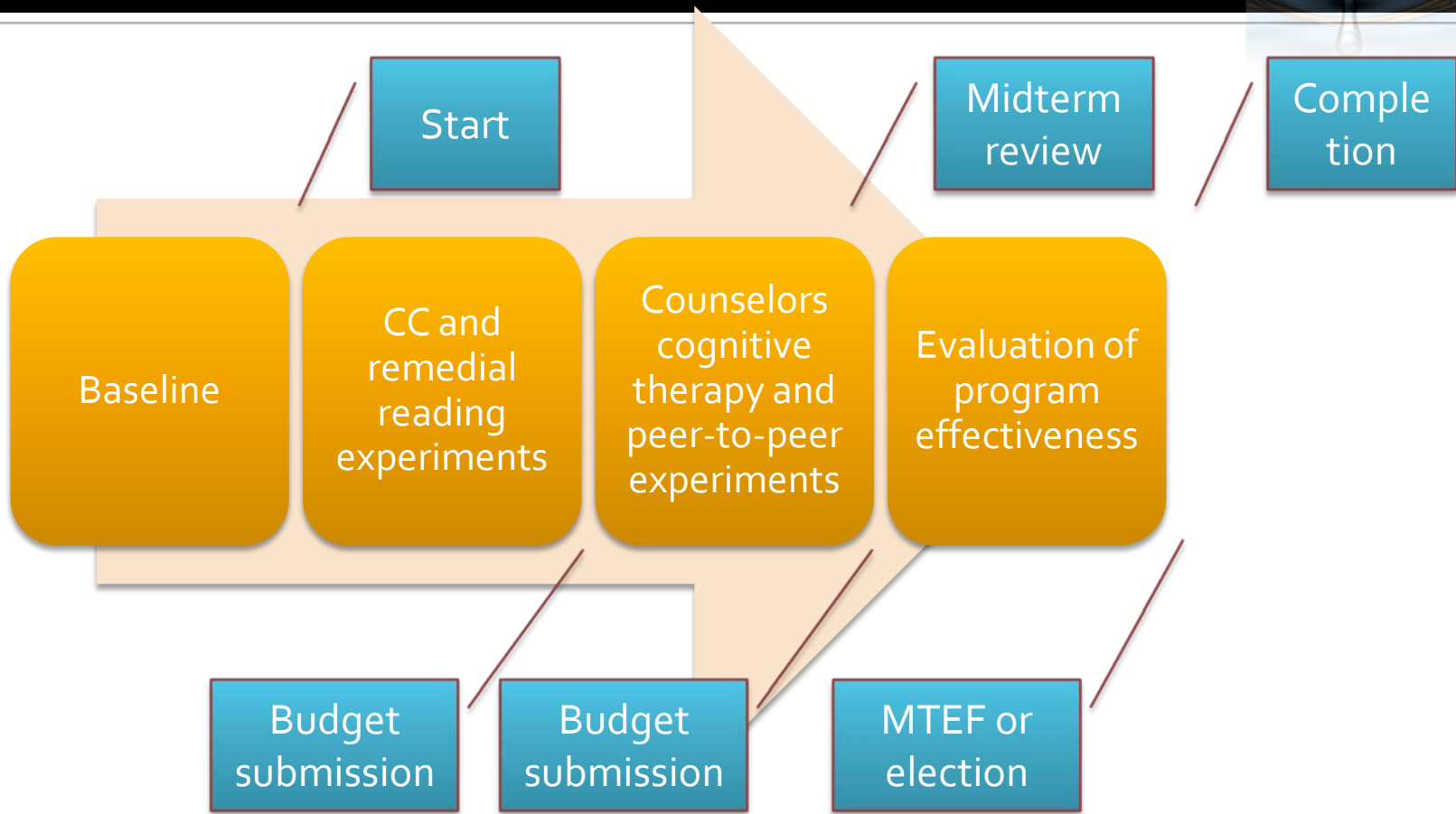


Time the results to policy cycles



- For Bank projects, time evaluation results to start date (baseline analysis), midterm (results of some initial experiments to inform changes) and completion (endline analysis of program effectiveness)
- For government programs, yearly budget discussions, MTEF, and elections
- For effectiveness, it depends! Develop the whole causal chain and time the evaluation at the different stages when something should happen (enrolment, attendance, achievement, job search after graduation, employment status, income, length on the job—these are not unique times)

Timeline example



Operational questions: managing for results



- Question design-choices of program
 - Institutional arrangements, Delivery mechanisms, Packages, Pricing/incentive schemes
- Use random trials to test alternatives
- Focus on short term outcomes
 - take up rates, enrolment, attendance, use of advisory, adoption of positive behaviors
- Follow up data collection and analysis
 - 3-6-12 months after exposure
- Measure impact of alternative treatments on short term outcomes and identify “best”
- Change program to adopt best alternative

Policy questions: accountability



- How much does the program deliver?
- Is it cost-effective?
- Use most rigorous method of evaluation possible
- Focus on higher level outcomes
 - Search time, employment, income, length on the job
- Measure impact of operation on stated objectives and a metric of common outcomes
 - One, two, three, five, ten year horizon
- Compare with results from other programs
- Inform budget process and allocations

Impact Evaluation Design



- Make sure the “evaluator” has the right skills and agrees to work with you throughout the evaluation
- Don't settle for second best
 - Discuss program until you find operational opportunities to develop very good designs.
 - Common operational opportunities are budget and capacity constraints.

Identification strategies



- Random assignment among applicants used in evaluation of job training programs—best—also random assignment of sites to different treatments
- Random encouragement designs also possible if unrestricted entry—less of an ethical debate and can be implemented when program is at scale
- Regression discontinuity when entry rules are quantitative (age cut off, means testing)
- Non-experimental unlikely to succeed: difficult to control for motivation, self-confidence, resourcefulness that determine success in the labor market

Data collection



- Survey instruments:
 - The ones used by other job training programs help you benchmark your own results
 - Include all causal chain variables (if it goes wrong, where does it go wrong?)
 - Include all other relevant variables for analysis (covariates)
- Baseline before roll out and even when doing randomized evaluations—
 - provides information on targeting, balance btw treatment and control
 - insures against the randomization going sour
- Cut costs: piggy back on other ongoing surveys

Have field coordinator on the ground with the right skills



- Train team
- Supervise data collection
- Supervise roll out to ensure design is not compromised—very important that treatment group receives treatment and control does not
- Record any occurrence that should be incorporated in the analysis
 - Program assigned to control
 - Treatment replaced by other

Roll out intervention



- Give clear instructions to all involved in the program and explain why rules are needed
 - People who understand why will stand by
- Intense monitoring to record what is being delivered, when and to whom
- Is anything else being delivered to the treatment or the control?
 - NGO decides to train the applicants that were not admitted to your training program. Now you can only evaluate your program versus theirs.

Follow up data collection and analysis



- Are the effects statistically significant?
 - Basic statistical test tells whether differences are due to the program or to noisy data
- Are they policy significant?
 - If the job training program is expensive and has positive effect but it's tiny, may not be worthwhile
 - Or perhaps, you measured too early. Job training programs may have little short term and bigger longer term effects. Wait?
- Are they sustainable?
 - Conversely, if youth learn little and forget fast, maybe the program needs a facelift

Policy feedback loop



- Incorporate results into program
 - If you know something works better, it is unethical to leave program unchanged!
 - Otherwise it is also a waste of resources paying for the evaluation
 - Thi will be hard to do if you have not involved every relevant person from the beginning
- Write up to align the politics
 - One liners for the Minister/President
 - Summary for budget/MTEF discussion
 - Article for the newspaper

Start discussions again



- Q&A