CCRC COMMUNITY COLLEGE RESEARCH CENTER
TEACHERS COLLEGE, COLUMBIA UNIVERSITY

MATHEMATICA Policy Research

General Principles for Creating Comparison Groups

Russell Cole
Mathematica Policy Research

TAACCCT Evaluation and Measurement Convening
August 7 and 8, Chicago, Illinois

What is internal validity?

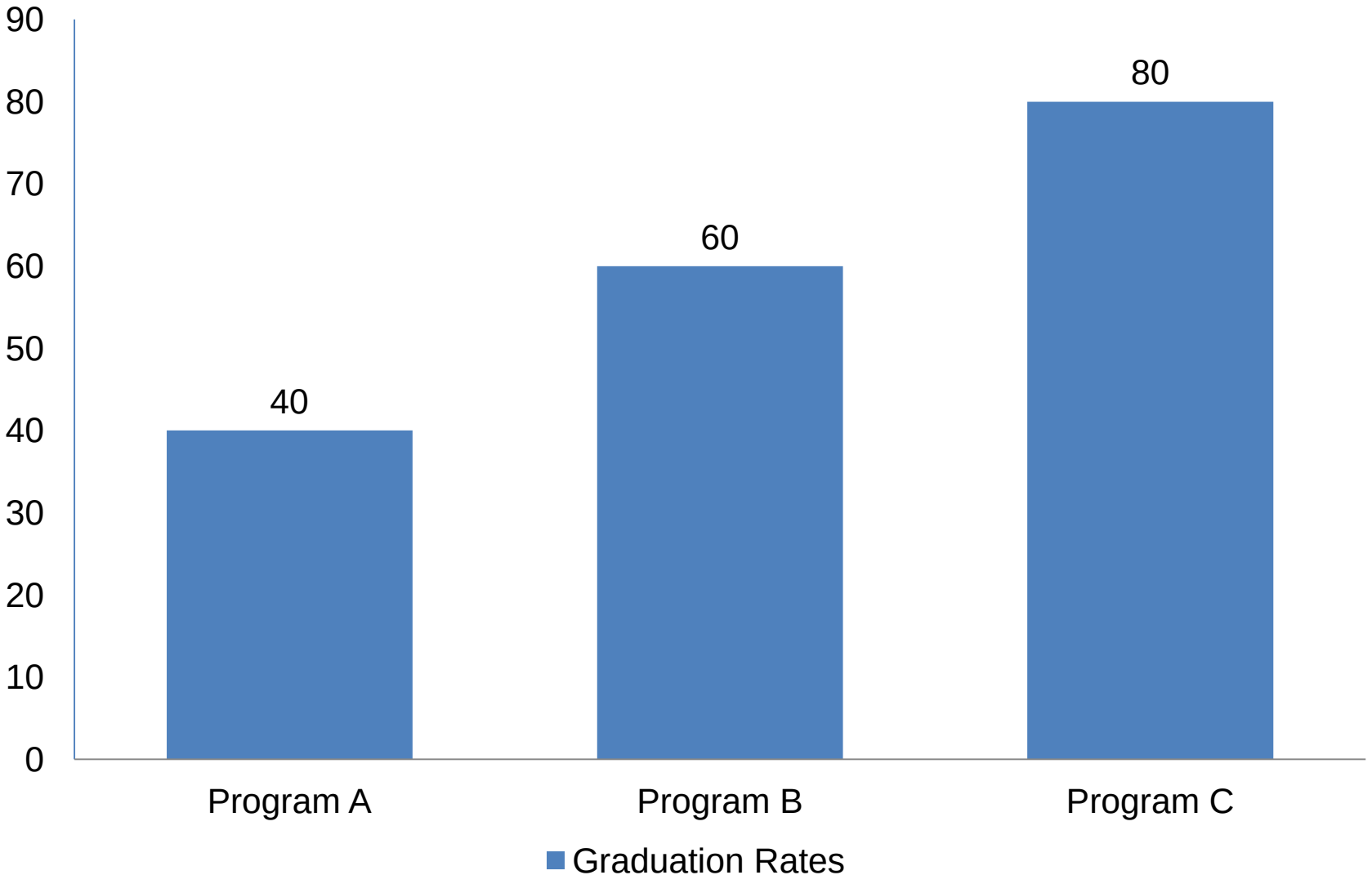
- How do we know that a program is successful?
  - Anecdotes?
  - Surveying program completers?
  - Interviewing employers?
- DOL/funders/administrators/policy-makers may want evidence of the success that is more "rigorous"
- A rigorous, well-implemented evaluation can provide credible/compelling evidence of a program's impact

➢ ***An evaluation that is internally valid provides credible/compelling evidence of a program's impact.***

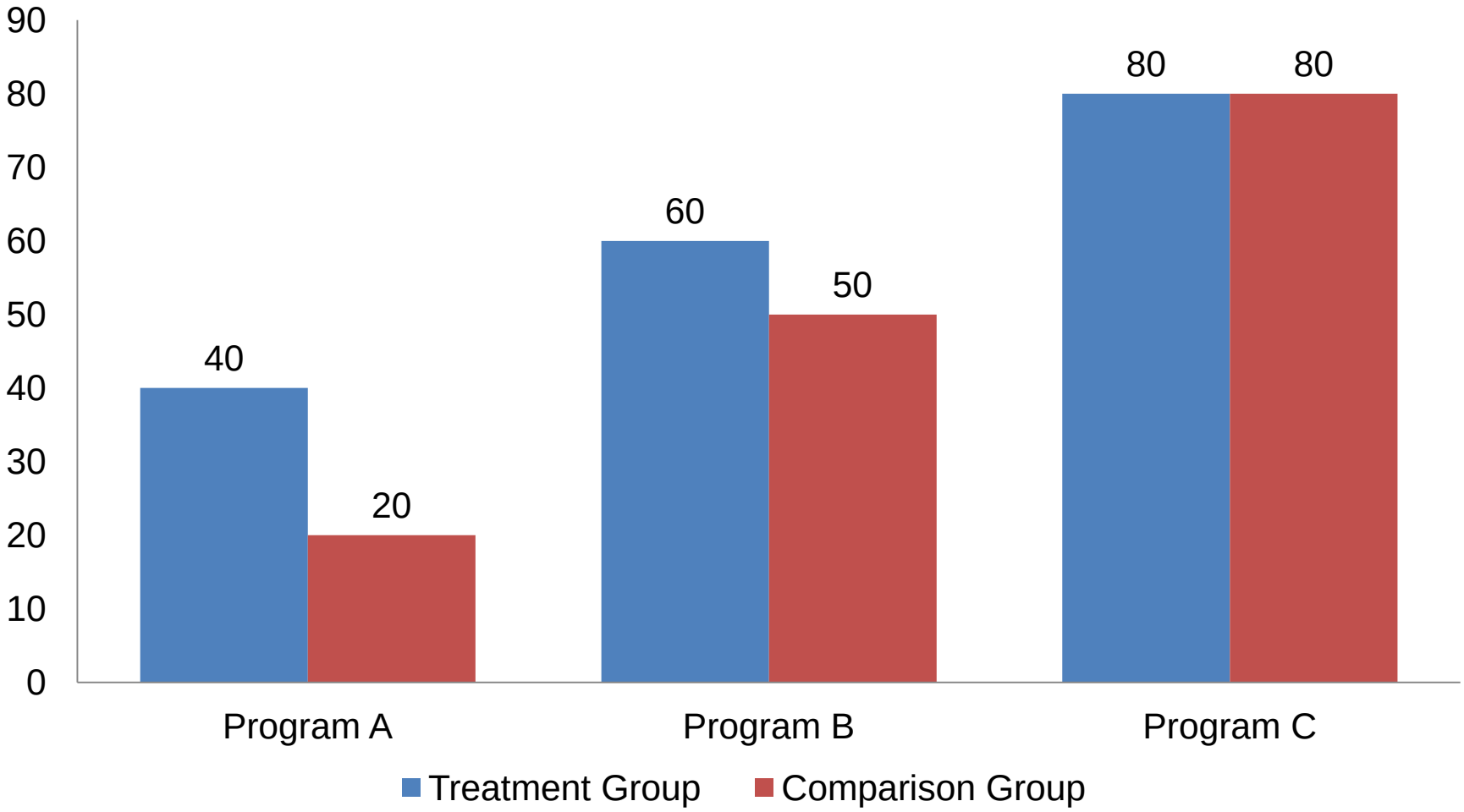
Goals

- Illustrate importance of comparison group
- Identify components of a rigorous evaluation that funders/administrators will find as compelling evidence of success
- Determine potential threats to validity of TAACCCT evaluations
- Establish solutions to improve the internal validity of each evaluation

Treatment group results can be deceiving



… What would have happened without the program (similar comparison group)



A comparison group is a step in the right direction…

- But having a comparison group is not a panacea…
- The credibility of our treatment vs. comparison "effect" depends on the extent to which the only difference between the two groups is participation in the program

➢ ***A comparison group is necessary but not sufficient to appropriately articulate the effects of a program on outcomes.***

A randomized controlled trial (RCT) serves as the "gold standard" for demonstrating the impact of a program

- Students randomly assigned to receive the program or not
- Only difference between the students is receipt of the program
- And thus, differences in outcomes (e.g., graduation rates) are solely due to differences in the program that the students receive

➢ ***RCTs allow for the best evidence of program effectiveness because they ensure that students are "similar" across conditions.***

What do we need to do to convince a skeptical critic?

- We need to do more to convince our audience that the evidence is compelling…
- Though the results may be seen as providing a lower tier of evidence

➢ ***It is possible to demonstrate credible program impacts without a RCT – but there will be caveats to the findings.***

Components of an internally valid comparison

- A rigorous impact evaluation should be able to mitigate the following threats to an internally valid comparison *(Campbell & Stanley, 1963)*:
  - Selection
  - History
  - Instrumentation

➢ ***There are three key threats to internal validity that we should keep in mind.***

Selection threat

- Students in treatment and comparison groups are very different from each other
  - e.g., treatment group has better employment history than comparison group
- At end of program, we observe differences in student earnings, and it's impossible to distinguish whether differences are due to
  - the true effect of the program, or
  - differences in the students at baseline that persist

Mitigating the selection threat

- Choose a good comparison group that has students that are similar to the treatment students
- Show that the students are similar to each other at baseline
  - Provide means/standard deviations for treatment and comparison groups on variables that are expected to be related to the outcome
  - Especially "pretest" types of measures of student academic outcomes and previous employment/earnings
- Statistically control for baseline differences in final impact analyses

History threat

- External events cause the observed changes in earnings
- Only a problem in studies where prior year cohort(s) is/are compared against a current cohort
- Example:
  - Treatment group = 2012 graduating cohort
  - Comparison group = 2011 graduating cohort
  - Context: Economy improves in 2012, and everyone in the treatment group gets a high-paying job
- The observed differences we see in earnings are due to
  - the true effect of the program, or
  - the external event of general economic improvement

Mitigating the history threat

- Use available data from another set of students (not those in the treatment or comparison groups)
  - For example, two cohorts of students in different programs from treatment and comparison groups

- Compare differences in outcomes over time for this additional set of students, relative to differences observed in the treatment and comparison groups

➢ ***This is really an exercise in convincing a critical reader that an earlier cohort is a valid comparison group.***

Instrumentation threat

- Differences in how the outcome of interest is measured across treatment and comparison groups confounds the observed difference
  - Treatment group = wage data obtained through DOL
  - Comparison group = wage data obtained through survey
- Observed differences in the outcome due to
  - The true effect of the program, or
  - The differences in the outcomes obtained across the two sources

Mitigating the instrumentation threat

- Don't use two different data sources (or different methods) for obtaining outcome measures

- If it's necessary to use two data sources, try to obtain data from both sources for some students
  - Show that the data are similar across both sources (e.g., correlation of outcomes across sources, magnitude of difference in outcomes across sources)

➢ ***Like the history threat, this is really an exercise in convincing a critical reader that instrumentation differences are not a problem.***

Best practices for comparison group studies

(Based on the WWC Standards)

- Show impacts on outcomes that are reliable
- Demonstrate the equivalence of the analytic sample at baseline (mitigate the selection threat and history threat, if applicable)
  - Statistically adjust for any baseline differences in impact analyses
- Do not have a systematic difference between the treatment and comparison groups
  - No systematic difference in data collection elements (mitigate instrumentation threat)
  - No "confounding factors" that align with the treatment being tested

➢ ***Follow WWC standards for compelling research evidence!***

Questions?

Russell Cole

rcole@mathematica-mpr.com