What interventions for more women in science? Tools, knowledge and know-how for the creation and evaluation of public policies.

Guillaume Hollard\textsuperscript{1} Pedro Vergara-Merino\textsuperscript{2}

\textsuperscript{1}CREST-CNRS \& Ecole Polytechnique
\textsuperscript{2}CREST - GENES

February, 2023
A widely shared **consensus** on the existence of a leaky pipeline in western countries

Large attrition of the fraction of women along the hierarchical ladder

Here is an instance for High Schools in France
Women in Science

A sharp contrast between:

- A large consensus on where we stand (i.e. the leaky pipeline)
  - Found in almost all Western countries
  - Very well documented

- A myriad of propositions to change things
  - A multitude of policies (role models, affirmative action, targeted fundings, etc . . .)
  - Virtually everyone has a favorite proposition that “would work”!

Why such a contrast?
Outline of the talk

I address two main questions:
- Is there anything like a science of women in science?
- Why is it a difficult question?

I will:
- Provide examples of interventions (mostly in High Schools)
- Introduce the main tools of quantitative policy evaluation
- Point at promising interventions (but the talk is not a review)
Identifying (and defining) “what works” should be considered as a scientific question and addressed as such

- Should include tools to evaluate actions

- Should pay greater attention to things such as:
  - Heterogeneity of treatment effects
  - Scaling (what can be generalized beside a specific context)
  - Cost-benefit analysis

As we will see, understanding gender differences is difficult
Evidence from a large scale reform

French’s 2019 High School reform:

- From three pre-designed tracks (Scientific, Litterary, Economics and Social Sciences) . . .
- . . . To a choice of three (and later two) “specialization courses”
- A student can still choose “classic” combinations but they have more flexibility

What is the effect on the gender in Math?
How much should we trust our intuition?

What is the effect of the reform on female students taking Math courses?

Figure 1: Share of French female students taking Math courses
Source: Notes DEPP 21.41 and 22.19
Don’t trust your intuition too much...

What is the effect of the reform on female students taking Math courses?

Figure 2: Share of French female students taking Math courses
Source: Notes DEPP 21.41 and 22.19

The reform created a massive, and unexpected!, gender gap (heterogeneous treatment effect)
Role models have been often suggested as a way to increase the number of women in STEM. The main idea is as simple as nice. By meeting a person who succeeded, starting at the same place, participants may revise their expectations on their ability to succeed.
Scaling: An example

- "For Girls in Science" sponsored by Fondation L’Oréal uses successful female scientists as role modle in high school.
- Breda et al. (2021) found positive (but small) effects of role models on female students choosing scientific careers.
Scaling: An example

- “For Girls in Science” sponsored by Fondation L’Oréal uses successful female scientists as role models in high school.
- Breda et al. (2021) found positive (but small) effects of role models on female students choosing scientific careers.

However, when looking carefully:

- Several logistic and budget constraints to implement this policy nationwide
- Effect mainly driven by two role models (out of 56)!
- Difficult to imitate these effects with other role models

\[\Rightarrow\] Local effects that are too difficult to reproduce!
Programme Talens by ENS (Ly, Maurin, and Riegert 2015). In 2011-2012:

- 14 tutoring sessions for 260 underprivileged students
- A yearly cost of 240,000€
Heterogeneity and spillovers: An example

**Programme Talens by ENS** (Ly, Maurin, and Riegert 2015). In 2011-2012:

- 14 tutoring sessions for 260 underprivileged students
- A yearly cost of 240,000€

No average effect of the policy! But why?
Heterogeneity and spillovers: An example

Programme Talens by ENS (Ly, Maurin, and Riegert 2015). In 2011-2012:
- 14 tutoring sessions for 260 underprivileged students
- A yearly cost of 240,000€

No average effect of the policy! But why?
- Positive effects for the best students and negative effects for the rest (heterogeneous treatment effects)
- Students spend more time in public transportation and studying for the tutoring sessions and less time for other subjects! (spillovers and crowding out)
In sum: An empirical question!

- Because it is difficult to anticipate what works, policies should be tested in the field
- The science of women in science is an experimental one!
- From "what works?" to "what works best?" (i.e. how to best allocate resources to reduce the gender gap?)

A perfect example of a policy evaluation
Addressing empirical questions in social sciences

The key difficulty is that we wish to establish causality (not just correlation)

- Great developments of empirical tools to evaluate causality in the last two decades
- Golden standard is RCT
  - Randomly select a control and a treatment group
  - Measure variable of interest prior and post intervention
  - Compare the effect of treatment to that of control
- In short, causality needs randomness!
What works?

Ideally, we are looking for interventions that would:

- Have a large positive impact
- Few negatively affected persons
- Are easy to generalize
- Are not too costly (compared to expected benefits)
- Has been evaluated using a RCT

Too demanding?
Women and (under-)confidence

Girls are less likely to say they are good at Maths
for any given level of Maths skills

Fraction of women (red) and men (green) who say they are good at Math across ability

Data Source: TIMMS 2019. Based on 250,926 8th graders across 46 countries.
Notes: The x-variable is the global percentile rank of the 1st plausible value in mathematics. The y-variable is the share answering ‘Agree a lot’ or ‘Agree a little’. The chart is unweighted.
Women and (under-)confidence

Women are often found to be less self-confident than men.

Here is a comparison between expected grades and actual grades for the French National Exam (Bac).

[Graph showing the comparison between expected and actual grades for different levels of honors, with a trend line for men and women.]
What works? Reducing underconfidence

- It is possible to providing feedback to students on their relative rank in the national test score distribution in France.
- A random sample of students received objective information regarding their ability.
What works? Reducing underconfidence

- It is possible to providing feedback to students on their relative rank in the national test score distribution in France.
- A random sample of students received objective information regarding their ability.
- This intervention closes about 60% of the gender gap in the likelihood of applying to elite programs (CPGE)!
- It is simple, cheap, and easily scalable.

Source: Hakimov, Schmacker, and Terrier (2023)
Targeted information consists in providing objective information to women (their position in the overall distribution, their likelihood to succeed, etc).

In particular, providing information at the individual level seem more efficient.

The largest effect are found when students are informed about opportunities they didn’t know about.

Note that negative effects are unlikely.
Learning from past experience is difficult

- Many interventions are not evaluated at all!
- Relevant indicators are not collected
- Heterogeneous treatment effects is the rule
- Evaluation can be biased (selection bias, false positive, etc)
- Intuition based on a single example may be misleading
Conclusion

However, there are reasons to be optimistic!

- Increasing number of RCTs
- Quantitative evaluation of public policies improved greatly
- There is something like a "culture" of evaluation
- Great scientific successes (Cf Esther Duflo)
- The science of women in science can be cumulative and rigorous!
Thank you for your attention!

Don’t hesitate to reach out!

guillaume.hollard@polytechnique.edu