Centre of Excellence for Development Impact and Learning

March 2022

Design Paper 8

Machine learning methods to uncover mechanisms underlying the impacts of two long-term evaluations of youth skills training programs in Uganda (8-year follow-up)

Laura Chioda and Paul Gertler

London International Development Centre

UK aid from the British people

Oxford Policy Management

About CEDIL

The Centre of Excellence for Development Impact and Learning (CEDIL) is an academic consortium supported by the UK government through UK Aid. The mission of the Centre is to test innovative methodologies in evaluation and evidence synthesis and promote evidence-informed development. CEDIL-supported projects fall into three programmes of work: evaluating complex interventions, enhancing evidence transferability, and increasing evidence use.

CEDIL design paper

CEDIL design papers present the protocols of studies supported by the CEDIL programme. They describe the objectives, the methodology and the implementation plan of impact evaluations, evidence syntheses, and methodological papers supported by CEDIL.

About this design paper

This design paper was submitted to CEDIL by the Machine learning methods to uncover mechanisms underlying the impacts of two long-term evaluations of youth skills training programs in Uganda (8-year follow-up) Project S.180 team.

Please direct any comments or queries to Laura Chioda and Paul Gertler at UC Berkeley, USA.

Suggested citation: Chioda, L., Gertler, P. (2022) ‘Machine learning methods to uncover mechanisms underlying the impacts of two long-term evaluations of youth skills training programs in Uganda (8-year follow-up)’, CEDIL Research Design Paper 8. Centre of Excellence for Development Impact and Learning (CEDIL), London and Oxford.

This project was funded by the Centre of Excellence for Development Impact and Learning (CEDIL), supported by UK aid from the UK Government. The views expressed in this research project paper do not necessarily reflect the UK Government's official policies or CEDIL.

CEDIL The design paper [L.180]

Title: Machine learning methods to uncover mechanisms underlying the impacts of two long-term evaluations of youth skills training programs in Uganda (8-year follow-up)

Authors: Laura Chioda (UC Berkeley) and Paul Gertler (UC Berkeley)

The design paper, per CEDIL guideline, is divided into three sections dedicated to the following topics: policy relevance, innovation, and technical design.

Policy relevance

Between 2010 and 2030, global labor markets will face the daunting task of generating three-quarters of a billion new jobs. As calculated in Bloom et al. (2018), this is the estimated number of jobs required to absorb the projected 21% increase in the world's working-age population, accommodate trends of increasing labor force participation (mainly among the female population) and automation, and achieve target unemployment rates (4% for adults; 8% for youth). Sub-Saharan Africa (SSA) faces an especially substantial challenge, with job needs by 2030 more than doubling the number created between 1990 and 2010.

Youth unemployment in Uganda is the highest in Africa. ActionAid places the country's youth joblessness rate at 62% – but the African Development Bank reports that the figure could be as high as 83%. Furthermore, the share of Ugandans under the age of 30 is the highest in the world at 70%. Against this backdrop, the question as to whether more successful entrepreneurs can be created by endowing them with the right skills becomes even more pressing.

Increasingly, education, and in particular entrepreneurship/skills education, has been recognized as a powerful driver of economic growth; many developed and developing countries have invested heavily in educational programs to increase their "stock" of entrepreneurial skill (Deming, 2017; OECD, 2015). Gaining an exact understanding of which skills are most important for the development of entrepreneurs and how to bolster them has therefore become a central research and policy priority. Indeed, a 37-country study reports that roughly one-third of the 35,000 employers interviewed have trouble finding workers with the right skills. A recent review of 28 studies relying on employer surveys across multiple countries finds that socio-emotional (soft) skills are the first priority in 76.5 percent of the studies that rank employer skill preferences. [1] However, whether such skills can be taught is an open question, as well as a policy priority.

How malleable soft skills are in adulthood and whether training programs that aim to increase the stock of these skills can generate improvements in productivity and life outcomes have only begun to be explored (Campos et al., 2017 and Groh et al., 2016). Mckenzie (forthcoming) reviews the evidence

[1] Manpower (2010), "Supply/Demand: 2010 Talent Shortage Survey Results," Manpower Group: Milwaukee, WI.

emerging from X studies focusing on business training programs for the typical micro and small firms in developing countries. He revisits and reassesses the evidence for whether small business training works, incorporating the results of more recent studies. Emerging evidence is provided on five approaches for improving the effectiveness of traditional training by incorporating gender, kaizen methods, localization and mentoring, heuristics, and psychology. Training programs that incorporate these elements appear to deliver improvements over traditional training programs on average, although with considerable variation. Given that training delivers some benefits for firms, the challenge is then how to deliver a quality program on a cost-effective basis at a much larger scale.

The 8-year follow-ups of two innovative youth skills-development and entrepreneurship interventions, Educate! and SEED, in Uganda, both implemented at-scale as randomized controlled trials (RCTs). which would serve to speak to these knowledge gaps. In particular, this work aims to shed light on the underlying mechanisms and components through which these interventions operate and yield lasting impacts, and to inform the debate on the optimal combination of soft- and hard-skills in the design of entrepreneurship training programs. The two RCTs permit to assess and compare the long-term impacts of three youth skills curricula that feature different combinations of soft and hard skills; namely, SEED-hard (25% focus on soft-skills), SEED-soft (75% focus on soft-skills) and Educate! (~90% devoted to soft-skills). In particular, adapting existing Machine Learning methods, we will conduct causal mediation analysis to study how the programs shape skills, how these are differentially rewarded in the labor market, and their social spillovers (e.g., risky behaviors, partnership quality, and intimate partner violence). Our work will combine the machine learning (i.e., generalized random forest) and causal mediation analysis literatures to go beyond the "effect of a cause" (i.e., the treatment effect) and investigate the "cause of the effect", i.e., the channels through which the effect on final outcomes is manifested.

Our work addresses several key policy-relevant knowledge gaps identified by the literature: (1) improving the understanding of which skills—and which combination of skills—are important for leadership and entrepreneurship and how to teach them; (2) the sustainability of these interventions' impacts; (3) documenting any spillovers beyond the usual economic outcomes, such as risky behavior and IPV; (4) identifying the mechanisms at play underlying effects as well as the subgroups that are most likely to benefit from these types of programs.

Interventions

1. RCT #1: The Educate! Experience program is a two year-long upper-secondary school-based intervention. It aims to train the social entrepreneurs and leaders of the 21st century.

The Educate! NGO aims to enhance skills among youth to help them engage and succeed in both formal employment and entrepreneurial activities in East Africa. The Educate! Experience program is implemented during the last two years of secondary school and delivered within existing secondary schools (government, private, and community schools) by practically-trained youth mentors, who use hands-on teaching methods and practical applications in classrooms and in a Student Business Club. The program's goal is to develop leadership, workforce-readiness and entrepreneurship skills in secondary school students. It teaches youth soft skills including both interpersonal skills—e.g., communication and teamwork—and intra-personal skills—e.g., self-confidence, critical thinking, creativity and grit. It also teaches hard skills such as business planning, budgeting, savings, etc.

The Educate! Experience program consists of three components: Social Entrepreneurship and Leadership Course (SELC), Mentoring, and Student Business Club (SBC).

Social Entrepreneurship and Leadership Course (SELC). The curriculum is taught in English for 80 minutes once per week, and four to nine times per school term for five terms (35 lessons in total). The SELC focuses on developing socially responsible leadership skills, business/entrepreneurship skills, community awareness/engagement, group and individual "personal projects" such as community initiatives and businesses, and group mentorship. Roughly, 80% of instructional time is devoted to soft skills versus 20% of instructional time is dedicated to hard-skill topics.

Mentoring: One-on-One and Group Sessions. Educate! mentors hold one-on-one mentoring sessions, in English, outside of the scheduled SELC lesson time. The goal is for these sessions to happen once or twice a term with each session lasting approximately 15 minutes. These sessions focus on the personal development of the students and are an opportunity to build supportive relationships between the mentors and students. Once per term the mentor holds a group mentorship session to discuss any issues with the entire class.

Student Business Club (SBC). The Student Business Club is focused on business development and meant to help scholars design projects that generate income. Members are responsible for developing and managing the club projects with the guidance of the Educate! mentor. The club starts working by writing a constitution and electing the leadership board. The members decide themselves what kind of business to start and raise funds for this purpose (often through fundraising, their own allowances, contribution from parents, etc.). The Student Business Club meets outside of the scheduled SELC classes. There is no target for the number of times the SBC should meet. The number of meetings depends on the interests and needs of the students. The club is shut down right before the students graduate, with remaining products sold and profits divided between the members. While the mentor oversees decision-making and operations, the students independently manage the entire business creation process from start to end.

2. RCT #2: Skills for Effective Entrepreneurship Development (SEED) is an in-residence 3-week mini-MBA modelled after western business curricula adapted to the Ugandan context. Youth were recruited from schools during their 6th (last) year of secondary school and attended the program during the summer. The study team designed two separate curricula: a hard skills-focused MBA and a soft skills-focused MBA. The hard-skills MBA features a mix of approximately 75% hard-skills and 25% soft-skills; the soft-skills curriculum has the reverse mix. All students get a basic overview entrepreneurship, lectures in how to develop a business plan, and daily work on their own business plan.

Topics covered as part of the **hard skills modules include** (1) accounting covering financial statements (income statements and balance sheets) and the cost structure of an enterprise, (2) business creation covering generating business ideas, selecting a suitable market, money needed to start an enterprise, and legal forms of business ownership (3) management covering sales, keeping track of business operations, hiring and managing people, and selecting suppliers; and (4) finance covering managing money, investment, and obtaining money to start a business. This hard-skills curriculum provides foundational skills so that the students know how to generate a business idea, how to find capital, how to get the business off the ground, and how to manage it.

Topics covered as part of th**e soft skills modules include** (1) learning to control one's self through self-management, goal setting, risk taking, decision making and personal power; (2) persuasion through communication strategies, leadership and team building, group decision making strategies, social networking and public speaking, (3) negotiation through win-win judgment, personal power, win-win judgement ad negotiation strategies. This soft skills curriculum provided foundational skills so that students have more self-awareness, more ability to regulate emotions and delay gratification, and more ability to influence others and make sound decisions.

The SEED curricula were designed to enable highly participatory classroom environments through a variety of exercises and teaching methods. The soft skills curriculum was based on a social leadership and entrepreneurship curriculum developed and used by Educate! in their programming. The hard skills curriculum was based on the Know About Business curriculum developed and used by the International Labor Organization in Africa. Both soft and hard skills curricula integrated content from the MBA program of the Haas School of Business at UC Berkeley. The curricula were extensively pilot tested to ground them in the local language and conceptual environment, to make them relevant and salient to local business institutions, and to ensure that course content was appropriate and well understood, and that the teaching methods were interesting and engaging.

The hard and soft residential mini-MBAs are centered on 17 sessions and detailed lesson plans were developed for each session. Lesson plans include a series of detailed steps for the teacher to follow, framed by objectives and rationale to encourage a sense of ownership by the teacher. The plans covered learning objectives, key concepts, case applications, presentation materials and a series of interactive exercises. The lesson plans were designed to be used an intense Western-Socratic based course designed to enable a learning process that encourages sharing, discovery and reflection while maintaining a comparable classroom environment between curricula and across all training sites.

Finally, there were also a series of planned social activities that were implemented consistently across all training sites. Social activities took place nearly every evening and included debates, a cultural night (in which students representing different cultures rehearsed and performed dances in costume), movie nights, a spelling bee, a lip-syncing competition ("miming"), stand-up comedy, quiz nights, and a social dance. In addition, sports activities took place each afternoon before dinner. Basic sports equipment (volleyballs, netballs, and soccer balls) was purchased for each school.

Objectives. We will study long-term, gender-disaggregated impacts, and the mechanisms underlying the impact for the two interventions. The two studies will permit comparisons between curricula featuring different intensities of soft- and hard-skills. Not only will the two RCTS provide concrete guidance for policy and curriculum design, but it also represents a unique opportunity to fill existing knowledge gaps. Furthermore, the large sample, the focus on entrepreneurial and leadership mindsets, as well as the characteristics of the population targeted by the intervention (male and female youth at the onset of their productive lives) make this work both research- and policy-relevant in light of the challenges that developing countries will face to create millions of jobs by 2030 to absorb a rapidly growing global workforce.

The analysis will focus on key determinants/mediators of the observed changes in outcomes and their relative importance. We are particularly interested in shedding light on the mechanisms underlying observed impacts and on the sources of heterogeneous responses to the interventions. To this end, we will build on existing machine learning methods (Athey, Tibshirani, Wager,2019, Wager and Athey, 2018, Athey and Imbens, 2016). **Data-driven HTE** approaches are agnostic as to the sources of heterogeneity and help isolate the characteristics of the most and least affected subpopulations. We plan on extending these methods to conduct causal mediation analysis, which will be valuable in welfare analysis and to shed light on the mechanisms that explain long causal chain.

We are interested in understating how these programs shape skills, how these skills are rewarded in the labor market, and the extent of social spillovers (e.g., risky behaviors, fertility, partnership quality, and IPV), which may highlight possible trade-offs between strictly economic returns vs. socially desirable outcomes. These findings will speak not only to the sustainability of impacts over time, but will also provide a more complete picture in terms of cost-effectiveness of these programs. As highlighted by Bouguen, Huang, Kremer, Miguel (2018), programs targeting human capital investments in both health and education appear to have high rates of return, making them potentially attractive for public policy. However, to our knowledge, this is the first study that endeavors to document the long-term impact of soft-skills interventions targeting youth.

The two RCTS reliance on data collected at multiple points in time (baseline, 4-year follow-up, 8-year follow-up) is an important feature of this project. It enables us not only to increase the studies statistical power, but also to study the trajectory of impacts. As McKenzie and Woodruff (2017) noted, most of the existing literature on business training does not examine the effects of the training at different points in time. A 3ie (2015) assessment stressed the "need for future impact evaluations of youth and transferable skills programs to measure outcomes further along the causal chain, […] such as wellbeing, health, income, employment and livelihoods." The few studies that have documented impacts over time often find that effects disappear in the longer term (Blattman et al., 2018 and Blattman et al. 2019, Bloom et al., 2018a, Ubfal et al., 2019). Our work will speak to these knowledge gaps and offer concrete policy and program advice.

The Educate! and SEED's medium run impacts are promising, but the timing of the two interventions calls for longer-term data collection efforts to measure the temporal evolution of returns to skill-upgrading and their social spillovers. For instance, one of the interventions, SEED, demonstrated that training was effective in improving both hard and soft skills, but that only soft skills were directly related to improvements in self-efficacy, persuasion, and negotiation. Both hard and soft skills were rewarded in the labor market after 3 years, however only soft skills increased enterprise starts and their profitability. In comparison, the other intervention, Educate!, led to additional investments in education, which in turn delayed full-time labor market participation. These medium-run follow-ups are not able to capture the full extent to which the labor market values these skills and additional educational investments. Are these gains sustainable? Only a longer-term follow-up will permit us to answer this key policy and research question and shed light on the returns to soft skills, while informing curriculum design.

Immediate and direct impact on program design and policy dialogue, thanks to our partnership with Educate! Educate! prepares African youth with the skills they need to succeed in today's economy. For 10 years, Educate! has rigorously designed, tested, and iterated on a skills-based model of education to improve the quality of secondary education systems to impact life outcomes of youth. Educate! has partnered with researchers to study its programming, including two RCTs and a quasi-experimental evaluation. Educate! was recognized by the World Bank (S4YE) and the UN, received the 2015 WISE award and the 2018 Klaus J. Jacobs Prize, was featured as a case study on scaling education by the Brookings Institution, and was highlighted by Bill Gates.

In keeping with a history of using evidence to inform program design, Educate! has been using these and previous results to update the program's curriculum. Educate! also plans to integrate aspects of another evidence-backed entrepreneurship model, called Skills for Effective Entrepreneurship Development (SEED), into its work.

Given our partnership with the Educate! NGO, this study has a tremendous the potential to influence the regional dialogue and influence the implementation of these programs at the national level. Educate! engages with and advises national governments to inform policy and support reform. It advises national governments on curriculum design and teacher training initiatives. Since 2012, Educate! has partnered with Uganda's government to integrate its curriculum and student business club structure into Uganda's entrepreneurship course nationally; in 2015, it served as a technical advisor to the Rwandan government on its secondary education reform; and in 2016, it began serving as a technical advisor to Kenya on the reform of its national curriculum. The immediate policy impact is significant since Educate! is today scaling its model to over 500,000 youth in Uganda, Rwanda, and Kenya, preparing to replicate its model across Sub-Saharan Africa (SSA), and deepening its engagement with governments across the region, advising them on education policy.

Technical Design

Study design Educate!: Between January and April 2012, six districts were selected out of the 111 districts in Uganda: Iganga, Jinja, Kampala, Masaka, Mbarara, and Mukono. These six districts are the most populous districts that have at least eight A-level schools with more than 40 students in their first year of upper secondary education (known as S5 students). Eight schools from each district were randomly selected to be included in the Educate! RCT study, bringing the total sample to 48 schools. Twenty-four schools, stratified by district, were randomly assigned to the treatment group (i.e., to participate in the Educate! program), while the other 24 schools were randomly assigned to the control group. In May 2012, a short survey was administered to all S5 students in the selected schools to ascertain interest in participating in a leadership and social entrepreneurship course, determine previous leadership and/or entrepreneurial experience, and assess literacy levels and cognitive ability. All 5,048 interviewed students were assigned a score based on the survey, and the top 45 S5 students in each school were invited to participate in the Educate! program for the rest of the academic year and during the following one, conditional on Educate! offering the program in their school. The final sample consisted of 1,942 study participants, including 976 control and 966 treatment (49.7 percent) participants, making the sample was well balanced between the two groups. The Educate! program was implemented and successfully completed during the 2012 and 2013 school years, (i.e., five full school terms).

Study design SEED: To ensure representativeness at the national level, students were recruited from 200 full time secondary schools randomly selected from the universe of Uganda's approximately 700 full time secondary schools. In October of 2012, promoters went to each of the 200 schools to recruit students in their 6th (last) year of secondary school. Students interested in the program were asked to fill out both a simple application form and a baseline survey. Response to the SEED program opportunity was overwhelmingly positive. Nearly, 8080 students applied to the program. However, 313 (3.9%) were dropped due to problems with the student ID, 113 (1.3%) because the student failed to indicate their gender, and another 223 were dropped because they were already participating in another entrepreneurship training program. The remaining 7,431 served as target population for the study.

Power calculations indicated that we needed 1200 students in each arm (hard skills treatment, soft skills treatment, and control group), or 30 students in each of the 80 classes. Taking into account attrition

SEED! – 3.5-year follow-up Summary Results.

SEED's 3.5-year follow-up demonstrated that training was effective in improving both hard and soft skills, but only soft skills were directly linked to improvements in self-esteem, persuasion, and negotiation. The skill upgrade translated into important economic and labor market benefits. Both hard and soft skills trainings are rewarded in terms of wages relative to the control group.

Youth in the hard and soft SEED trainings were more likely to start new businesses (9.2 and 10.1 pp respectively) relative to youth in the control group, 50% of which started a business. Youth in the soft skills-focused training were also more likely to start a formal business (9 pp) and these formal businesses are more likely to survive over time (6.7 pp) relative to those created by the control group. Youth in both treatment groups are also more likely to start and ensure survival of informal businesses relative to the youth who receive no treatment. Taken altogether, while SEED trainees are overall more likely to start businesses that survive over time, those who received the soft skills-focused training appear more likely to start higher quality businesses as proxied by their formality. Youth who undertook the soft skill training are more likely to create higher quality (formal) businesses. Benefits, as proxied by 3.5 years of cumulative earnings. SEED lead to 31.7% and 30.7% increases in earnings for youth who attended SEED-hard and SEED-soft trainings, respectively. On average this corresponds to an additional $306 per, which sets the Benefit/Cost ratio to 2.6:1.

This is a conservative benefit measure since it abstracts from any impacts resulting from social spillovers or in terms of mental well-being, which are both quite difficult to monetize. This said, the nonmonetary benefits of such skills training programs – and of SEED training, in particular – can be quite sizable. SEED graduates report being more satisfied with their quality of life (0.16sd. and 0.173sd for hard and soft-skills treatments, respectively); as documented earlier, they also report less stress.

Costs. Total intervention costs were estimated at $118 per student. These costs include the salaries and benefits for trained teachers (10%), curriculum development and testing (2%), teacher recruitment and training (4%), site management staff costs (4%), venue, transports, meal, staff plus other students costs (67%), training materials and supplies (5%), supervision and support to teachers (2%), and administration country office (5%).

estimates from similar programs and the long period of time between recruitment and program implementation, the evaluation team over-enrolled by one-third and accepted approximately 1600 students in each treatment arm, or 40 students per class. The control group remained at about 1200 students. Therefore, we randomly sampled 4,400 students out the 7,436 valid applicant pool, stratified by school and gender, to participate in the study. We then randomly assigned 1600 to each to the treatment groups and 1200 to the control group stratified by school and gender.

In order to minimize travel time, recruitment schools were clustered into 4 geographic regions and students are participating in the study would be sent to a randomly selected school in their region. Of the 200 schools visited during data collection and recruitment, 20 eventually served as training sites. The distribution of training sites was chosen based on the number of individuals in the sample surveyed in each region in order to achieve fairly even numbers across regions. Roughly 40% of the sample attended school in the West, 40% from the East (including 20% in the Jinja area and 20% in the Mbale area), and 20% from the North. As a result, 8 sites were located in the West, 8 in the East, and 4 in the North for a total of 20 training sites.

Power Calculations. Four and 3.5 years after the intervention, for Educate! and SEED respectively, approximately 90% of the original samples were successfully tracked over time. The resulting samples of

Educate! – 4-year follow-up Summary Results.

Our four-year follow up of Educate! in Uganda indicate that graduates saw meaningful, lasting increases in their soft skills after participating in the program. The Educate! Experience program has been extremely successful in improving youth's intra- and inter-personal soft skills. Educate! graduates appear to focus more on long-term goals and report being more in control of aspects of their lives, as well as more empowered to implement actions towards achieving their plans.

Overall, Educate! graduates did not appear to possess more business knowledge than the comparison group. However, they were more likely to identify business opportunities and demonstrated better aptitude for deliberate dialogue. While the study found no impacts on political participation, community engagement, or trust in institutions, graduates exhibited significantly more prosocial attitudes — defined as an individual's care for the community and intent to help others.

The shift toward long-term planning is also accompanied by additional investments in education, especially among women, who were 11% more likely to be enrolled in tertiary education. In general, those who participated in the Educate! program were more likely to graduate from secondary school and Educate! graduates were more likely to select business and STEM majors in tertiary education, pathways to possible improvements in future economic outcomes.

The program delivered positive and lasting social spillovers as well. Youth who participated in the program reported having fewer sexual partners, delayed family formation, and had fewer children than those in the control group. Students who participated in Educate! expressed more gender egalitarian views and were less likely to justify IPV; women in the treatment group were also 18% less likely to report threats and incidences of physical violence than women in the control group.

These preliminary results found no relationship between Educate! training and employment rates, wages, or earnings. At the time of the follow-up, 35% of the sample was still enrolled in tertiary education; as such, it was too soon to fully assess economic impacts. The research team is currently fundraising to carry out an 8-year follow-up to assess long term impacts on an array of outcomes, including labor market outcomes.

3,890 for SEED and 1,597 for Educate! are well-balanced across treatment and control groups, yielding studies that are well-powered: Minimum Detectable Effects are quite small (Educate! ~0.12sd; SEED, 0.08-0.11sd). Note that, when splitting the sample by gender, the within-gender variance in outcomes is expected to be smaller than the variance across the full sample (i.e. across genders), allowing for a well-powered within-gender analysis. Overall, the study is well-powered relative to other studies of a similar nature. E! Intra-cluster correlation is quite low, for power calculations was set to 0.01 based on previous rounds of data collection.

Theory of change for both RCTs.

The theory of change underlying the two RCTs for the full sample of youth can be summarized as follows.

- *The problem*:
    - Multidimensional poverty among youth and life outcomes (lack of skills, low quality labor force participation, poor life-outcomes, and risky behavior)
- *Target Population*:
    - Youth before their transition from school to the labor market
- *Entry points*:
    - SEED: summer following the end of secondary school
    - Educate!: last year of secondary school
- *Inputs*:
    - Entrepreneurship training; hard and soft skills (the two RCTs feature three curricula – Educate!, SEED-Hard, and SEED-Soft. They differ in their relative focus on hard vs soft skills)
        - Soft skills: goal-setting, persuasion, leadership and team building, value of cooperative solutions, negotiations, emotional regulation
        - Hard: business development, budget management, and access to capital
        - Educate! only: strong focus on social leadership and social entrepreneurship
- *Outputs*:
    - Youth attend training
- *Outcomes*:
    - Improved soft skills and (hard) technical knowledge indicators
    - Increased SES outcomes (schooling, economic participation, income generation potential)
    - Reduced risky behaviors (e.g. sexual behaviors) and delayed family formation
- *Long Term Outcomes*:
    - Well-being indicators (income, mental health)
    - Income generating potential and improved livelihood and aspirations
    - Social spillovers (social norms, reduced IPV)

Summary of theory of change related to intimate partner violence (IPV) outcomes

Both interventions operate in the short and medium runs on proximate causes of IPV, such as control over resources, perceived social norms, partner quality.

Specific IPV-Relevant Channels.

- a traditional economic/control over resources channel, as the program aims to expand participants' education and economic opportunities;
- improvements in both male and female participants' bargaining and soft skills could have spillovers onto personal dimensions of their lives, including conflict resolution, sexual behaviors, fertility, and overall decision making;
- female business-development and mentors/teachers may shift the aspirations of women who may otherwise lack economically active and successful female role models; similarly, men's exposure to their partners' peers in non-traditional gender roles and to female mentors/teachers may shift their perceptions about women's productive potential and gender roles, thereby reducing backlash and the likelihood of IPV;
- an expansion of aspirations, improved bargaining skills, better education and enhanced income potential could affect the quality of partners in the marriage/partner market.

These channels and intermediate outcomes may weaken the tolerance for IPV, lead to better and more cooperative decision making within the partnership, better outcomes in the market for partners (if partners are selected based on more egalitarian gender roles), better educational and economic outcomes, and to stronger preferences for more cooperative interactions and better decision making, which may in turn further contribute to reductions in the incidence of violence.

Risks and Mitigation Strategies: As acknowledged and briefly discussed in the proposal's narrative, a key risk/concern of any data collection and of a long-term study is the ability to successfully track study participants. 89% of the baseline sample was tracked after 3.5 years. The same re-contact rates are feasible for our 8-year follow-up for the following reasons.

- At each round of data collection, including our 3.5-year follow-up, we have collected extensive contact information. This includes not only respondents' addresses and phone numbers (main and secondary phone numbers, up to 4 different contacts), but also the contact information of family, friends, and colleagues, which have been proven extremely effective at all points of our data collections/tracking.
- The research team has extensive tracking experience from the recent studies. We employ a two-phase tracking strategy, as was done for the 3.5-year follow-up. In addition, for the 8-year follow-up, we will carry our ***two tracking exercises***. We are currently wrapping up the first phone tracking survey, with expected date of completion at the end of this month or early December. The 1sr tracking exercise is comprised of two phases:
- **Phase 1:** All study participants and their alternate contacts are contacted at regular intervals (approximately every three hours, three calls per call-day for each phone number) during a call day for up to three call days over 6/7-day periods. At the end of phase1, we reached over 80% of the original study participants.
- **Phase 2:** Typically, this would be an in-person short-distance tracking. However, mindful of Covid-19-related risks, we have opted to implement this phase 2 also as a phone-based survey. Restricting attention only to the hard-to-reach contacts (those not reached during the 1st phase), we have randomly re-assigned these IDs to different enumerators and for a new round of calls. We were successfully able to contact 40% of these hard-to-reach IDs, which will lead to an overall successful

contact rate in the 88-90% range.

We have not observed statistically significant differences in contact rates across treatment statuses. The best way to minimize (nonrandom) attrition bias is to achieve very high recontact rates. The above successful contact rates are in line with those from the 3.5-year follow-up. It is worth noting that the power calculations[2] in the previous section are for 89% of the original sample tracked during the 3.5-follow-up, which makes them extremely relevant in light of the results of the 8-year follow-up tracking.

- During tracking, we monitored contacts rates by treatment assignment and, should differential (problematic) re-contact rates emerged, we would have adapted call protocols and contemplated a third phase to recover IDs to achieve balance across treatment assignment. The advantage of planning for a second tracking exercise in Q1 2021 is that we will also be able to formally check balance of baseline characteristics and adapt our second tracking to intensify efforts to balance our sample.
- **Several elements explain our successful tracking exercise.** We have extensively and carefully piloted several aspects of our phone survey: call protocols, use of SMS ahead of the calls and as appointment reminders, in addition to adequate incentives for respondents. We will conduct a 2nd tracking exercise in early 2021, shortly before the 8-year follow-up, to verify and update location and contact information. The 8-year follow-up will be implemented via two separate 30/40-minutes phone surveys. From our experience with piloting activities and implementing the phone tracking survey, we have learnt that two separate surveys minimize the burden on respondents and can be quite effective in achieving high response rates when coupled with adequate survey incentives to compensate respondents for their time.

A second tracking in planned for the first quarter of 2021. Both tracking exercises are planned/implemented as phone-based surveys; all the training and piloting activities are also done remotely. The government of Uganda has expressed a willingness to continue permitting in-person data collection for the foreseeable future, provided safety measures are observed.

Ethics. We expect the possible risks to study participants to be minimal. As with past rounds of data collection, the study protocol and survey instruments will be reviewed by IPA's Institutional Review Board and the Mildmay Uganda Ethics Review as well as by the Uganda National Council of Science and Technology. The research team will take special precautions in administering the IPV module to study participants. Enumerators will be trained in accordance with WHO guidelines, including a basic introduction to IPV and overall orientation on the concepts of gender and gender discrimination/inequality. Priority will be given to (i) ensuring participant safety – the interview will be conducted in a private area; enumerators will be trained to change topics in the event of an interruption and/or if they notice someone listening in; to (ii) minimizing participant distress – enumerators will be trained to recognize signs of distress and take appropriate steps to support the respondent and/or

[2] **Power calculations**. Three and a half years after the intervention, we successfully tracked 89% of the original sample. The resulting sample of 3,890 youth (35% female) was well-balanced across the two treatment arms and the control group. Minimum Detectable Effects are quite small (0.08-0.11sd). Note that, when splitting the sample by gender, the within-gender variance in outcomes is expected to be smaller than the variance across the full sample (i.e., across genders), allowing for a well-powered within-gender analysis. Overall, the study is well-powered relative to other studies of a similar nature.

terminate the interview; enumerators and respondents will be gender-matched; and to (iii) provide referrals – the research team will refer participants to toll-free hotlines to the Communication for Development Foundation of Uganda for information and counselling.

Outcomes of interest and measurements. Some of the key outcomes of interest include:

We are interested in studying how these programs shape skills, how these skills are rewarded in the labor market, and the extent of social spillovers (e.g., risky behaviors, fertility, partnership quality, and IPV), which may highlight possible trade-offs between strictly economic returns vs. socially desirable outcomes. All our measures have been validated and tested.

1. Skills (soft and business-related): Practical business skills include business planning, management, accounting and finance, financial literacy, savings behavior and use of financial institutions; soft skills include creativity, patience, self-efficacy, depression, grit, and big five.

2. SES and labor market outcomes: education, social standing, entrepreneurial success (starting, investing in and expanding business, business profitability, cost data), wages, total earnings from all sources of income; aspirations, time use, access to lending (formal and informal), indicators of extensive and intensive margins of economic participation, employment formality and benefits , etc. ***This is not a COVID-19 study***, but we will also try to collect measures of economic resilience and coping strategies, since our treatments may have endowed program participants with skills to better cope with and weather the crisis.

3. Partnerships, family formation, and IPV-related outcomes:[3] incidence of physical, psychological, and financial violence; partners' self-reported decision-making surrounding economic decisions, freedom of movement, social interactions and sexual behaviors, fertility, HIV testing, contraceptive use; consumption and female labor force participation; attitudes toward traditional gender roles and division of labor within the household; use of and justifications for physical and psychological violence; disclosure of violence and willingness to reach out to others and share personal experiences; community expectations about prevention and social norms. Self-reported assessments of social norms will be complemented with experimental (scenario) vignettes. Our instruments comply with best practices as summarized by Glennester et al., (2018), and WHO-recommended instruments to measure gender empowerment and sensitive topics. (Note: We will adopt IPV instruments previously validated in the Ugandan context, including typical measures found in the Uganda DHS, as well as key indicators from the SASA! evaluation in Kampala (Abramsky et al. 2014, 2016). This permits comparability of measures across studies, which will be important in benchmarking results across regions and interventions.)

4. Soft skills and psychological assessments: Big5, grit, self-efficacy, negotiation/persuasion pro-sociality, stress management, depression, etc. Because our focus is on underlying mechanisms and on the role of soft skills in shaping ultimate outcomes, it is important to obtain reliable measures of soft skills. In addition to standard self-reported validated psychological instruments, we plan on including task-based measures of soft skills.

[3] We will adopt IPV instruments previously validated in the Ugandan context, including typical measures found in the Uganda DHS, as well as key indicators from the SASA! evaluation in Kampala (Abramsky et al. 2014, 2016). This permits comparability of measures across studies, which will be important in benchmarking results across regions and interventions.

Social Desirability Bias Concerns. For some of the questions (e.g. around IPV, risky behavior, and number of partners) one could be concerned that social desirability bias may lead to more favorable responses in the treatment group. The following steps mitigate concerns about desirability bias:

- The Educate!/SEED intervention do not touch upon any of the gender-related and risky behavior outcomes, nor those regarding IPV and social norms, either directly or indirectly. The two programs focus exclusively on leadership and entrepreneurship skills;
- As mentioned above our instruments comply with the best practices as summarized by Glennester et al. 2018, as well as WHO-recommended instruments to measure gender empowerment and sensitive topics. For instance,
    - They feature several checks and validations across measures to check consistency. In addition to stated preferences, we also ask directly about couples' decision-making surrounding expenditure and labor market participation decisions, as well as fertility and contraceptive use.
    - We combine responses from several survey questions into indices (e.g. social acceptability of IPV, incidence of physical and emotional violence).
    - Triangulation can help capture outcomes that are challenging to measure or susceptible to response bias.

Innovation and Methodology

The estimating equations are informed by the fact that our two studies are RCTs. We will be able to estimate the causal impacts of the Educate!/SEED treatments by comparing the sample means of the treatment and control groups or, equivalently, by estimating the following linear regression models:

$$y = \alpha + \beta T + \varepsilon \qquad \text{(Educate!)}$$

$$y = \alpha + \beta_1 T_{soft} + \beta_2 T_{hard} + \varepsilon \qquad \text{(SEED)}$$

where y is the outcome for a given individual; and T is a treatment indicator. For SEED, we will capture both hard and soft skills treatment impacts. The βs denote the coefficients of interest. In the case of Educate!, standard errors will be clustered at the school-level, which is the level at which treatment was assigned.[4] To boost efficiency of our estimated average treatment effects, we will also run specifications including baseline controls. This step may not only increase the precision of estimates but can also address possible imbalance of certain baseline characteristics across treatment and control groups. To guard against unscientific data mining and specification searching, we adopt data-driven approaches such as Double/Debiased ML.[5]

[4] Because assignment to treatment took place at the school level, standard errors should be clustered at the unit of assignment. This permits heteroskedasticity and within-cluster error correlation. We use wild bootstrap procedures with Rademacher weights to obtain more accurate cluster-robust inference that allows for a relatively small number of clusters (Cameron and Miller 2015).

[5] Chernozhukov et al. (2014).

In addition to estimating treatment effects, the analysis will focus on key determinants/mediators of the observed changes in outcomes and on their relative importance. We are particularly interested in shedding light on the mechanisms underlying observed impacts and on the sources of heterogeneous responses to the interventions. To this end, we will build on existing machine learning methods (Athey and Imbens, 2016; Wagner and Athey 2018). Here again, data-driven heterogeneous treatment effect (HTE) approaches are agnostic as to the sources of heterogeneity and help isolate the characteristics of the most and least affected subpopulations. We have extended these methods to conduct causal mediation analysis, which will be valuable in welfare analysis, and to shed light on the mechanisms that explain long causal chains. In summary, we plan on going beyond the "effect of the cause", i.e., the treatment effect, and investigate the "cause of the effect", i.e. the channels through which the total effect on outcomes is manifested.

Methodology – Machine Learning Methods: Causal Trees and Random Forests

Mid-level theories start with an empirical phenomenon (as opposed to a broad abstract entity like social systems) and abstract from it to create general statements that can be verified by data. They can be thought of as a consolidation of tendencies or mechanisms. To identify these mechanisms and empirical regularities that will become building blocks for mid-level theory, we plan to exploit recent developments in Machine Learning (ML) methods. Exploring treatment heterogeneity can provide valuable information as to which mechanisms drive results; for instance, which skills matter in determining economic outcomes or social spillovers, or whether the programs act as complements or substitutes for initial skills or for SES at baseline. Whereas ad hoc searches for particularly responsive subgroups may mistake noise for a true treatment effect, the proposed approaches[6] adapt machine learning methods to discover particular forms of heterogeneity by seeking to identify subgroups that have different treatment effects in a principled way. The data-driven approaches to study heterogeneity in treatment effects have the advantage of being agnostic as to the sources of heterogeneity and therefore avoid any concerns about ex-post data mining; they allow to study forms of heterogeneity that were not specified in a pre-analysis plan, without invalidating confidence intervals.

The key references are Athey and Imbens (2016), Wager and Athey (2018) and Athey, Tibshirani, and Wager (2019), who extend regression tree[7] and random forest algorithms to the problem of estimating average treatment effects for different subgroups and derive estimates of personalized treatment effects that change smoothly as function of baseline characteristics.[8] These approaches are tailored for applications where the unit of observation may have many attributes relative to the number of units observed.

The rest of this section is organized as follows. We first provide a quick overview of these methods; we then explicitly draw a parallel between them and mid-level theories; and conclude the section by providing two examples to better illustrate the parallel and demonstrate how the findings in this paper will permit

[6] Athey and Imbens (2016), Wager and Athey (2018), and Athey, Tibshirani, and Wager (2019).

[7] The general idea behind a regression tree is that it segments the predictor space into a number of simple regions.

[8] These approaches overcome the adaptive nature of traditional ML methods, which use the training data for model selection, leading to biases that disappear only slowly as the sample size grows.

to generalize and inform policy decision making in settings that differ from those in which our two RCTs were implemented.

- **Causal trees**: combining well-established tools from the potential outcomes and ML literatures, **Athey and Imbens (2016**) build on earlier work by Su et al. (2009) and Zeileis, Hothorn, and Hornik (2008). The method is based on the regression trees machine learning method, but they use a different criterion to build the tree: rather than focusing on improvements in mean-squared error of the prediction of outcomes (*y*), they focus on mean-squared error of treatment effects. In other words, the criterion used for tree construction is to estimate *heterogeneity in treatment effects* rather than heterogeneity in *outcomes*. The output of the data-driven approach is a set of subgroups, selected to optimize treatment effect heterogeneity (by minimizing expected mean-squared error of treatment effects), together with treatment effect estimates (and associated standard errors) for each subgroup. As a result, the global model is mapped into two data-driven components: one is the recursive partition (tree structure), the other is a simpler model for each cell (leaf) of the partition.
- **Random Forests and Generalized Random Forests (Wager and Athey, 2018 and Athey, Tibshirani, Wager, 2019)** build on Athey and Imbens (2016) and propose a method for estimating HTEs based on random forests, by generating many different trees and averaging over them, except that the component trees are now Athey and Imbens (2016) causal trees. Random forests are collections of randomized regression trees. Each regression tree forms predictions by recursively partitioning the data. Random forests are among the most popular machine learning algorithms and have been widely used for prediction. Relative to a causal tree, which identifies a partition and estimates treatment effects within each element of the partition, the causal forest leads to estimates of causal effects that change more smoothly with covariates and, in principle, distinct average treatment effects can be derived for each individual; i.e., their method delivers conditional average treatment effects (CATEs). Wager and Athey (2018) show that the predictions from causal forests are asymptotically normal and centered on the true conditional average treatment effect for each individual. They also propose an estimator for the variance, so that confidence intervals can be obtained.

Here we provide an illustration of the types of results and statements we will be able to make using these methods, relying on data from our Educate! 4-year follow-up. Figure 1 presents the individual CATEs for one of our outcomes of interest: completion of secondary education.

Important remarks:

- ATE: Educate! graduates are 3.68 percentage points (pp, or 4 percent) more likely to complete secondary school relative to the control group (88 percent graduation rate), a statistically significant effect (one sided p-value = 0.03). The point estimate for the female subsample is three times as large as that for the male subsample (6.6 pp or 8 percent) and is statistically significant (one sided p-value = 0.01), thereby virtually closing the gender gap in graduation.

- The quintile CATEs are all greater than zero and statistically distinct from it, implying that the program positively impacted the likelihood of graduating for all quintiles of the distribution. We also observe a great deal of heterogeneity in the distribution of CATEs (see Figure 2).
- Generalized Random Forests will also permit us to study how heterogeneity in the CATEs relates to the nodes (baseline characteristics) that have been used to build the random forests. Intuitively, each tree (in the forest) induces a partition of the data that will identify subsets of the data and combinations of the baseline characteristics, Xs, for which ATEs within cells are significantly different from each other. A forest will average over all the tree-induced partitions.

Figure 1: Educate! Individual CATEs for Completed Secondary (Generalized Random Forest)

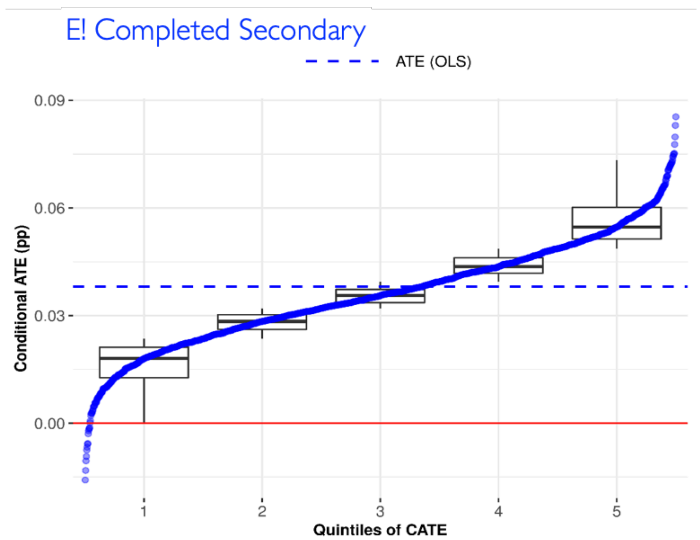
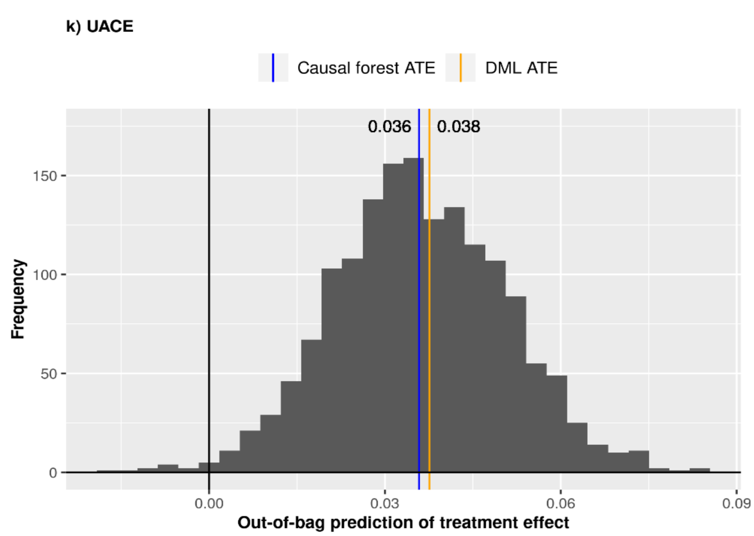


Figure 2: CATE distribution, Average CATE, and ATE



It is sometimes helpful to have a more expressive summary of the CATE function $\tau(x) = E[Y(1) - Y(0) \mid X = x]$ than the average treatment effect. One such summary would be a function that relates CATEs to set of features; for example, a subset of the Xs used to train the forest. In other words, we may be interested in understanding how the CATEs for a given outcome vary as function of all or a subset of baseline characteristics.

- For instance, implementers and policymakers might not only be interested in whether the Educate! program was effective in improving students' graduation rate (ATEs, and CATEs) achievement.
- They may also be interested in whether the program is likely to be a complement to or a substitute for their household socio-economic status (SES).
    - Does the program benefit youth from low or high SES households? Does the program level the playing field in terms of economic opportunities, or does it amplify baseline inequities?
    - Is the effect of the intervention moderated by pre-existing IQ, or does the program benefit youth irrespective of their innate abilities (proxied by Raven's tests)?
    - Do other covariates moderate treatment effects?

Generalized Random Forest and the HTEs approach outlined here will permit us to answer such questions.

CATEs and Heterogeneity and External validity. A key concern with randomized experiments is external validity. "External validity concerns inferences about the extent to which a causal relationship holds over variation in persons, settings, treatments, and outcomes." (Shadish et al 2002, p. 83). If applied in a different setting, will the treatment deliver the same effect? While numerous factors vary across settings, one common way in which settings differ from one another is that the individuals that populate them may be different. If these differences can be captured with observable pre-treatment variables, then it is in principle possible to address this element of external validity (as in Hotz et al., 2005). In particular, if we obtain an estimate of the treatment effect (CATEs) for each potential value of the covariates measured at baseline, then we can estimate average treatment effects in any population by accounting for the differences in distributions of baseline characteristics. That is, if $HTE(y|X=x)$ is known, it is straightforward to compute the average treatment effect $E(HTE(y|X=x))$ if the distribution of X is known (Athey and Imbens 2017).

- For instance, our figures 3 and 4 suggest that underlying heterogeneity results from the skill program being particularly effective for youth with high anxiety and low socio-economic status. The partitioning of the data space thus points us to a simpler model of program effectiveness that permits to understand not only to what extent the findings are generalizable, but also suggests possible theories of underlying mechanisms and associated testable predictions. That is, the program was able to deliver large impacts by improving soft skills, but also acted as a substitute, i.e., it compensates for low SES, to prior family background characteristics.

Figure 3: CATEs for Completed Secondary as a Function of Youth's Family SES at Baseline

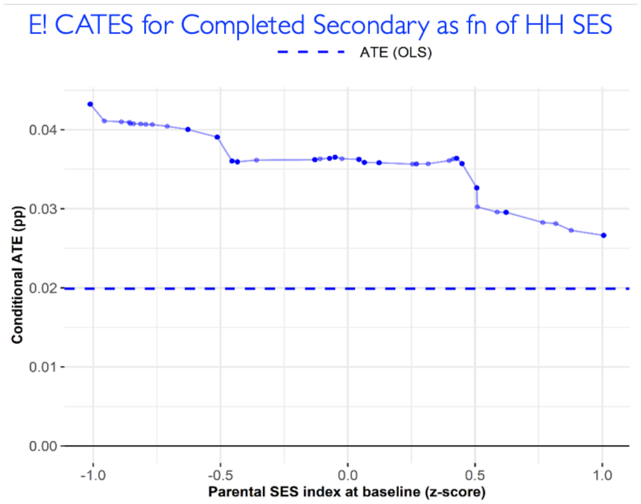


Figure 4: CATEs for Completed Secondary as a Function of Youth's Anxiety at Baseline

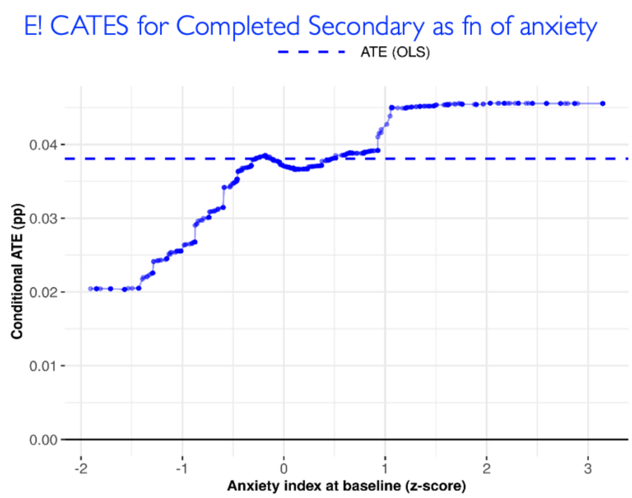
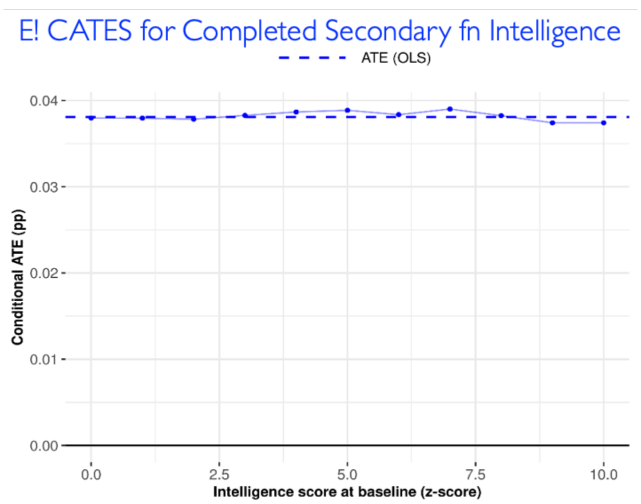


Figure 5: CATEs for Completed Secondary as a Function of Youth's Intelligence (Raven) at Baseline



Causal Mediation Analysis and Underlying Mechanisms

Following the notation in Masset and White (2019) and restricting attention for simplicity to labor market outcomes, assume earnings (y) is a function of the inputs (i.e., intermediate outcomes/mediators) affected by the program (h; Educate! or SEED-Hard or SEED-Soft), which is a complex intervention (i.e., it involves mentors, teaches hard skills, business idea development, an array of soft skills development, pedagogical interactive models, etc.). The object of interest of traditional RCTs is a valid estimate of the impact of the program that answers the question as to whether the complex bundle (h) was effective in improving y (in our example, earnings).

A natural query in this setup relates to mechanisms that underlie the total effect on economic outcomes. For instance, which set of skills were important in shaping economic outcomes? Were there changes in extroversion, risk aversion, grit, executive functions, stress management skills? Or did changes in intra-personal skills such as negotiation, persuasion, or acquired knowledge in business practices matter in shaping outcomes of interest? What is the relative importance of these skills, and what changes in skills delivered the largest impacts on y based both on baseline characteristics and on changes in skills? Our methodological contribution will not only permit to identify which skills matter, but also to study how impacts vary with impacts on intermediate outcomes. Key references are Huber (2020), Huber (2019), Farbmacher, Huber, Laffers, Langen, Spindle (2020)

Mediation analysis typically aims at decomposing the average treatment effect (ATE) of a binary treatment, denoted by T, on an outcome variable, Y, into a direct effect and an indirect effect operating through a mediator, M.

Using the potential outcomes framework's notation, the ATE for outcome Y is given by

$$\text{ATE}(Y) = E[\ Y(1, M(1)) - Y(0, M(0))\ ]$$

which can be decomposed into both a direct effect (of treatment T on outcome Y) and an indirect effect (of the mediator M on Y):

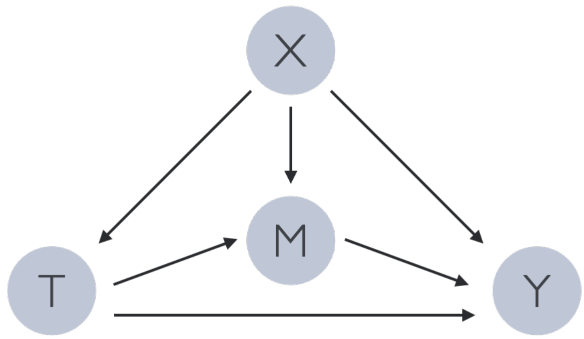
- **Direct Effect**: $\theta(T) = E[Y(1,M(T))-Y(0,M(T))],\ T\in\{0,1\}$
    - equals the difference in mean potential outcomes when switching treatment status while keeping the potential mediator fixed, which blocks the causal mechanism via M;
- **Indirect Effect**: $\delta(T) = E[Y(T,M(1))-Y(T,M(0))],\ T\in\{0,1\}$
    - equals the difference in mean potential outcomes when changing the potential mediator's values while keeping the treatment fixed to block the direct effect.

The vast majority of identification strategies relies on selection-on-observables-type assumptions, implying that the treatment and the mediator are conditionally exogenous when controlling for observed covariates.[9]

[9] Conti, Heckman, and Pinto (2016), Huber (2015), Huber, Lechner, and Mellace (2017) and Huber, Lechner, and Strittmatter (2018).

Figure 6 summarizes causal pathways under conditional exogeneity with arrows representing causal effects. Each of T (treatment), M (mediators), and Y (outcome) might be causally affected by distinct and statistically independent sets of unobservables, but none of these unobservables may jointly affect two or all three elements (T, M, Y) conditional on X.

Figure 6: Causal paths under conditional exogeneity given pre-treatment covariates (Farbmacher et al. 2019)



An important (implicit) assumption underlying these studies is that the covariates to be controlled for can unambiguously be preselected by the researcher. As Farbmacher et al. (2019) point out, this assumes away any uncertainty regarding the selection of covariates to be included in the model and entails incorrect inference under the common practice of choosing (or refining the choice of) covariates based on their predictive power. **To overcome this limitation**, Farbmacher et al. (2019) combine causal mediation analysis based on efficient score functions (Tchetgen Tchetgen and Shpitser, 2012) with the **Double/Debiased Machine Learning (DML)** of Chernozhukov et al. (2018) for a data-driven method of controlling for observed confounders to obtain valid inference under specific regularity conditions. This permits to relax the assumption that that confounders (X) need not be known a priori, and the set of potential confounders can be even larger than the sample size. This is particularly useful in high dimensional data settings with a vast number of covariates, as in our case, that could potentially serve as control variables, which can render researcher-based covariate selection complicated, if not infeasible.

Farbmacher et al. (2019)'s identification strategy is based on the assumption that confounding of the treatment-outcome, treatment-mediator, and mediator-outcome relations can be controlled for by conditioning on observed covariates, X.

Consistent with a large part of the mediation literature, Farbmacher et al. (2019) invoke **sequential conditional independence assumptions** with respect to the treatment and the mediator for identification. Formally, their identification of the causal mediation parameters hinges on the following assumptions:

- ***Assumption 1: conditional independence of the treatment (T) and potential mediators (M) or outcomes given X.*** ***In experimental settings, such as ours, this assumption is easily satisfied in the absence of attrition.*** In non-experimental data, the plausibility of this assumption hinges critically on the richness of X. *Notice that in our case, SEED and Educate! are both RCTs and we have also a rich set of baseline characteristics.*

- ***Assumption 2: conditional independence of the mediator.*** This rules out confounders that jointly affect the mediator and the outcome conditional on T and X. If X is measured exclusively pre-treatment (as is common to avoid controlling for variables potentially affected by the treatment), this implies the absence of post-treatment confounders of the mediator-outcome relation. This assumption is more delicate and is what confers the "**sequential conditional independence assumption**" its characterization as a strong assumption.
    - It is worth noting that this assumption is more easily threatened if M and Y are measured at different points in time and a long period of time has elapsed between their measurements. ***In our RCTs, M and Y are measured contemporaneously***. *Nonetheless this remains a strong assumption.*
- ***Assumption 3: Common support assumption*** *that is* Pr(*T=treatment status* | M=m*,* X=x)>0 such that treatment status is not predicted with probability 1, conditional on M and X.

We are particularly interested in how these programs shape skills and are differentially rewarded in the labor market, as well as in subsequent social spillovers they may engender (e.g., in terms of risky behaviors, fertility, partnership quality, and IPV), which may highlight possible trade-offs between strictly economic returns vs. socially desirable outcomes.

Figure 7: Richness of our Data – Summary.

Baseline Characteristics (X)	Mediatiors (M)	Outcomes (Y)
Family Background - SES - Wealth - Business ownership - Youth (age, gender) IQ measures - Memory & Raven	Self-Reported Soft Skills - Big 5 - Grit - Stress - Prosocial attitudes/behaviors - Self-Efficacy	Educational Investments (E!) - Years of schooling - Business & STEM majors
Soft Skills - Anxiety - Confidence - Risk Aversion - Creativity - Empowerment - Leadership - Locus of control - Patience - Prosocial attitudes/behaviors	Task-Based Soft skills - Creativity - Negotiation - Persuasion	Economic Outcomes - Labor force participation - Business creation/performance - Earnings
	Hard-Skills - Business knowledge - Win-Win opportunities - Identifying business ideas	Social Spillovers - Family formation - Risky behaviors - Gender norms - IPV

Causal mediation analysis will permit us to understand which skills, as they are impacted by our interventions, are responsible for shaping school investments, labor market outcomes, and social spillovers. We will be able to gain insight into which dimensions of skill upgrading are responsible for which outcomes and whether different skills matter for different outcomes.

Beyond sequential ignorability (SI) assumption. At the core of causal mediation analysis is the problem of how to identify the causal effect of mediators (i.e., intermediate outcomes) on final economic outcomes. Traditionally, the mediation literature rests on a sequential ignorability (SI) assumption; that is, sequential conditional independence assumptions with respect to the treatment and the mediator and their respective counterfactual outcomes. A substantial portion of the recent literature on mediation stems from the recognition that sequential ignorability represents a strong assumption and novel contributions have focused on estimation, sensitivity checks, and bounds on mediation effects under weaker conditions (for a recent review, see Huber, 2019).

We will also attempt to add to the causal mediation analysis literature in the context of ML methods, exploring two separate strategies.

- Strategy 1: we will maintain the sequential ignorability assumption and derive conditions under which it is possible to estimate the conditional average treatment effects as a function of mediators. That is, our objects of interest are the HTEs (or CATEs) for outcome Y, as function of mediators M for Y: HTE(Y| M, X).
- Strategy 2: Eventually, we will also attempt to relax the sequential ignorability assumption. We think the following steps hold some promise. Intuitively, our strategy proceeds in two steps. First, estimate HTEs (or CATEs) for intermediate outcomes as a function of baseline characteristics. The resulting HTE will by construction be a function of baseline characteristics and be a consistent estimator for HTE(Y|HTE(M|X)), where Y is the outcome of interest, X is a vector of baseline characteristics and M is a vector of intermediate outcomes denoting mediating factors. In our example, HTE(Y| HTE(M|X))=HTE(earnings |HTE(skills|X)).Therefore, in the second step, we are armed with a series of instruments (HTEs with respect to skills, M, affected by the interventions as a function of baseline characteristics) that will permit us to decompose the average treatment effect for the final outcome of interest (in our example, earnings or any other economic or social spillover outcomes) as a function of program-impacted skills (M). In summary, this will permit us to study the HTEs (or CATEs) for the final outcomes as a function of the HTEs (or CATEs) for the intermediate outcomes.

Innovation/conclusion

Whether soft skills are malleable in early adulthood and the degree to which training programs that aim to increase the stocks of these skills can generate improvements in productivity and life outcomes are questions that have only begun to be explored. Most of the existing literature only examines the short-run effects of business training. The few studies that have documented impacts over time often find that effects vanish in the longer term. To our knowledge, these two studies are the first to rigorously evaluate the long-term impacts of skill development programs for youth currently in secondary school or who have recently graduated. Our works features several contributions and innovations.

By applying regression trees and ML methods to study HTEs to inform the generalizability of the observed impacts and by identifying/studying (causal) mechanisms, these findings will contribute to the formulation of mid-level theories. Furthermore, they will offer direct guidance to policy makers as to program design and policy targeting while abstracting from program-specific features to identify which set of skills and/or program components are effective levers such that, if activated by other programs, they could move outcomes along the causal chain.

The proposed work also has the advantage of studying these mechanisms in the context of two distinct RCTs involving three curricula with different intensities of hard and soft skills, yielding concrete policy guidance on curriculum design for skills interventions in alternative contexts and on the generalizability of results to different populations.

The long-term nature of the two studies will permit us to study outcomes along long chains of causation. However, an additional strength of the proposed work is its reliance on a rich and novel set of instruments. These include self-reported and task-based measures of soft skills which will permit us to study impacts not only on traditional economic outcomes, but also spillovers such as on social norms concerning gender roles, couples' decision making, and intimate partner violence.

Finally, in line with recent trends increasingly moving towards greater research transparency, we plan on making data and replication code publicly available once our work is submitted for publication. Sharing research materials beyond published articles is valuable both because it allows others to examine the full body of evidence underlying reported results, and because it facilitates fuller re-use of collected data.

References – To be Completed

- 3ie (2015) "The State of Evidence on the Impact of Transferable Skills Programming on Youth in Low- and Middle-Income Countries." *3ie Scoping Paper 4*, New Delhi: International Initiative for Impact Evaluation (3ie).
- Abramsky T., Devries K., Kiss L., Nakuti J., Kyegombe N., Starmann E., Cundill B., Kaye D., Musuya T., Michau L., and Watts C. (2014). "Findings from the SASA! Study: a cluster randomized controlled trial to assess the impact of a community mobilization intervention to prevent violence against women and reduce HIV risk in Kampala, Uganda." *BMC Medicine*, 12(122): 1-17.
- Abramsky T., Devries K.M., Michau L., Nakuti J., Musuya T., Kyegombe N., and Watts C. (2016). "The impact of SASA!, a community mobilization intervention, on women's experiences of intimate partner violence: secondary findings from a cluster randomized trial in Kampala, Uganda." *Journal of Epidemiology and Community Health*, 70(8): 818–825.
- Athey S. and G. Imbens (2016) "Recursive partitioning for heterogeneous causal effects." *Proceedings of the National Academy of Sciences*, 113 (27).
- Athey S., J. Tibshirani, and S. Wager (2019) "Generalized random forests". *The Annals of Statistics*, 47(2):1148–1178, 2019.
- Blattman C., N. Fiala, and S. Martinez (2018) "The Long-Term Impacts of Grants on Poverty: 9-year Evidence From Uganda's Youth Opportunities Program." *NBER WP 24999*.
- Blattman C., S. Dercon, and S. Franklin (2019) "Impacts of Industrial and Entrepreneurial Jobs on Youth: 5-Year Experimental Evidence on Factory Job Offers and Cash Grants In Ethiopia." *NBER W P15788*.
- Bloom N., A. Mahajan, D. McKenzie, J. Roberts (2018) "Do Management Interventions Last? Evidence from India" *NBER WP24249*.
- Bloom DE., M. McKenna, and K. Prettner (2018a) "Demography, Unemployment, Automation, and Digitalization: Implications for the Creation of (Decent) Jobs, 2010-2030" *NBER WP 24835*.
- Bouguen, Huang, Kremer, Miguel (2018), "Using RCTs to Estimate Long-Run Impacts in Development Economics" *NBER WP 25356.*
- Cameron A, Miller D. (2015), "A practitioner's guide to cluster robust inference" *Journal of Human Resources* 50: 317–372.
- Campos F., M. Frese, M. Goldstein, L. Iacovone, H.C. Johnson, D. McKenzie, and M. Mensmann (2017) "Teaching personal initiative beats traditional training in boosting small business in West Africa." *Science*, 357(6357).
- Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018): "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, 21, C1–C68.
- Conti, G., J. J. Heckman, and R. Pinto (2016): "The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviour," *The Economic Journal*, 126, F28–F65.
- Della Vigna S. (2009) "Psychology and Economics: Evidence from the Field." *J. of Economic Literature*, 47(2).
- Deming D. (2017) "The Growing Importance of Social Skills in the Labor Market." *The Quarterly Journal of Economics*, 132(4).

- Egana Del Sol P. (2018) “Emotions, Education, and Entrepreneurship,” *Working Paper*.
- Farbmacher H., M. Huber, L. Laffers, H. Langen, M. Spindler (2020) “Causal mediation analysis with double machine learning” *Working Paper*.
- Glennester R., C. Walsh, L. Diaz-Martin, (2018) “A Practical Guide to Measuring Women’s and Girls’ Empowerment in Impact Evaluations,” JPAL.
- Groh et al. (2016), “The impact of soft skills training on female youth employment: evidence from an RCT in Jordan,” *IZA Journal of Labor & Development*.
- Haushofer, J. and E. Fehr (2014) “On the Psychology of Poverty.” *Science*, 344(186).
- Hotz, J., Imbens, G., Mortimer, J., (2005) “Predicting the efficacy of future training programs using past experiences at other locations”. *Journal of Econometrics*, 125, 241-270.
- Huber, M. (2014): “Identifying causal mechanisms (primarily) based on inverse probability weighting,” *Journal of Applied Econometrics*, 29, 920–943.
- Huber, M. (2015): “Causal pitfalls in the decomposition of wage gaps,” *Journal of Business and Economic Statistics*, 33, 179–191.
- Huber, M. (2019) “A review of causal mediation analysis for assessing direct and indirect treatment effects,” *Working Paper*.
- Huber, M., M. Lechner, and G. Mellace (2017): “Why Do Tougher Caseworkers Increase Employment? The Role of Program Assignment as a Causal Mechanism,” *The Review of Economics and Statistics*, 99, 180–183.
- Huber, M., M. Lechner, and A. Strittmatter (2018): “Direct and indirect effects of training vouchers for the unemployed,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181, 441–463.
- Masset E. and H. White, (2019) “To Boldly Go Where No Evaluation Has Gone Before: The CEDIL Evaluation Agenda,” *CEDIL White Paper*.
- Manpower (2010), “Supply/Demand: 2010 Talent Shortage Survey Results,” Manpower Group: Milwaukee, WI.
- McKenzie, D. (2020) “Small business training to improve management practices in developing countries: Reassessing the evidence for “training doesn’t work",” Oxford Review of Economic Policy, forthcoming.
- McKenzie D. and Woodruff C. (2017) "Business practices in small firms in developing countries." *Management Science*, 63(9).
- OECD (2015) “Skills for Social Progress: The Power of Social and Emotional Skills,” *OECD Skills Studies*: OECD Publishing.
- Shadish, W., Cook, T., Campbell, D., 2002. Experimental and Quasi-experimental Designs for Generalized Causal Inference. Houghton Mifflin.
- Su, X., Chih-Ling Tsai, H. Wang, D. M. Nickerson, and B. Li (2009) “Subgroup Analysis via Recursive Partitioning.” *Journal of Machine Learning Research*, 10:141–58.
- Tchetgen Tchetgen, E. J., and I. Shpitser (2012): “Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness, and sensitivity analysis,” *The Annals of Statistics*, 40, 1816–1845.
- Ubfa Dl., Arraiz, D. Beuermann, M. Frese, A. Maffioli, “The Impact of Soft-Skills Training for Entrepreneurs in Jamaica,” *IGIER WP 645*.

- Wager S and Athey S. (2018) "Estimation and inference of heterogeneous treatment effects using random forests." *Journal of the American Statistical Association*, 113(523).
- Zeileis A., T. Hothorn, and K. Hornik. (2008) “Model-Based Recursive Partitioning.” *Journal of Computational and Graphical Statistics*, 17(2): 492–514.

Centre of
Excellence
for Development
Impact and Learning

www.CEDILprogramme.org