



SESRI Policy & Program Evaluation Workshop

Doha, Qatar
January 19-22, 2015

Outline: Session 5

- Measurement: Benchmarking in Evaluation Design
- Quasi-experimental research designs: Combining evaluation with policy implementation
- Outcome Evaluation
- Causal Inference vs. Statistical Inference
- Interpreting tables
- Effect size

Measuring academic progress

- Yusef memorized one vocabulary word yesterday and two vocabulary words today: Yusef is learning faster today than he was yesterday.
- Yusef memorized one vocabulary word yesterday and two vocabulary words today, but the average student in Yusef's school memorizes five vocabulary words a day: Yusef is learning more slowly than the average student.
- Yusef memorized one vocabulary word yesterday and two vocabulary words today, but the average student of Yusef's age forgets half the words he learns each month: what will Yusef's retention rate be?

Measurement: Benchmarking

- What's "academic progress"? The concept requires a BENCHMARK.
- Benchmark: standard that serves as a point of reference
 - In education, a selected, agreed-upon set of concepts, skills, and/or facts is designated as appropriate for each grade or age level
 - A measurement technique (tests, portfolios, observations) is then used to assess command of that knowledge
- What other kinds of evaluations might use "benchmarking"?

Choosing a Benchmark

- Although benchmarking can change with respect to a chosen goal, it should never be random.
 - Internal validity: what information are you trying to capture?
 - External validity: who do you want to accept your measure?
- Example: If algebra is taught in 7th grade in School A, but in 10th grade everywhere else, including it in a 7th grade benchmark for School A increases the benchmark's internal validity at the cost of its external validity.

Multi-site Assessment

- Evaluation across multiple sites requires a consistent dependent variable
- International education reform – Academic progress across years, across schools, compared to international peers
- Benchmarking can create that consistency, but implies assumptions
 - Year to year academic progress – same school: benchmark on a standard school curriculum, assuming curriculum does not change significantly from year to year
 - Academic progress across schools: benchmarking on grade level standards, assuming that differences in non-school characteristics can be controlled for/excluded via experimental design
 - Academic standards internationally: benchmarking across grade and/or age, generally accepting lack of meaningful controls

K-12 Education in Qatar as a Public Policy Problem

- In 2001, 20% of students failed the secondary school exit examination, despite only needing to score 50% or higher to pass (Ministry of Education)
- In 2001, only 47% of individuals taking the Qatar University (QU) entrance examination achieved high enough scores to be accepted into one of QU's colleges (Ministry of Education)
- How do you solve this problem?

Qatari K-12 Education: Policy Options

- Form small groups and discuss one of these two options. How likely would parents/teachers/students be willing to accept these changes? What might happen in the implementation?
 - **Problem definition:** The Qatari curriculum is not sufficiently rigorous or advanced. **Solution:** Design and implement a more rigorous standardized curriculum in half of the Ministry schools; randomize schools to implement this curriculum.
 - **Problem definition:** Students do not learn effectively when teachers lecture without customizing the curriculum to their needs. **Solution:** Train teachers to use a more student-centered, interactive form of teaching in half of the Ministry schools; randomize teachers to use this interactive form of teaching.

“Education for A New Era”: Program Design

- Public Problem
 - The Qatari government perceived that students matriculating from Ministry schools required greater academic proficiency to succeed in postsecondary education or the Qatari labor market.
- Program Model
 - Create **Independent Schools** that will operate autonomously to design and teach a curriculum aligned to national standards.
- Hypothesis
 - Independent Schools, autonomously designing and teaching curriculum that is aligned to national standards, will improve Qatari students' academic performance.

Design Tenets of “Education for a New Era”

Autonomy

Schools are given autonomy to implement a curriculum aligned to national standards



Variety

Several types of schools emerge with different educational philosophies and pedagogies to administer curriculum



Accountability

Schools are assessed on whether or not they meet national standards

Information is disseminated to parents and society



Choice

Parents choose schools based on academic performance

Clicker Question 1

What was the “treatment” that the Independent Schools reform, **as DESIGNED**, provided? (What is the *mechanism* through which they will improve academic performance?)

- a) Innovative curriculum benchmarked to national standards
- b) Student-centered pedagogy
- c) Parental choice of school best suited for child’s learning style
- d) Autonomous, school-level decision-making about hiring and curriculum
- e) All of the above

The Independent Schools Reform: The Implemented Design

- The Independent Schools reform intended to create a system of autonomous schools and Ministry schools, each offering different curricula, among which parents could choose. The combination of national standards and competition for students would raise school and elevate student achievement.
- Choice and competition did not persist in the final design of the reform, due to the small number of Independent Schools and the decision to convert them from existing schools with carryover student bodies.
- The conflation of curriculum, pedagogy, and school autonomy in Independent Schools would make it difficult to assign causal significance to any individual factor.

Exercise

In small groups, design an education reform that would test ONE of the following hypotheses, while holding all other factors in Qatari education constant. Try to create an RCT, a pre-post comparison with matched groups, and/or an interrupted time series design to test your hypothesis.

- a) An innovative curriculum benchmarked to international standards will improve academic achievement.
- b) Student-centered pedagogy will improve academic achievement.
- c) Decision-making at the school-level about teachers and administration will improve students' academic achievement.

The Independent Schools Reform: Evaluation Design

- All schools are tested with a new exam, benchmarked on the new curriculum standards.
- The evaluation design is therefore treatment and comparison group, multiple waves
- Because there are multiple waves of implementation, some of the treatment group (independent schools) are in the sample for multiple years.
- However, the schools that do not convert to treatment schools until later waves are not in the early waves of the sample. In other words, there are no pre treatment measurements.

Evaluating the Reform: Outcome Evaluation (I)

Do students in Independent Schools show greater academic achievement than students in Ministry Schools?

- What measures are needed to answer this question?
- Is this the question that the Ministry of Education wants answered? What policy conclusions is it likely to form from the answer?
- Would these conclusions be accurate? Why or why not?

Evaluating the Reform: Outcome Evaluation (II)

Do student-centered curriculum and pedagogy improve student achievement?

- What measures are needed to answer this question?
- What are the potential proxies for “student-centered curriculum” and “student-centered pedagogy”? Are these good proxies?
- What policy conclusions are likely to be formed from the answer to this question?
- Would these conclusions be accurate? Why or why not?

Evaluating the Reform: Outcome Evaluation (III)

Are students in Qatar more prepared for college and the labor market since the Independent Schools reform?

- What measures are needed to answer this question?
- Is this a question about program outputs, or program outcomes?
- What explanations might be available to explain the results, positive or negative?
- Can any of these explanations be tested by program evaluation?

Interpreting the Outcome Evaluation

Table 7.1
Students Meeting and Approaching Standards, by School Type, 2006

Subject	Percentage of Students at Performance Level					
	Meets Standards			Approaches Standards		
	Independent	Ministry	Private Arabic	Independent	Ministry	Private Arabic
Arabic	6	3	3	30	20	19
English	7	1	1	19	5	8
Mathematics	0	0	0	48	19	19
Science	0	0	0	30	16	15

SOURCE: Gonzalez et al., 2009.

NOTE: Percentages shown are rounded to the nearest whole percent and represent the average of the grade-level averages for grades 4 to 11 for the given school type.

PISA Results, 2012

» Compare your country - PISA 2012

+

-

Qatar

x

Overview

Student engagement

+share

Main results

Overall performance

% of top performers

% of low performers

Boys vs. girls

Mathematics

Reading

Science



Social background

Impact on performance

Performance gap

Resilience

Immigrant students

Results

Trends



● = better than OECD avg.

● = around OECD avg.

● = not as good as OECD avg.

↗ = getting better

→ = neutral trend

↘ = doing less well

Outcome Evaluation: A Closer Look at the Results

- IS students: better-educated parents
 - Father/Mother no HSD: Gen 1 IS: 21/18% MS: 44/48%
 - Father/Mother BA/BA+: Gen 1 IS: 49/54% MS: 26/22%
- IS schools
 - Few secondary schools, and secondary schools converted from scientific complex schools
 - Teacher characteristics could not be linked to student performance
- Conclusion: **Selection bias** operating for both students and schools

Outcome Evaluation: Using Statistical Inference

Table 7.2
Results of 2006 Test Score Model: Effect Sizes for Grades 4, 5, and 6

	Arabic	English	Math	Science
School type (vs. Ministry)				
Generation I				
Assessments in Arabic	+0.15		+0.25	+0.25
Assessments in English		+0.33	n.s.	-0.58
Generation II				
Assessments in Arabic	+0.10		n.s.	+0.40
Assessments in English		+0.14	-0.15	-1.00
Demographic characteristics				
Qatari	-0.13	-0.04	-0.12	-0.16
Non-Arabic speaker	-0.14	+0.07	-0.10	-0.18
Male	-0.18	-0.13	-0.15	-0.17
Special needs	-0.19	-0.09	-0.10	-0.23
Father not secondary graduate	-0.15	-0.10	-0.12	-0.18
Mother not secondary graduate	-0.14	-0.12	-0.12	-0.12

NOTE: Interval regressions were used to control for prior test score and background variables. Reported results are significant at .05 level; "n.s." indicates that results were not significant.

An effect size of 0.10 under "Arabic" is $1/10^{\text{th}}$ of a standard deviation of the Arabic score on the QCEA.

Is that a big or small effect?

The Lessons of Outcome Evaluation

- Students in Independent Schools show a significant advantage in academic achievement when compared to students in Ministry Schools. Less than half of all students in either type of school “approach standards”.
- The size of that difference ranges from equal to double the size of the effect caused by demographic characteristics.
- The outcome results are almost certainly influenced by selection effects at both the school and student levels, for which the design does not control.

Re-Designing the Outcome Evaluation

- Design: Post-test, treatment and comparison groups
- Potential changes in evaluation design
 - Linking teacher pedagogy, teacher characteristics to students
 - Adding curriculum variables to distinguish curriculum from school autonomy
 - Adding a pre-test to distinguish the treatment effect from selection effects
- Potential changes in program design: Randomization
 - Randomly assign schools to become independent schools.
 - Randomly assign students to attend independent or ministry schools.
 - What are the differences in these two designs?



SESRI Policy & Program Evaluation Workshop

Doha, Qatar
January 19-22, 2015

Outline: Session 6

- Understanding Stakeholders
- Process Evaluation
- Linking Process and Outcome Evaluation

Stakeholders

- Who is affected by the policy?
 - Who might benefit?
 - Who might be hurt?
 - Whose routines might change?
- Who will pay for the policy?
 - Directly? Indirectly? Are the costs visible?
- Who wants the program? Why?
- Who does not want the program? Why not?
- Are there competing interests at play?

Stakeholders in Education Reform

- Who are the stakeholders in the Independent Schools reform?
 - Ministry of Education:
 - How does it benefit?
 - Are its interests, jurisdiction, or reputation harmed by the reform?
 - Do its current practices need to change as a result of the reform?
 - Will it incur additional costs because of the reform? Is there additional funding to meet these costs?
 - Does it want this reform? Did it initiate it? Is there any evidence that it supported or opposed it?
- What other stakeholders might there be for the Independent Schools reform?
- Turn to your neighbor. For one of the stakeholders we have named, consider the questions we asked about the Ministry of Education.

The Role of Stakeholders in Program Success (1)

Stakeholders' roles in a program are critical to program operation.

- Stakeholders determine program goals

At the beginning of the Independent Schools reform, the Supreme Education Council wanted to raise Qatari educational standards. Within the first few years, it decided that this meant *increasing the representation of Qatari teachers as well as raising the achievement of Qatari students*.

- Stakeholders may be the program targets

All of the professional development resources went to the Independent Schools teachers, who were the obvious targets of reform. **But the Ministry Schools**, whose students also had to take the new assessments, **were also targets** of the reform. Their teachers received no help in aligning the curriculum with the new standards. As a result, Ministry teachers reacted inconsistently to the new standards: some changed their teaching, while others did not.

The Role of Stakeholders in Program Success (2)

- Stakeholders influence whether program inputs lead to program outputs

The Independent Schools reform presumed that student enrollment would be motivated by matching student interests to the type of Independent School design and curriculum. Most parents, however, chose the school their children went to based on the school's location and convenience.

- Stakeholders judge whether program outcomes are satisfactory

Independent Schools were intended to break away from dependence upon the required textbooks in the Ministry Schools and design their own curriculum. But parents objected to the unfamiliar lesson plans and the lack of familiar texts. In response, the Ministry of Education created a resource list from which Independent Schools were required to select textbooks, which then had to be supplemented with additional material.

Program Evaluation: The Simple Definition

- How did the stakeholders react to the program?
 - Process evaluation
- Did the program targets act as planned and expected?
 - Process and outcome evaluation
- Were the program goals accomplished?
 - Outcome evaluation

Evaluation as Decision-Making

- Evaluation requires making informed choices in a defensible and systematic manner.
 - Which goals are you assessing?
 - What kinds of target behavior are you looking for? If it does not exist, who do you hold accountable?
 - Which stakeholders' points of view are you taking into account?
- Each choice you make as an evaluator will be embedded in the program model(s) you test, the measures you collect, and the questions they allow you to answer.

Process Evaluation in the Independent Schools Reform

- How was the reform implemented at the Ministry level?
- How was the reform implemented in Independent Schools? **Across different Independent Schools?**
- How was the reform implemented in Ministry Schools?
- How was the reform implemented at the teacher level?
- How was the reform implemented at the parent and student level?

Ministry Implementation

- Qatarization Policy

- Original policy: Independent School operators had autonomy to hire teachers .
- Revised policy: Both Ministry and Independent Schools were required to attempt to meet minimum hiring thresholds for Qatari teachers
- Complication: Qatari teachers preferred to teach in Ministry Schools, because seniority counted for promotion and the hours were family-friendly. Conversely, Independent Schools were hampered in hiring expatriate teachers with experience in the new curricula they were trying to introduce.

- Under what conditions is this policy

- Reflective of stakeholder interests?
- Inconsistent with the goals of the reform?
- A threat to the validity of the evaluation's results?
- Likely to produce unintended consequences?

Ministry Implementation

- Curriculum Policy
 - Original policy: Schools required to design original curriculum and select materials independently; expectation of divergence from Ministry curriculum
 - Revised policy: Schools required to choose texts from an approved Ministry list; texts must cover at least 70% of required material in curriculum but be supplemented by additional resources
- Under what conditions is this policy
 - Reflective of stakeholder interests?
 - Inconsistent with the goals of the reform?
 - A threat to the validity of the evaluation's results?
 - Likely to produce unintended consequences?

Independent vs. Ministry Schools

- H: The improved curriculum and teacher engagement in Independent Schools relative to Ministry Schools will result in higher student achievement.
- Independent School teachers
 - Long days, heavy workload, availability of professional development, resources
 - Awareness of new curricular standards, focus on need to meet them
- Ministry School teachers
 - New standards promulgated but curriculum, resources, professional development did not change
 - Advantages of professional security, promotion to administrative jobs, predictability accrued primarily to Qatari teachers
- To what extent are these findings consistent with the hypotheses? To what extent do they introduce alternative explanations?

Evaluating “Student-Centered Pedagogy”

- Ministry Schools: Lecture-based teaching, little interaction between teachers and students, no class discussion or interactive learning
- Independent Schools: Encouraged to adopt “student-centered pedagogy”: interactive classrooms, geared to student interests and student pace
- How would you measure “student-centered pedagogy”?

Measuring “Student-Centered Pedagogy”

- Curriculum Planning
 - Recorded lessons
 - Teacher interpretations/feedback
- Classroom observations
 - Checklist
 - Open-ended
- Teacher reports
 - Surveys
- Interviews
- Student reports
 - Surveys
 - Focus groups
- Parent reports
 - Surveys
 - School contacts

Explaining “Student-Centered Pedagogy”

- Both classroom/teacher observations and teacher surveys show much more student-centered pedagogy in Independent Schools. Why?
 - H1: Professional development provided in Independent Schools.
 - H2: Non-lecture-based curriculum used in Independent Schools.
 - H3: Teachers in Independent Schools feel a sense of ownership for the curriculum because they helped develop it, and convey that ownership to students.
 - H4: Independent Schools attract the most motivated teachers, who also prefer student-centered pedagogy.
- What information do we need to distinguish between these explanations?

Parents and Students as Stakeholders

- Parents and students report more involvement, more pride, and more satisfaction in the Independent Schools.
- What evidence would we look for in the process evaluation to help us conclude that that these are consistent with a “school effect”, rather than just a “novelty” effect?
 - Teacher and administrator activity
 - School characteristics
 - Sense of mission (“initiation” effect)

The Lessons of Process Evaluation

- Independent Schools flourished despite changes in policy that altered significant aspects of their design.
- Independent Schools benefited from significant staff motivation and dedication, which could have many causes not solely attributable to the autonomous school model.
- The achievement gains attributed to Independent Schools could be linked to several mechanisms. These mechanisms are conflated in the Rand evaluation because differences between Independent Schools, and differences between teachers/classrooms within schools, are not considered.