

IGL Working Paper Series

Impact Measurement based on Repeated Randomized Control Trials: The Case of a Training Program to Encourage Social Entrepreneurship

Thomas Åstebro and Florian Hoos

January 2021

IGL Working Paper No. 21/01

The IGL Working Paper Series is intended to make available early results of research undertaken or supported by Nesta, its partners, or members of the IGL Research Network in order to elicit comments and suggestions for revisions and to encourage discussion and further debate prior to publication. © Year 2021 by the author(s). Short sections of text, tables and figures may be reproduced without explicit permission provided that full credit is given to the source. The views expressed in this working paper are those of the author(s) and do not necessarily represent those of Nesta.

Impact Measurement based on Repeated Randomized Control Trials: The Case of a Training Program to Encourage Social Entrepreneurship

Thomas Åstebro
HEC Paris
astebro@hec.fr

Florian Hoos
HEC Paris
hoos@hec.fr

ABSTRACT

Designing effective entrepreneurship training programs is still a challenge despite the investments in training made by governments and private institutions, and its importance for economic growth. We report a case of impact measurement of a social entrepreneurship program based on repeated randomized controlled trials (RCTs), discuss challenges of conducting repeated RCTs, and implications for policy evaluation. Impact measures from the first edition of the program showed no detectable treatment effects. The second edition was adjusted by reducing leadership training and increasing traditional entrepreneurial skills training, and had strong treatment effects on entrepreneurial activities, the creation of a new venture during the program, and subsequent start-up activity. Employing sequential field experiments can improve entrepreneurship training programs despite the challenges of executing RCTs in the field.

Keywords: social entrepreneurship; field experiment; policy evaluation; randomized controlled trial; training.

1. INTRODUCTION

Entrepreneurs are widely seen as important in creating economic growth and employment. Social entrepreneurship is a recent and rapidly growing form of entrepreneurship that is widely recognized in the media and by policy makers (e.g., Economist, 2006; EC, 2014; Forbes, 2014). Social entrepreneurship is characterized by a deep commitment to a social cause, and the objective to develop new business models based on the coexistence of economic, social, and ecological impacts (Miller & Wesley, 2010; Nga & Shamuganathan, 2010).¹ Hence, social entrepreneurs primarily seek to solve societal problems through innovative entrepreneurial activities (Austin et al., 2006).² As reflected in the sustainable development goals of the United Nations,³ numerous politicians and business leaders have called for social entrepreneurs to step forward (e.g., Miller & Wesley, 2010; Nga & Shamuganathan, 2010; Parrish, 2010; Yunus, 2010). Providing adequate training for future social entrepreneurs has therefore become an important challenge (Litzky et al., 2010; Nga & Shamuganathan, 2010; Tracey & Phillips, 2007). Social entrepreneurship training includes training not only in entrepreneurial skills but also social leadership skills and social entrepreneurial identity development (e.g., Litzky et al., 2010; Smith & Woodworth, 2012). This multidimensional training approach takes into account

¹ Corner and Ho (2010) identify social entrepreneurship (SE) as an innovative type of entrepreneurship: “scholars suggest that opportunities for SE are likely to be distinct from opportunities in the commercial sector and need to be examined in their own right.”

² Social entrepreneurs focus most often on market-oriented approaches to implement social change and seek to generate revenues to finance their activities (Mair et al., 2012). In this paper, the researchers involved made no judgment as to which types of projects were included, and the organization that ran the program was flexible in admitting various types of social entrepreneurial undertakings.

³ See <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>.

the duality of social ventures with both a social and a business identity (Short et al., 2009).

A growing number of educational institutions, incubators, and accelerators have tried to develop social entrepreneurship training programs.⁴ However, our collective knowledge on what constitutes *effective* social entrepreneurship training remains limited and is often based on qualitative accounts (e.g., Kwong et al., 2012; Miller et al., 2012; Smith et al., 2012), even though “identifying a variety of effective pedagogical approaches will become increasingly important for business education in particular and society in general” (Smith & Woodworth, 2012, p. 390). Even in the general field of entrepreneurship, relatively little is known about the effectiveness of entrepreneurship training (see Section 2). Despite the many efforts to foster entrepreneurship,⁵ it is apparently a difficult skill to teach.

We describe a social entrepreneurship training program and measure its impact in two sequential randomized control trials (RCTs). We discuss the challenges and limitations of conducting repeated RCTs in a collaboration between academics and practice and how it affects policy evaluation. We conducted our research jointly with a French social enterprise that offered a training program with the goal of encouraging youths to become social entrepreneurship leaders in France. The program cost of approximately 12,000 euros (\$13,100)

⁴ See, e.g., the NYU Wagner School’s Social Entrepreneurship undergraduate minor, the Duke Fuqua School of Business’s MBA concentration in social entrepreneurship, and the Stanford Graduate School programs in social entrepreneurship. See also the Skoll Foundation’s website on social entrepreneurship training programs: <http://archive.skoll.org/2011/01/18/training-the-next-generation-of-social-entrepreneurs/>.

⁵ Governments and donors around the world probably spend billions of dollars on entrepreneurship training programs (Fairlie et al., 2015). For example, according to Lyons and Zhang (2017), in 2012 the US Small Business Administration alone spent \$127 million on training entrepreneurs. And McKenzie (*forthcoming*) estimates that at least \$1 billion is spent annually training at least 4 million to 5 million potential and existing entrepreneurs in developing countries.

per participant was largely paid with government and philanthropic funding. The program duration was 12 days, with a six-month period afterward of individualized follow-up coaching. The training consisted of two major components: accelerating social entrepreneurial activity and promoting leadership skills associated with a social entrepreneurial identity.

To our knowledge, this study is the first to report the use of repeated RCTs to measure the impact of a social entrepreneurship training program. We show that RCT-based impact measurement can be a way for researchers to use scientific methods for improving entrepreneurial training in a collaboration between academics and practice. To this end, using survey data, we measure improvement in entrepreneurial and social entrepreneurial activity, as well as progress in the participants' social leadership skills, social career intentions, and sustainable behavior. Most important, we track career development based on LinkedIn data, which suffers less from the execution problems that come with collecting survey data.

In evaluating the impact of the first round of the program, we found no detectable treatment effects. Based on what we learned in the first evaluation, the program was redesigned in the second round. Decisions on the redesign were made by a French social enterprise through a process that was constructive, but less than ideal from a scientific perspective. The main change was an increase in analytical entrepreneurial skills training and a decrease in leadership skills training and, at the same time, intensification of the individualized coaching. In the second RCT, the redesigned program had a large treatment effect on entrepreneurial activities and the creation of a new venture during the program, as well as on subsequent entrepreneurial career development. However, because both changes were made simultaneously, we could not disentangle the contribution of the training component from that of the coaching. Nevertheless, the program still had no detectable impact on either leadership skills or social entrepreneurial identity measures.

This case study has implications for the use of RCTs in conducting academic research and

for social entrepreneurial practice. First, it highlights the advantages as well as the challenges and limitations of impact measurement when working with practitioners. Recent literature on impact measurement identifies that “perhaps the most sophisticated approaches are experimental and quasi-experimental research designs, such as randomized control trials (RCTs)” (Kroeger & Weber, 2014, p. 517). But RCTs also have many practical execution challenges. Nevertheless, although practitioners may sometimes have different goals than researchers, RCTs performed in the field can still produce substantial lessons for both practice and science.

Second, our field-experiment evidence offers some insights for firms designing internal training programs for stimulating innovative and social entrepreneurial behavior, as well as incubators, accelerators, and other organizations wishing to impart social entrepreneurial skills and identity to their participants. With some methodological caveats noted, our results indicate that teaching analytical skills, in particular business model design and the lean-startup method, and affecting action-oriented behavior appear to matter more than teaching social leadership skills or targeting identity-related social entrepreneurial dimensions, such as pro-social behavior or social career intentions among those interested in social entrepreneurship. If larger-scale research confirms our findings, then our study has implications for policy making, including funding decisions for (social) entrepreneurial training.

2. LITERATURE REVIEW

This literature review clarifies the differences between social and traditional entrepreneurship and informs the choices of impact measures in the field experiment. For the latter, we draw mainly on traditional entrepreneurship training policy evaluations. We also develop some new impact measures in social entrepreneurship and leadership.

Researchers largely agree that what distinguishes social entrepreneurship from traditional entrepreneurship is the focus on a social mission (e.g., Dacin et al., 2010; Dees et al., 2001;

Peredo & McLean, 2006; Smith & Woodworth, 2012). Social entrepreneurs are said to hold a strong belief in individual responsibility and the capacity to contribute meaningfully to the creation of social value and common welfare (Dees et al., 2001; Litzky et al., 2010; Nga & Shamuganathan, 2010). As a result, social entrepreneurship is characterized by a duality between a business identity and a social identity. Referring to this duality, Short et al. (2009, p. 184) stress the “nonfinancial aspects that meet the affective needs of venture members, such as identity, ability to exercise influence for positive social change, and the perpetuation of the venture’s mission.” As a result, one of the main challenges in designing effective social entrepreneurship training programs is to address this duality and to find the right balance between developing entrepreneurial skills and a socially oriented identity. The latter could be expressed, for example, through taking social leadership and making pro-social choices.

The social entrepreneurship literature stresses social leadership skills as necessary for success as a social entrepreneur and the specific socially oriented identity of the people considered social entrepreneurs (e.g., Short et al., 2009; Smith & Woodworth, 2012). We view the social entrepreneurship training studied in this paper as having an identity-based approach (cf. Smith & Woodworth, 2012). Identity-based approaches to social entrepreneurship training aim to develop desires in trainees that are consistent with a social entrepreneur’s identity. Granted, many characteristics of social and commercial entrepreneurs are similar, such as the ability to find and exploit opportunities and a customer-centric approach to doing business (Bacq et al., 2016). However, we join previous literature in stating that social entrepreneurship training typically offers “an opportunity to begin identifying with social entrepreneurs and innovators, and, consequently, develop desires consistent with that social identity” (Smith & Woodworth, 2012, p. 391; see also Howorth et al., 2012). Among the dimensions that shape a socially oriented identity are altruistic behavior (Tan et al., 2005), a passion for social impact (Plaskoff, 2012), a caring attitude (André & Pache, 2016), and a specific transformational

leadership style (Litzki et al., 2010). The last of these has been used as a pedagogical dimension in social entrepreneurial learning contexts, so that students believe that they can become social entrepreneurs (Litzki et al., 2010) and take action accordingly (e.g., Howorth et al., 2012; Smith & Woodworth, 2012). As Smith and Woodworth (2012) put it, “course instruction can be a catalyst that channels and enhances students’ desires to make a difference in the world—helping them identify with the social entrepreneurship community” (p. 391). We measure three dimensions of a social entrepreneur’s identity: social leadership skills, sustainable behavior, and social career intentions.

At the same time, previous literature on social entrepreneurship stresses that many business challenges inherent in social entrepreneurship can be addressed through the acquisition of skills. These challenges are, among others, overcoming resource constraints specific to social entrepreneurs (Desa & Basu, 2013), the acquisition of traditional entrepreneurial knowledge that is adequate for social entrepreneurial contexts (Ko & Liu, 2015), the skills necessary to scale a social venture (Smith et al., 2016), the challenge of choosing and exploiting the right ecosystem (Thompson et al., 2018), and the challenge of managing a double-bottom or triple-bottom line, in which the social entrepreneur must balance the economic, social, and environmental impacts of a social venture (Desa & Basu, 2013). Sometimes, these entrepreneurial challenges are connected and together represent the business challenges inherent in social entrepreneurship. For instance, Desa and Basu (2013, p. 26) argue that “the tensions between social mission and financial return can create resource mobilization challenges.” We call the necessary social entrepreneurial skill development to engage in social entrepreneurial activity “entrepreneurial and social entrepreneurial activity”. In sum, previous literature suggests that, to be effective, social entrepreneurship training take into account the duality between teaching traditional entrepreneurial skills and a socially oriented identity.

Even though some studies on social entrepreneurship make concrete training propositions

and evaluate some aspects of the social entrepreneurship learning experience, they are largely based on qualitative accounts (e.g., Kwong et al., 2012; Miller et al., 2012; Smith et al., 2012) and often offer “how to” guides (e.g., Litzki et al. 2010; Zietsma & Tuck, 2012), rather than examining the causal effects (impacts) of different elements of training.

Because of the lack of experimental studies on social entrepreneurship training, we rely on experiences in experimental studies on traditional entrepreneurship training that occurs around the globe. The training appears to differ between Western and emerging/developing countries. Programs in the former typically offer a potential career path to those interested in becoming entrepreneurs or try to promote entrepreneurial skills and intentions in general, often among young people. In contrast, in the latter, programs are intended most often to increase the level of self-employment among poor people, often coupled with financial assistance. The factors that lead to positive outcomes of entrepreneurship training in developing and emerging countries are better understood (see, e.g., Anderson et al., 2018; Banerjee et al., 2015; Campos et al., 2017; Cho & Honorati, 2014; Gertler, 2017; McKenzie, *forthcoming*; McKenzie & Woodruff, 2014; van der Sluis et al., 2005).⁶

As the program we evaluate is performed in a Western country, from this point on, we review only program evaluations in the West. They are quite limited in number and focus on entrepreneurship training of various kinds at schools (e.g., Elert et al., 2015; Oosterbeek et al., 2010; Peterman & Kennedy, 2003; Rosendahl-Huber et al., 2014; Souitaris et al., 2007; von Graevenitz et al., 2010), with evaluation of only some entrepreneurship training programs

⁶ Briefly, short programs have a limited impact, whereas intensive programs appear to have more substantial effects. Training appears to be more effective when it is presented in simple ways, such as by rule of thumb. Complementing in-class teaching with personalized follow-up visits or mentoring appear to increase or have positive effects where there is none otherwise (Lafortune et al., 2018; McKenzie, *forthcoming*).

outside schools (e.g., Fairlie et al., 2015).⁷ The results are decidedly mixed. Most studies examine the effect on a limited set of outcomes in the near future, such as the intention to become an entrepreneur from coursework, or being mentored in an entrepreneurial project while in school. For example, the Junior Achievement Company Program consists of five to eight months of weekly coaching sessions of one to two teenage students by a seasoned entrepreneur with additional support from a teacher. This program has been evaluated in Sweden, the Netherlands, and Australia, with inconsistent results (for details, see Appendix A; Elert et al., 2015; Oosterbeek et al., 2008; Peterman & Kennedy, 2003). An additional study worth mentioning is Lyons and Zhang (2017), who, in a non-experimental setting, examine the effect of a Canadian university-based “high-tech” entrepreneurship training program for undergraduates. This program appears similar in scope and in the number of participants to the French program that we examine, although it focuses on “high-tech” entrepreneurship, instead of social entrepreneurship. Lyons and Zhang observe non-uniform positive effects on subsequent entrepreneurial career paths, with lower future entrepreneurial activity among applicants with prior entrepreneurship experience. The causal identification, unfortunately, is not clear.

It is unclear whether these programs should be expected to have any large treatment effects or, when detected, whether they persist. On the one hand, it could be argued that teaching entrepreneurship is a matter of passing on some simple tricks of the trade, such as how to write a business plan (a very popular subject) and how to present oneself well in a venture pitch competition (also popular). However, except as part of earning a degree in business, these types of courses might do little to prepare a person for entrepreneurship. On the other hand, teaching

⁷ Appendix A summarizes studies in Western countries that are relevant to our research along the dimensions in terms of length of program exposure, intensity, subject matter, and type of outcome measurement.

entrepreneurship in the field might be very important if it has a significant positive effect on the economy, is difficult to learn, requires extensive training, and includes entrepreneurial character development, such as leadership, as key to its success. To benchmark expectations, McKenzie (*forthcoming*) studies in-field instructor-led business training programs of between three and 12 days in developing countries and reports an average treatment effect on profits and sales of between 5 and 10 percent.

A few studies expand the scope of outcomes to examine changes in various noncognitive skills (e.g., Oosterbeek et al., 2010; Rosendahl-Huber et al., 2014). The extent to which the program content or program type (formal academic and school education vs. training) influences the outcomes of entrepreneurship training is not yet well understood (Betcherman et al., 2007; Martin et al., 2013).⁸ And although Heckman and Kautz (2013) forcefully argue that noncognitive character skills, such as agreeableness and conscientiousness, are very important for labor market and other life outcomes,⁹ it is not clear which skills predict entrepreneurial success or whether they can be affected by a short-term training program or course, such as those that have been evaluated so far.¹⁰ For example, widely different results

⁸ Martin et al. (2013) find some evidence in their meta-analysis of entrepreneurship education and training that “academic-focused” entrepreneurship education leads to better entrepreneurial outcomes than “training-focused” interventions, whereas the meta-analysis by Betcherman et al. (2007) supports the conclusion that no major differences are found across different categories of entrepreneurship interventions in terms of impact or cost effectiveness. These reviews include few studies containing causal identification.

⁹ For example, Heckman and Kautz (2013) cite meta-analyses showing that measures of character skills rival IQ and measures of socioeconomic status in predicting longevity.

¹⁰ Martin et al. (2013) conducted a meta-analytic review of the outcomes of entrepreneurship education and training (EET) and point out that “the EET literature includes many studies that do not meet a high standard of rigor, and there is no clear indication of a trend toward increased methodological rigor at this time. Further, our results

are observed in the studies by Oosterbeek et al. (2010) and Rosendahl-Huber et al. (2014). Oosterbeek et al. (2010) find no treatment effects on 11 noncognitive skills and a significant reduction in entrepreneurial intentions, while Rosendahl-Huber et al. (2014) find small but statistically significant positive treatment effects on seven out of ten similar skills, as well as a negative significant treatment effect on entrepreneurial intentions.¹¹

Other types of entrepreneurship training programs, such as accelerators, may also have some similar training components (see, e.g., Gonzalez-Uribe & Leatherbee, 2014; Hallen et al., 2017; Yu, 2019), which would be interesting to examine in an RCT. Accelerators run 30- to 90-day programs that offer seed funding, mentoring, training, coaching, networking, peer interaction, and coworking space. However, they have a different design from the more general Western training programs discussed, in particular the French program, in at least three dimensions. Although the ages of participants are about the same across various accelerators and the French program, the French program focuses on social entrepreneurship projects and on leadership training. Further, we were able to perform random allocation to treatment.

suggest that the lower rigor studies tend to overestimate the impact of EET. Many studies do not incorporate both pre- and post-EET intervention measures and treatment and control group comparisons. To improve the value of the EET literature in the future, studies should be designed to include both of these elements (ideally at several points in time post-intervention). Such methodological rigor will greatly improve the ability of researchers to make accurate claims about the impact of EET on entrepreneurship related outcomes” (p. 212).

¹¹ The quite common negative effects of training on entrepreneurial intentions for Western-based students (Oosterbeek et al. 2010; Rosendahl-Huber et al., 2014; von Graevenitz et al., 2010) might be attributable to the fact that the courses are mandatory. Participants whose attendance is compulsory but lack a strong interest in entrepreneurship might become even less interested after taking such a course, as entrepreneurship is a risky and difficult career choice. Where the sample is first prescreened, as in Anderson et al. (2018), Lyons and Zhang (2017), and in this study, the effect of the treatment on intentions might be different.

However, with the clear exception of Chatterji et al. (2019), all the studies on accelerators that we know of measure an all-or-nothing bag of treatments on projects nonrandomly allocated to treatment, making it very difficult to pick out the treatment effect of any specific training component.

In summary, past research shows mixed, null or negative treatment effects of entrepreneurial training programs on skills, entrepreneurial activities, and future career choices in the West. No treatment results exist for social entrepreneurial training programs. It is argued that social entrepreneurship training must address the development of a socially oriented identity, including social leadership skills, in addition to developing traditional entrepreneurial skills. The literature strongly influenced our discussions with the French social enterprise on the training components, informed the design of our field-experiment study, and our choice of impact measures.

3. THE TRAINING PROGRAM AND THE CASE OF A COLLABORATION BETWEEN ACADEMIA AND PRACTITIONERS

3.1. The 2014 program: Organization, program design, and application process

The program was first offered by the organization in its founding year, 2014. The organization has since won several awards (including the 2015 Google Impact Challenge), has been promoted widely, and has been mentioned by French politicians as a flagship program for boosting social entrepreneurship in France. The mission of the organization is to activate the potential of talented young people and empower them to change society in a positive direction through entrepreneurship.

Inspired by the Indian program Jagriti Yatra (<http://www.jagritiyatra.com>), the organization wanted to offer an intensive social entrepreneurship program with a strong leadership skills training dimension and a social entrepreneurial activity component. The central portion of the program was a 12-day bus trip with stops in different French cities. The

50 participants were coached and met experts in leadership, social entrepreneurship, and related topics. The bus trip was followed by a six-month coaching program. The main focus of the coaching period was to provide networking opportunities, coaching, and other support for developing the ventures. The organizers received support for the program design and execution from a large and dedicated group of advisors with either senior business experience or teaching experience at business schools. In particular, three professional coaches were in charge of the program content and were facilitators of events during the bus trip. For program details, see Appendix B1.

We offered to conduct a rigorous scientific evaluation of the impact of the training program for the organization, for which randomization to treatment is required. After negotiations, the organization agreed to implement stratified random allocation of qualified applicants into treatment and control groups. At this point, the organization was excited about obtaining a rigorous scientific impact evaluation as it was convinced of the value of the program.

The first round of the program was held as follows. First, a call for applications was posted on the organization's website, describing the ideal participant as follows: "We are looking for 50 budding entrepreneurs motivated to put their talents in the service of society and to bring change through entrepreneurship and social innovation. Regardless of your social origin, schooling, diploma (or lack thereof), hobbies or skills, we are looking for young people who are extremely motivated, enthusiastic, optimistic, and eager to commit."¹²

The organization aims to attract French youth from various backgrounds, in order to represent them fairly based on their socioeconomic background, gender, and education. The

¹² Translation by the authors from the original 2014 documents in French. All future descriptions are similarly translated by the authors. The 2015 documents were largely unchanged from those in 2014.

major criterion for participation was a motivation to contribute to a better and more sustainable society. To attract participants, the organization reached out via social networks (Facebook, Twitter, alumni networks of universities), e-mail lists from partner organizations (e.g., Ashoka, Make Sense), in-person presentations at French institutions of higher education, and presentations at numerous public events on social entrepreneurship, sustainability, and related topics.

The pedagogical concept of the program is based on three pillars: *inspiration*, *introspection*, and *taking action*. The inspiration phase informed participants about social entrepreneurship and the problems it faces and allowed them to meet famous social entrepreneurs and to gain a sense of pressing social problems in French society. In the introspection phase, participants worked on their profiles, learned about their personal strengths, and identified a social issue that is important to them. Participants were asked to develop a social entrepreneurial idea or personal desire to contribute to a fairer and more sustainable society. In the phase for taking action, participants worked on developing their projects. At a public event in Paris, which was also the last day of the bus trip, the participants presented their projects.

The total program budget in 2014 was 640,000 euros (approx. \$700,000 at the exchange rate on February 11, 2020, of which 275,000 euros were in-kind donations. The organization estimated the cost of the training program at about 12,000 euros (\$13,100) per person.¹³

Between March 17 and April 30, 2014, the organization received 397 applications. Each applicant answered open-ended questions using a web survey (see Appendix C). Applicants

¹³ The typically evaluated three- to 12-day business training program in developing countries costs between \$21 and \$740 per participant (McKenzie, *forthcoming*). A comparable training program in the United States costs between \$850 and \$1,300 per participant (Fairlie et al., 2015).

were informed which four characteristics were the most important for selection: the capacity to dream, a willingness to change society, leadership potential, and communication ability. Their free-form answers were independently evaluated by as many as three judges, who had received instructions from the organization on how to rate applicants on these four characteristics. The items defining the constructs for rating applicants and their weight are described in Appendix D.¹⁴ The weight of constructs and items was predetermined by the organizers. The organization (not the judges) calculated a “suitability” score using a linear additive weighting rule and ranked the applicants based on an average of the three judges’ evaluations.

After receiving a list with the top 100 applicants sorted by the suitability score, we randomly assigned them to treatment and control groups using stratified random sampling.¹⁵ At this point, the organization pushed back. The organizers were under pressure to run a successful program, and they argued that it would be unfair to randomly assign, for example, applicant 77 into the program, whereas applicant 76 might end up in the control group not receiving the training. The researchers reinforced the value of randomization for the ability to state whether the program was successful or not, and ultimately the organizers agreed to accept this design feature.

¹⁴ Based on our feedback, constructs were obtained from research on leadership, and items were taken from the International Personality Item Pool (IPIP) (<http://ipip.ori.org>).

¹⁵ In this approach, we partition the sample into K strata. In each stratum, the applicants are rank ordered based on their suitability score. In each stratum, half the top applicants in a stratum are randomized to be in the treatment group, and the other half are randomized to be in the control group. Appendix G, Table G1, indicates the sampling strata, the number of applicants, the number of individuals among the top 100 in each stratum, and the actual number in each stratum sampled. Stratified random sampling is intended to increase the efficiency of the experimental design. For details on the statistical properties of this approach, see Athey and Imbens (2016).

3.2. Measurements and design in 2014

We collected identical data before the program and after a six-month coaching period based on survey measures from both treated and control groups. We also manually collected LinkedIn data years after the completion of the program from both treated and control groups (described in more detail in section 4.4). Although we had to develop some measures on our own, the literature review is the basis for most of the measures used in the survey.

The program is intended to encourage young people to become leaders in social entrepreneurship and consists of two major training components: encouraging entrepreneurial activity and acquiring social entrepreneurial leadership skills that contribute to the development of a socially oriented identity. Consequently, we tried to match outcome measures to the goal of the program and its two major components.

First, we track their entrepreneurial activity. We ask whether a new business was started between $t = 0$ and $t = 1$: “Did you start working on a new business idea during the last few months (since June 2014 [2015], when the participants in the 2014 [2015] Tour were announced)?” and we code that as new business = 1 if participants answered Yes, and 0 if they answered No. We also ask how many out of the 22 entrepreneurial activities the participant engaged in during the same period (see Appendix E).¹⁶ We sum all affirmative answers and call this scale “traditional entrepreneurial activities.”

Importantly, we also track social leadership skills development. We use two standard leadership skill measures. First, we use the transformational leadership style (TLS) scale (Podsakoff et al., 1990) because it has been described as the natural leadership style of social

¹⁶ The measure is available from the Panel Study of Entrepreneurial Dynamics

(<http://www.psed.isr.umich.edu/psed/data/>) and asks yes/no questions for 22 different entrepreneurial activities.

See Appendix B.

entrepreneurs (Litzki et al., 2010). The three aspects of transformational leadership behavior (Podsakoff et al., 1990) are articulating a vision, providing an appropriate role model, and fostering the acceptance of group goals.¹⁷ The Cronbach's α for the three subscales is between 0.54 and 0.82 after treatment. Second, we use a scale concerning the motivation to lead (MTL) (Chan & Drasgow, 2001) because one of the main aspects of social entrepreneurship is the desire to foster social change through leadership. TLS, by contrast, reflects how people lead others (Podsakoff et al., 1990). MTL is a three-part measure of valence associated with the act of leading others, the person's noncalculative beliefs about the outcomes associated with success, and finally social norms related to the act of leadership.¹⁸ The Cronbach's α is between 0.71 and 0.77 for the three subscales of MTL after treatment.

To track their pro-social activity and identity, we capture social entrepreneurship skills acquisition and several identity measures. First, we ask the participants to answer yes or no as to whether they engaged in any of the five activities to become a social entrepreneur (see Appendix F). The program had various components concerning social entrepreneurial skills to engage in social entrepreneurial activity. For example, on one afternoon, the participants had "informal meetings with four different famous French social entrepreneurs," and an evening event included a "creativity workshop on innovative solutions for a better society facilitated by an external party through a gamification approach." As we did not find any prior studies that

¹⁷ In contrast to a transactional leadership style, in which leaders motivate subordinates by providing or withholding extrinsic rewards (MacKenzie et al., 2001); "transformational leaders get their followers to 'buy into' their visions and internalize them so that the followers become intrinsically motivated to strive for common goals and visions" (Goodwin et al., 2001, p. 772).

¹⁸ A sample item of affective identity is "I am the type of person who likes to be in charge of others"; a noncalculative belief is "I would agree to lead others even if there are no special rewards or benefits with that role"; and a social-normative motivation is "I have been taught that I should always volunteer to lead others if I can."

measured these activities, we develop this scale based on what typically constitutes social entrepreneurial activity (e.g., Dees et al., 2001; Kwong et al., 2012). We sum all affirmative answers and call this scale “social entrepreneurial activity.”

Second, the program also contained modules promoting the acquisition of a socially oriented identity, which should be manifested as sustainable behavior—for example, by working for a social or nonprofit organization or donating money to charitable organizations. To measure the impact of this promotion, we form a scale with 11 questions, such as “I have systematically recycled waste in my daily life,” add up the number of affirmative answers and call this measure “sustainable behavior.” See Appendix F. The Cronbach's α (a measure of scale reliability) for this scale is 0.65 for post-treatment values.

Similarly, researchers have suggested tracking students' career intentions as a measure of the more immediate impact of social entrepreneurship interventions on the development of their socially oriented identity (Kickul et al., 2012). Consequently, we construct four items representing social entrepreneurial intentions, such as an individual's desire to make a contribution to society and to serve a social mission in his or her job (Kickul et al., 2012; Kwong et al., 2012; Smith & Woodworth, 2012). We label this scale “social career intentions.” The Cronbach's α for this scale is 0.81. To contrast the development of social career orientations with general career intentions, we also construct a scale for traditional career intentions. Six items are sourced from established ideal employer studies and employer branding surveys (McKinsey, 2009; Universum, 2009) and include items such as the desire to have prestige, the importance of the competitiveness of the employer, and salary. The Cronbach's α for this scale is 0.77. For details, see Appendix F.

Measures were taken days before the applicants were informed about their selection (or exclusion) ($t = 0$) as well as six months after those who participated completed the bus tour ($t = 1$). For entrepreneurial and social entrepreneurial activity, the question at $t = 0$ refers to the

number of activities in which the participant engaged in the prior 12 months. The question at $t = 1$ refers to the number of activities in which the participant engaged between $t = 0$ and $t = 1$, a period of approximately eight to nine months, depending on precisely when the individual responded. Sustainable behavior was measured in a similar way.

We also collect data on standard demographic items in order to describe the participants and the control group. Questions are asked on age, gender, education, and various household and employment status characteristics. And as mentioned before, we collect data from LinkedIn in October 2017. In addition to manually retrieving data from LinkedIn on career outcomes, we also tried to contact the few participants who did not have a LinkedIn account by phone to obtain the same data.¹⁹ The data collection process for the 2014 cohort is described in Figure 1.

INSERT FIGURE 1 ABOUT HERE

3.3. The 2015 program: Changes of the program design and application process

As described in detail in Section 4, the results of the 2014 round were surprisingly modest on all measured dimensions. We saw no detectable improvements in social leadership skills, entrepreneurial activity, or social entrepreneurial identity development compared to the control group. This led to a redesign of the key program features. The organization hoped that, with the redesign, the program would have a stronger impact on engaging in social entrepreneurial activity, while still providing some basic training in leadership and social entrepreneurial identity. The organization revised the selection process, the program content during the bus trip, and the follow-up coaching for the 2015 program. Regarding selection, the organization

¹⁹ We collected these data in October 2017, approximately two years (25 months) after the 2015 bus tour and approximately three years (37 months) after the 2014 bus tour.

decided that, to increase the quality of the applicant pool, it would require applicants to have at least some previous entrepreneurial experience.²⁰ Further, because in the 2014 round the attrition rate in the coaching stage was high, the organization decided to select only applicants who committed to attending the follow-up coaching program.²¹ Regarding program content, the 12-day bus trip was redesigned to include more analytical entrepreneurship tools, in particular business model design and the lean startup method. Participants were also given more time to work on their projects. Leadership skill training, however, was significantly reduced. Finally, the follow-up coaching became more sophisticated, with a stronger focus on developing entrepreneurial skills and more pedagogically structured interactions with experts and peers to accelerate projects. The major changes are outlined in Table 1.

INSERT TABLE 1 ABOUT HERE

We had numerous discussions with the organizers about the unexpected lack of impact in the first round of the program and offered some recommendations based on our knowledge of the entrepreneurial training literature. However, the design changes implemented were determined largely by the organization. Our initial goal was to change either the 12-day training design or the following coaching period, but not both. However, the organization wanted to redesign both parts, and we therefore focused on defending our recommendation to implement random assignment to treatment again in order to allow identification of the causal impact of the program. Without random allocation to treatment, the results might show the effect of, for

²⁰ The applicants were obliged to indicate having a previous entrepreneurial orientation or experience in responses to four questions. For example, one question was: “Tell us about the achievement or initiative you are most proud of (a project you initiated/led or joined).”

²¹ Note that these two changes might change the composition of the sample but have no effect on the randomization to treatment.

example, the stricter selection criteria.

The organization received 200 applications for the 2015 round. The application procedure was similar to that in 2014. In particular, at the beginning of the online application the organizers explained that to receive consideration, applicants had to be 18- to 30-year-old French speakers, who “have a strong desire to change society through entrepreneurship and who want to get started based on [the organization’s] program.” As before, applicants were informed that four characteristics were most important for selection. Some slight modifications were made in the construction of the suitability score (see Appendix B2). The organization then established an ordinal ranking of the top 100 applicants based on the final suitability score.

We had another discussion with the organizers about randomization, but we were unable to persuade them to randomly assign all the top 100 applicants into treatment and control. The final result of our discussions was that the organization selected the best 25 of the top 100 applicants to be included in the program based on their suitability score. Because these applicants were not randomly allocated, we generally do not analyze them. Instead, the next 50 ranked applicants were jointly agreed to be randomly assigned to either treatment or control groups using the same stratified sampling approach as in 2014, and these are the applicants we use in the second RCT analyses. We call this group the “Mid50.”²² The organization was now confident that the best applicants were selected into the program.

4. METHODOLOGICAL CHALLENGES AND IMPACT MEASUREMENT RESULTS

4.1. Sample and nonrandom attrition checks

For the 2014 cohort, we obtained answers at both $t = 0$ and $t = 1$ from 38 members of the

²² Appendix G, Table G2, describes the sampling scheme in the 2015 round.

treatment group (76%) and 23 members of the control group (46%). For the 2015 cohort, we obtained responses from 26 members (100%) of the treatment group and 15 members (60%) of the control group, respectively.²³

The average age of the 2014 (2015) cohort is 26.6 (25.2), 47.5 percent (43.9%) are male, 21.3 percent (29.3%) had faced adversity of some kind in their life, 37.7 percent (51.2%) have a business education, 29.5 percent (14.6%) are currently students, 24.5 percent (19.5%) are employed full time, 29.5 percent (14.6%) have been self-employed at some point, 4.9 percent (5.0%) had started a business with employees before joining the program, and 9.8 percent (2.0%) have parents who earned more than 150,000 euros (\$164,000) in the year before the study. Other background statistics are reported in Appendix G, Table G3.

Because nonrandom response rates are possible, randomization and balance may have been compromised. In Appendix G, Table G3, we check whether randomization worked and whether balance is maintained. It shows the composition of the samples in the treatment and control groups on demographics and sampling variables at $t = 0$ for those responding both at $t = 0$ and $t = 1$. In the 2014 cohort, no statistically significant pre-treatment difference is seen between the treatment and control groups, whereas in the 2015 cohort the two groups show one significant difference: the participant's father had had different kinds of jobs. Because this is one significant result out of 26, it might be a false positive, which should occur at random in one out of 20 tests at $p < 0.05$. Nevertheless, we perform regressions in which we control for differences across participants in their likelihood of responding to both of the two surveys.

In addition, the two cohorts differ statistically on three variables: the 2015 cohort engaged in more social entrepreneurial activity than the 2014 participants, have stronger traditional

²³ By chance, 27 were allocated to the treatment group and 23 to the control group. One of the 27 in the treatment group did not answer at $t = 1$, and therefore we have 26 remaining observations in the treatment group.

career intentions, and are younger at the baseline (at $t = 0$). These differences are not jointly statistically significant, however $\chi^2(26, N=102) = 30.17, p = 0.26$. We conduct a statistical analysis on how these differences might affect a comparison of results across the two cohorts.

After detailing our estimation approaches, we report the survey-based treatment effects for the 2014 cohort, the 2015 cohort, and the 2017 career outcomes for both cohorts and, finally, we conduct a robustness analysis.

4.2. Analytical methods

Although balance among respondents appears to be maintained across the treatment and control groups for those who respond, the probability of response to the surveys, however, is correlated with treatment and with some covariates, as described in Appendix G, Table G4. These differences might also affect balance and thus both estimates and standard errors. Nonrandom attrition can be addressed in at least two ways (Athey & Imbens, 2016). First, we can adjust for covariate differences that determine responses in regressions. Second, we can weight responses with the inverse of the probability of response (and sampling probability), using a weighted least squares (WLS) regression. We apply both techniques, which requires individual-level analysis.

If we were not concerned with nonrandom attrition, the method that yields the most efficiently calculated treatment effects use the sampling stratum as the unit of analysis (Athey & Imbens, 2016), to compute a weighted mean across strata, which is then compared between the treatment and control groups using a simple t -test. Consequently, in Appendix G, Tables G7 and G8, we present results following Athey and Imbens's (2016) stratum-level analysis. In contrast, in Appendix G, Tables G5 and G6, we present regression results without making any adjustments for attrition (ordinary least squares [OLS]). We report these attrition-unadjusted results because adjusting for variables that have not been specified in advance is poor statistical practice and might introduce data-mining bias (Altman, 2005). However, taking the potential

effect of attrition seriously, we focus on results in which we adjust for both covariates and the probability of response and sampling using WLS. In these regressions, each observation is multiplied by a combined weight representing the inverse of the sampling proportion across strata computed from Appendix G, Tables G1 and G2, and by the inverse of the predicted response probability given covariates, derived from Appendix G, Table G4 (Holt et al., 1980). In the text, we note when the results are consistent across all three methods (adjusting for covariates and the probability of response with WLS, no adjustment using OLS, and stratum-level analysis) and call these results “robust.”

We regress y_{i1} on D_i , a dummy variable that has a value of one if the individual was in the treatment group, and zero if the individual was in the control group. In addition, the estimation equation includes the lagged outcome y_{i0} on the right hand side, providing the ANCOVA estimator. The specification allows the model to determine the structure of the relationship between the baseline and end-line levels of the outcome, rather than imposing a relationship as in the difference-in-differences estimator. Regression models of this kind are also more efficient than the difference-in-difference estimator. For example, “...with a single baseline and follow-up, one would need twice the sample size when using difference-in-difference to get the same power as obtained with ANCOVA.” (McKenzie, 2012, p. 212). For the purposes of controlling for potential nonrandom attrition, the estimation model includes additional controls as follows:

$$y_{i1} = \alpha + \delta D_i + \beta y_{i0} + \lambda X_{i0} + \varepsilon_i \tag{1}$$

where X_{i0} is a set of control variables, and each observation is appropriately weighted for attrition and sampling. The potential control variables include all five sampling variables and all ten demographics. Unless noted otherwise, the variables are standardized so that the magnitude of δ can be directly compared across different measures.

4.3. Treatment results captured by the surveys

We start by analyzing the 2014 survey data, reporting results for the variables that measure traditional entrepreneurial behavior: traditional entrepreneurial activity and the creation of a new business. We then report the effects on social leadership skills: motivation to lead and transformational leadership style. Finally, we add a discussion on treatment results on four measures on pro-social activities and identity: social entrepreneurial activity, sustainable behavior, social career intentions, and traditional career intentions.

We report WLS (OLS) treatment estimates, with and without conditioning on all relevant baseline covariates weighted (not weighted) by the share of sampling and nonresponse. The regression results for the 2014 round are reported in Table 2 (WLS) and in Appendix G, Table G5 (OLS).

INSERT TABLE 2 ABOUT HERE

In 2014, about half the treatment group (0.50) and half the control group (0.52) started a new venture during the observation period (means are from Appendix G, Table G7). The average amount of traditional entrepreneurial activity was 3.8 out of 22 in the treatment group and 3.6 out of 22 in the control group. Table 2 consequently reports no significant treatment effects on traditional entrepreneurial behavior. These results are robust to including/excluding covariates (Table 2) and removing weights (Appendix G, Table G5).

In social leadership skills, we find no detectable robust treatment effects on those skill developments either. For the social behavior and identity measures, we find that the average number of social entrepreneurial activities was 1.7 out of 5 in the treatment group and 1.4 out of 5 in the control group. Those in the treatment group performed 8.1 sustainable activities between $t = 0$ and $t = 1$, while those in the control group performed 8.2 out of 11. Hence, we find no detectable treatment effects on these measures. After covariates and weighting observations for response attrition are controlled for, the conclusion of no detectable treatment effects on all outcomes remain.

The results for the 2015 round are reported in Table 3 and show that about 75 percent of those in the treatment group (0.73) and 40 percent of those in the control group (0.40) started a new venture during the observation period (means are from Appendix G, Table G8). The average number of traditional entrepreneurial activities was 7.5 out of 22 for the treatment group and 4.3 out of 22 for the control group. Table 3 consequently reports significant treatment effects for traditional entrepreneurial activity and the creation of a new business. The results are robust to including/excluding covariates (Table 3) and different analytical methods (Appendix G, Tables G6 and G8). Traditional entrepreneurial activity increased over 60 percent in the treatment group compared to the control group, a high elasticity. The 95% confidence interval [CI] for this estimate is [0.26, 1.13]. In addition, the probability of starting a new business in the treatment group increases by 0.30 (95% CI is [0.14, 0.52]) percentage points over the control group's rate, an increase of 75 percent, an even larger effect.

The effect sizes are significantly larger in the 2015 round than in the 2014 round on traditional entrepreneurial activity but do not pass the standard tests for significant differences across cohorts for new business creation (*p*-values are listed at the bottom of Table 3).

We continue to report no detectable treatment results for all measures of social leadership skills and pro-social activity and identity. The average number of social entrepreneurial activities was 1.4 out of 5 in the treatment group and 1.3 out of 5 in the control group. Those in the treatment group displayed 7.0 sustainable behaviors between $t = 0$ and $t = 1$ while those in the control group displayed 7.6 out of 11. Table 3 reports no detectable treatment effects for social entrepreneurial activity or sustainable behavior. Finally, no treatment effects were found on social or traditional career intentions. Our results are robust to the analytical method used (see Appendix Tables G6 and G8).

INSERT TABLE 3 ABOUT HERE

4.4. Treatment results on career development

The survey data provided immediate feedback on the impact of the two programs. Combined with reflections from the organization and the researchers, this led to the revision of the program in the second round. We also collected career development data in 2017 to examine long-term impacts. These data were not available at the time of the discussion on how to revise the program in 2015, but they provide an important benchmark for measuring the sustainability of the impact of the training.

Career data were collected from LinkedIn, two and three years after the program concluded in 2015 and 2014, respectively. These data do not suffer much from attrition. Where attrition is found, it is due to the absence of data posted on LinkedIn by an individual, rather than our conducting the RCT. Because these data are likely unconfounded by the sampling method, we examine the data with Athey and Imbens's (2016) stratum-level analysis, producing simple *t*-tests of mean differences between the treatment and control groups.

Table 4 shows that, by 2017 and for the 2014 cohort, we find a significant difference between the treatment and control groups only in the number of socially responsible activities they had engaged in. In the 2015 cohort, the means often show greater differences between the two groups. A significantly higher number of members of the treatment group than the control group were founders or co-founders of a startup in the 2015 cohort. Further, in 2017 the number of startups, the number of socially responsible startups, and the number of social media (i.e., LinkedIn) connections is significantly higher for the treatment group than the control group in the 2015 cohort. The difference between the treatment and control groups with respect to career choices for being a founder, the number of companies founded, and the number of socially responsible startups founded are predominantly statistically significantly different between the 2014 and 2015 cohorts. Two *p*-values are below 0.05, and one is below 0.10 (Table 4, rightmost column).

Although the 2015 participants are significantly more active social entrepreneurs, their

firms do not have impressive employment growth, and they do not grow more quickly than the control group. Those in the treatment group who started a company employ on average 1.3 workers, i.e., themselves, as the founder of the business, and one part-time employee. Employment growth is similarly tepid by those that have started a business in the 2014 treatment group, even after having had one additional year to develop their business. For illustrative purposes, in Appendix I we describe four case studies of projects in the treatment group with the greatest employment growth, external funding, or revenues.

INSERT TABLE 4 ABOUT HERE

So far, we have analyzed the difference in entrepreneurial activity overall. One might wonder about differences in the traditional entrepreneurial activities between those in the treatment group and those in the control group in the 2015 round. An analysis of them indicates the mechanism of the business skills treatment. In Appendix G, Table G9, we perform additional *t*-tests on differences across each of the 22 traditional entrepreneurial activities. We find significant and large differences between the treatment and control groups for 11 of the 22 activities: “Have begun the preparation of a business plan,” “Have sent a formally written business plan to other people,” “Have tested your product or service that you want to sell,” “Have started marketing or promotion,” “Have submitted an application for a patent or copyright,” “Have purchased or rented equipment, facilities, or property,” “Have talked with potential customers,” “Have defined the market opportunities,” “Have developed financial projections or break-even analyses,” “Monthly revenue ever exceeded monthly expenses for the new business,” and “Potential customers can contact you by phone or through e-mail or a website.” In general, then, the treatment group in the 2015 round is more active across a range of entrepreneurial activity and, when measured, applied analytical entrepreneurial tools more often.

4.5. Robustness analysis

We examine several threats to the validity of our inferences: low power, major differences between the 2014 and 2015 rounds, nonrandom selection into the 2015 round, multiple hypothesis testing, *p*-fishing (*p*-hacking), and truncation bias. The largest threats to the validity of inferences come from low power and sampling difference. However, correcting for conducting multiple *t*-tests does not change inferences, we find no truncation bias in the data, and the purposeful selection of the top 25 suitable applicants by the organization to the 2015 treatment group does not bias the analysis of the career development data. We discuss the power issues and potential differences in cohort characteristics here. Details on the other robustness examinations are in Appendix H.

There is a reasonable concern that we falsely accept the null hypothesis of no treatment effect because of low statistical power due to a relatively modest sample size, with 61 observations in the 2014 round and 41 in the 2015 round, respectively. Although response rates increased in 2015, the sample size still dropped because of the nonrandom selection of 25 participants by the organization, leaving just 50 qualified individuals to be randomly allocated between the treatment and control groups.

To analyze statistical power, we start by calculating the recommended sample size to avoid falsely accepting the null hypothesis of no treatment effect on the 2014 cohort. For example, the average treatment effect size is 0.055, and the average standard deviation is 1.12 (s.e. $0.143 \cdot \sqrt{61}$) in Table 2, column 1. We expect, on average, the treatment to be positive. Hence, for a one-sided test with a default power of 80 percent, the recommended sample size to achieve 80 percent power is 2,561 observations.²⁴

The power is low for two reasons. First, the effect sizes in the 2014 round are small, just 5.5 percent of one standard deviation. The first conclusion is therefore that when our tests do

²⁴ power onemean 0 0.055, sd(1.119) onesided.

not detect significance, it is partly because the effect sizes are close to null. The second is that the sample sizes are small, producing large CIs. A large enough sample would eventually produce significance even at the 5.5 percent effect size, but this was not possible given the limited size of the program. Given the limited sample, the effect sizes, instead, need to be much greater than 0.06 to obtain clear inferences. Furthermore, given the cost of the program, the ability to reject a false negative at an effect size of 0.06 is neither feasible nor relevant. A much larger treatment effect was needed by the organizers to justify running the program. Indeed, we obtain considerably larger effect sizes for the entrepreneurial activity measures in the 2015 round.

In order that we do not falsely reject the null, we can calculate either the minimum required effect size given a certain sample size and expected power or the minimum sample size given a target effect size and expected power. The approaches are equivalent. We use the latter. With a target effect size of, for example, 0.62 and a s.e. of 0.22 for the number of traditional entrepreneurial activities (Table 3), the recommended N to obtain the default power of 0.8 is 32, which we exceed by a good margin. Similarly, the minimum sample size for a standard power of 0.8 for the difference in the proportion of participants starting a new business across the treatment and control groups in the 2015 round is $N = 17$, which we also surpass by a good margin.²⁵ The significant t -tests for the LinkedIn data have similar differences in proportion, implying that sufficient power is reached for those tests as well because in most of those tests N is at least 85, much greater than 17. We thus feel confident that these tests are not underpowered, in view of the various results from the 2015 round on traditional entrepreneurial activities and career outcomes. However, for results with an average effect size of 0.055, we can only conclude that they are not detectably larger than zero.

²⁵ power oneproportion 0.40 0.73, power(0.8).

Another concern might be that the two sets of results differ substantially, perhaps because of the changes in selection criteria between the two rounds produce different samples, thereby explaining the difference in treatment effects between them. Although all descriptors are not jointly statistically significant across the two rounds, Appendix Table G3 reports statistically significant differences between the two rounds in terms of three pre-treatment characteristics. This could explain the difference in treatment effects between the 2014 and 2015 rounds.

We use the method developed in Allcott (2015) to reweight the 2015 data so that observations look similar to the 2014 data for all pre-treatment characteristics. We then examine whether we still see a significant effect for a sample that looks the same as the top 100 sample in 2014 when exposed to the treatment in the 2015 round. The reweighting is performed as follows. We first run a logit regression for being selected into the top 100 applicants for the 2014 sample on all available pre-treatment characteristics X_0 and Y_0 , call them Z_0 , and save the regression coefficients. We then impute the probability of being selected to the top 100 for the 2015 sample, using these saved regression coefficients. We then calculate the selection weight as $p(Z_0)/[1-p(Z_0)]$ and multiply this weight with both the stratified sampling weight and the nonresponse weight previously calculate for the 2015 sample. The combined weight is then used in a WLS treatment effect regression, as in equation (1). The results of the reweighted treatment effects are presented in Table 5. Treatment effects are not dissimilar to those reported in Table 3, leading us to conclude that the slight compositional differences in the two cohorts do not explain the differences in the size of the treatment effects on entrepreneurial activity between 2014 and 2015.

INSERT TABLE 5 ABOUT HERE

5. DISCUSSION

We report a case in which we measure the impact of a training program to encourage social entrepreneurship among youth in France twice. The program was very costly to run, so the

impact measurement was important. The study, a collaboration between academia and practitioners, involved an initial impact assessment using an RCT, learning from that and making program adjustments, and a repeated RCT after substantial changes were made in the program. We also measured career development impacts a few years after each program had concluded.

Although entrepreneurship training in developing countries is evaluated more often using RCTs (see, e.g., McKenzie, *forthcoming*; McKenzie & Woodruff, 2014), social entrepreneurship training evaluation is largely based on qualitative accounts. Our intention was to enrich this field with a carefully conducted, scientifically valid impact measurement study. We report some notable execution hiccups that are not uncommon when conducting RCTs in the field but might not be well known by management or entrepreneurship scholars. We also report the evidence on the impacts obtained.

The study provides an example of how repeated RCTs can play a significant role in increasing the contribution of academic work in specific contexts and, in combination with other methods, can contribute to building academic knowledge and make useful predictions (for recent discussions about the advantages and drawbacks of RCTs, see Banerjee, 2020; Deaton & Cartwright, 2018; Duflo, 2020; Kremer, 2020; Nielsen & Miraglia, 2017). We provide an example of how repeating an RCT can be a powerful tool for improving training and making policy recommendations. Repeating the RCT is a key element that sets this study apart from one-time intervention studies (Duflo, 2020). At the same time, our intervention is context dependent and based on a small sample, and the results therefore need to be interpreted carefully (see also Deaton & Cartwright, 2018). The results are simply one piece of the puzzle to increase our knowledge about the potential effects of social entrepreneurship training programs. In that sense, our study forms part of a cumulative process in which evidence from RCTs, together with evidence based on other methods, help to increase knowledge about “why

things work,” not just that some things work in some contexts (Banerjee, 2020; Deaton & Cartwright, 2018). Our RCTs focus on estimating the impacts in a single context. We do not claim to be able to generalize from this context. But we hope that this article stimulates others to apply repeated RCTs, as they are clearly underrepresented in policy evaluation studies. Repeated RCTs add useful knowledge on how organizations learn in the field and add knowledge about why some programs work (see also Banerjee, 2020; Duflo, 2020).

Why should we conduct RCTs in the first place when qualitative accounts are typically much simpler and less costly to obtain? The benefit of RCTs is that if the execution is by the book, the usual complaints about omitted variable bias disappears and a causal claim about what works, or does not work, can be made. This is truly important, as in social entrepreneurship “influential organizations and associations are carefully promoting social entrepreneurship by providing compelling anecdotal evidence of heroic individuals changing the world” (Dacin et al., 2011, p. 1), thereby providing a different and perhaps less critical view of the impacts of training. In our case, despite a very positive perception of the first round of the program in the media, by policy makers, and by the general public, the first RCT did not lead to the same rosy conclusions. In fact, thanks to using a scientific approach in impact assessment, the program was considerably redesigned before it was offered again, and we could make some inferences about how that redesign affected its impact in the second round. We thus see our repeated RCTs as an example of how policy and public and private spending can be rigorously evaluated, and if necessary, entrepreneurial training programs can be redesigned and improved.

As mentioned, the implementation was not without difficulties, which to some extent limits the inferences that can be made. Execution problems in RCTs are common (e.g., Duflo, 2020). In development studies, perhaps the most frequent problem in RCTs is that participants either do not show up for training or do not complete it. Angus Deaton has made a large point

of this, as it may affect the ability to draw causal inferences (e.g., Deaton & Cartwright, 2018). In short, if participation is nonrandom, it does not matter whether allocation to treatment is random.

In terms of the challenges we faced, lack of participation was not the problem. The biggest hurdle was, rather, gaining approval from our partner to randomize applicants to treatment. We persuaded the organization to accept randomization in the first round, but as no measurable impacts were detected in it and pressure was mounting on the organization, we could not convince it to accept full randomization in the second round. In line with Duflo (2020), we experienced that randomization is easier to implement in the absence of prior data, and the expectation is that a program has a significant positive impact. However, when test results do not provide support for anticipated effects, it becomes more difficult to repeat the test. Researchers are then in peril. Although formal collaboration agreements can be crafted between academia and practitioners, it is unclear how much leverage they create. We agree with Kremer (2020), who explains that RCTs require a significant amount of time in the field and in discussions with the potential participants, trainers, and other relevant people on the ground, on whom conducting RCTs depend. This interrogation process equips experimentalists with rich knowledge and a sense of context that informs the design of an RCT. It is also a source of qualitative data helping to interpret results and to learn, by which new hypotheses can be generated and later tested. For example, in our case this involved interaction with the organization to determine the changes that needed to be made in the training program, the introductions by the organization that enabled us to conduct case studies, and the considerable support needed to expand the share of respondents.

A second execution problem was the nonrandom attrition in survey responses. The inference problem resembles when people nonrandomly show up for training. The difficulty of obtaining survey responses is not new and not unique to RCTs. Personal reminders from the

organization had an important impact on boosting response rates. Compared to online surveys, for example, our response rates are extremely high. But the unevenness of replies suggests that one should be cautious when drawing conclusions: in the control groups, the response rate was 46 percent (2014) and 60 percent (2015), whereas in the treatment group the response rate was substantially higher, 76 percent (2014) and 100 percent (2015). The estimation section describes some state-of-the-art statistical adjustments made to control for nonrandom attrition. But attrition might also be a function of selection on unobservable characteristics that is harder to deal with, and throwing in regressors blindly might also be inappropriate. We therefore report results that analyze treatment effects on entrepreneurial career outcomes collected from LinkedIn. This approach reduces the effort needed for participants to respond to surveys, reducing attrition to a few percentage points, and it does not make attrition dependent on whether an individual is allocated to the treatment or control group.

There are a few execution issues. First, the new General Data Protection Regulation (GDPR) require asking for consent to collect data and specifically to link that with public data, and that might reduce potential participants' willingness to take part in the research. Second, one cannot quickly collect short-term outcomes from public data. The entire setup of test-learn-retest in this study depended on quickly executed and tailored end-line surveys. Nevertheless, the holy grail everyone seeks in such training is persistent long-term impacts. A combination of quick end-line surveys for rapid adjustment and the use of long-term publicly available outcome data collected later therefore seems ideal for policy evaluation.

The third execution issue relates to the many changes made between the two rounds of the program. The organization was under pressure to show results in the second round and simply could not accept making only one change in the program. We could control for the new selection criteria imposed using advanced statistical techniques (Allcott, 2015). But because of the simultaneity in content changes, we could not disentangle the contribution of the intense

12-day training component from that of the individualized coaching.

The results from the repeated RCTs that we report should be interpreted in light of the challenges outlined above. The initial 2014 training program had two main components: accelerating social entrepreneurial activity and promoting leadership skills associated with social entrepreneurial identity. We find no detectable treatment effects from that program. These findings led the organizers to change the selection process, the program's content, and the follow-up coaching in the 2015 round. Less time was spent on developing leadership skills. The revised training included more training in analytical entrepreneurial skills applied to their venture idea, primarily using the "business model canvas" and "lean startup" methodology. Finally, coaching was better structured pedagogically.

We repeat the RCT on the second round in 2015. The results this time are encouraging. The treatment has a broad effect on a range of traditional entrepreneurial activity and when measured with the use of multiple analytical tools. Career development data show that those in the treatment group in the redesigned program are more likely to become a founder of a startup, to start more socially responsible startup, and to have more social media connections than those in the control group. The program changes focused more on entrepreneurial skills training, more on taking action, and with better individual coaching therefore have a measurable impact on both immediate entrepreneurial activity and career development. There were still no detectable treatment effects on social leadership skills and various pro-social identity measures.

With the scientific caveats in mind, these patterns have implications for social entrepreneurship program design, firms designing internal training programs for stimulating innovative and entrepreneurial behavior with a pro-social dimension, as well as incubators, accelerators, and other organizations wishing to impart social entrepreneurial skills to their participants, in addition to government institutions and philanthropy to support the most promising programs for increasing the number of successful social entrepreneurs.

As Kremer (2020, p. 1976) states, “field experiments are useful not just for testing the researcher’s preexisting hypotheses, but also for generating new hypotheses.” To that end, our study found that training and using entrepreneurial analytical skills and affecting entrepreneurial activity mattered considerably more than teaching social leadership skills and fostering a socially oriented identity. These results can be interpreted in at least two ways. The first is that the participants in this study were young people with social leadership skills and pro-social identity that are already rather well developed and thus only marginally changeable. The second is that social leadership training requires considerably more effort. Organizers of social entrepreneurial programs should note that, despite considerable training and inspirational efforts, we observe no treatment effects on social entrepreneurial intentions, activity, or sustainable behavior across the two rounds of the program. More research is necessary to make recommendations about the right balance between entrepreneurial skills and social leadership skills training, as well as the intensity of training in both components, to make social entrepreneurship training successful.

What differentiates this training program from other entrepreneurial training programs, and how can these differences explain some of the results? Here we must resort to pure speculation. As opposed to participants in accelerators, the participants are not selected to create high-tech or fast acceleration businesses. Participants instead target social causes such as recycling, helping the poor or disenfranchised, and sustainable development. The double-bottom line in social entrepreneurship might therefore mean slower growth, as reflected in the decidedly tepid growth in the number of employees in these projects. And as already mentioned, the screening process for the program might also mean that the participants already have well developed social leadership skills and social entrepreneurial identities.

Our methodological approach, based on repeated field experiments, shows how entrepreneurship training programs can be evaluated in order to recommend improvements, in

a context in which evaluation has largely been based on qualitative accounts. More research using RCTs will help the field to better balance the use of different research methods and to use the powerful RCT-inherent design element that allows for replication (Banerjee, 2020; Duflo, 2020).

REFERENCES

- Allcott, H. (2015). Site selection bias in program evaluation. *Quarterly Journal of Economics*, *130*(3), 1117–1165.
- Altman, D.G. (2005). Adjustment for covariate imbalance. In P. Armitage & T. Colton (eds.), *Encyclopaedia of biostatistics*, 2nd ed. (pp. 1273–1278). Chichester: John Wiley.
- Anderson, S.J., Chandy, R., & Zia, B. (2018). Pathways to profits: Identifying separate channels of small firm growth through business training. *Management Science*, *64*(12), 5461–5959.
- André, K., & Pache, A.-C. (2016). From caring entrepreneur to caring enterprise: Addressing the ethical challenges of scaling up social enterprises. *Journal of Business Ethics*, *133*(4), 659–675.
- Athey, S., & Imbens, G. (2016). The econometrics of randomized experiments. Working paper, Stanford University.
- Austin, J., Stevenson, H., & Wei-Skillern J. (2006). Social and commercial entrepreneurship: Same, different, or both? *Entrepreneurship: Theory and Practice*, *30*(1), 1–22.
- Bacq, S., Hartog, C., & Hoogendoorn B. (2016). Beyond the moral portrayal of social entrepreneurs: An empirical approach to who they are and what drives them. *Journal of Business Ethics*, *133*(4), 703–718.
- Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B., & Udry, C. (2015). A multifaceted programme causes lasting progress

- for the very poor: Evidence from six countries. *Science*, 348(6236). DOI: 10.1126/science.1260799.
- Banerjee, A. (2020). Field experiments and the practice of economics. *American Economic Review*, 110(7), 1937–1951.
- Betcherman, G., Godfrey, M., Puerto, S., Rother, F., & Stavreska, A. (2007). *A review of interventions to support young workers: Findings of the youth employment inventory*. Retrieved from <http://www.youth-employmentinventory.org/downloads/1.pdf>.
- Campos, F., Frese, M., Goldstein, M., Iacovone, L., Johnson, H.C., McKenzie, D., & Mensmann, M. (2017). Teaching personal initiative beats traditional training in boosting small business in West Africa. *Science*, 357(6357), 1287–1290.
- Chan, K.Y., & Drasgow, F. (2001). Toward a theory of individual differences and leadership: Understanding the motivation to lead. *Journal of Applied Psychology*, 86(3), 481–498.
- Chatterji, A., Delecourt, S., Hasan, S., & Koning, R. (2019). When does advice impact startup performance? *Strategic Management Journal*, 40(3), 331–356.
- Cho, Y., & Honorati, M. (2014). Entrepreneurship programs in developing countries: A meta regression analysis. *Labour Economics*, 28(c), 110–130.
- Corner, P.D., & Ho, M. (2010). How opportunities develop in social entrepreneurship. *Entrepreneurship Theory and Practice*, 34 (4), 635–659.
- Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review, Papers and Proceedings*, 97(2), 31–47.
- Dacin, P.A., Dacin, M.T., Matear, M., & Tracey, P. (2011). Social entrepreneurship: A critique and future directions. *Organization Science*, 22(5), 1203–1213.
- Dees, J.G., Emerson, J., & Economy, P. (2001). *Enterprising nonprofits: A toolkit for social entrepreneurs*. New York, NY: Wiley.
- Desa, G., & Basu, S. (2013). Optimization or bricolage? Overcoming resource constraints in

- global social entrepreneurship. *Strategic Entrepreneurship Journal*, 7(1), 26–49.
- Duflo, E. (2020). Field experiments and the practice of policy. *American Economic Review*, 110(7), 1952–1973.
- Economist (2006, February 23). The rise of the social entrepreneur. Retrieved from <http://www.economist.com/node/5517666/>
- Elert, N., Andersson, F., & Wennberg, K. (2015). The impact of entrepreneurship education in high school on long-term entrepreneurial performance. *Journal of Economic Behavior and Organization*, 111, 209–223.
- European Commission (EC). (2014). Social entrepreneurship. Retrieved from http://ec.europa.eu/internal_market/social_business/index_en.htm
- Fairlie, R., Karlan, D., & Zinman, J. (2015). The effects of entrepreneurship training: Evidence from a randomized control experiment. *American Economic Journal: Economic Policy*, 7(2), 125–161.
- Forbes (2014). 30 under 30: Social entrepreneurs, class of 2014. Retrieved from <http://www.forbes.com/special-report/2014/30-under-30/social-entrepreneurs.html>
- Gertler, P. (2017). *Making entrepreneurs: The return to training youth in hard versus soft business skills from an at-scale field experiment*. Paper presented at the sixth HEC workshop on Entrepreneurship and Development, HEC Paris, Jouy-en-Josas, December 15–16.
- Gonzalez-Uribe, J., & Leatherbee, M. (2018). The Effects of Business Accelerators on Venture Performance: Evidence from Start-Up Chile. *Review of Financial Studies*, 31(4), 1566–1603.
- Goodwin, V.L., Wofford, J.C., & Whittington, J.L. (2001). A theoretical and empirical extension to the transformational leadership construct. *Journal of Organizational Behavior*, 22(7), 759–774.

- Heckman, J.J., & Kautz, T. (2013). Fostering and measuring skills: Interventions that improve character and cognition. Working paper. Retrieved from <http://www.nber.org/papers/w19656/>
- Holt, D., Smith, T.M.F., & Winter, P.D. (1980). Regression analysis of data from complex surveys. *Journal of the Royal Statistical Society, Series A*, 143(4), 474–487.
- Howorth, C., Smith, S.M., & Parkinson, C. (2012). Social learning and social entrepreneurship education. *Academy of Management Learning and Education*, 11(3), 371–389.
- Imbens, G., & Wooldridge, J. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 1–81.
- Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, 15(4), 309–334.
- Karlan, D., & Valdivia, M. (2011). Teaching entrepreneurship: Impact of business training on micro finance clients and institutions. *Review of Economics and Statistics*, 93(2), 510–527.
- Kickul, J., Janssen-Selvadurai, C., & Griffiths, M.D. (2012). A blended value framework for educating the next cadre of social entrepreneurs. *Academy of Management Learning and Education*, 11(3), 479–493.
- Ko, W.W., & Liu, G. (2015). Understanding the process of knowledge spillovers: Learning to become social enterprises. *Strategic Entrepreneurship Journal*, 9(3), 263–285.
- Kremer, M. (2020). Experimentation, innovation, and economics. *American Economic Review*, 110(7), 1974–1994.
- Kroeger, A., & Weber, C. (2014). Developing a conceptual framework for comparing social value creation. *Academy of Management Review*, 39(4), 513–540.
- Kwong, C.C.Y., Thompson, P., & Cheung, C.W.M. (2012). The effectiveness of social business plan competitions in developing social and civic awareness and participation.

- Academy of Management Learning and Education*, 11(3), 324–348.
- Lafortune, J., Riutort, J., & Tessada, J. (2018). Role models or individual consulting: The impact of personalizing micro-entrepreneurship training. *American Economic Journal: Applied Economics*, 10(4), 222–45.
- Litzky, B.E., Godshalk, V.M., & Walton-Bongers, C. (2010). Social entrepreneurship and community leadership: A service-learning model for management education. *Journal of Management Education*, 34(1), 142–162.
- Lyons, E., & Zhang, L. (2017). Who does (not) benefit from entrepreneurship programs? *Strategic Management Journal*, 39(1), 85–112.
- MacKenzie, S.B., Podsakoff, P.M., & Rich, G.A. (2001). Transformational and transactional leadership and salesperson performance. *Journal of the Academy of Marketing Science*, 29(2), 115–134.
- Mair, J., Battilana, J., & Cardenas, J. (2012). Organizing for society: A typology of social entrepreneuring models. *Journal of Business Ethics*, 111(3), 353–373.
- Martin, B.C., McNally, J.J., & Kay, M.J. (2013). Examining the formation of human capital in entrepreneurship: A meta-analysis of entrepreneurship education outcomes. *Journal of Business Venturing*, 28(2), 211–224.
- McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, 99(2), 210–221.
- McKenzie, D. *forthcoming*. Small business training to improve management practices in developing countries: Re-assessing the evidence for “training doesn’t work.” *Oxford Review of Economic Policy*.
- McKenzie, D., & Woodruff, C. (2014). What are we learning from business training and entrepreneurship evaluations around the developing world? *World Bank Observer*.
- Retrieved from

https://openknowledge.worldbank.org/bitstream/handle/10986/22564/wbro_29_1_48.pdf;sequence=1/

- McKinsey & Company and e-fellows.net. (2009). Most wanted: The ideal employer survey. McKinsey & Company, Düsseldorf.
- Miller, T.L., Wesley, C.L., & Williams, D.E. (2012). Educating the minds of caring hearts: Comparing the views of practitioners and educators on the importance of social entrepreneurship competencies. *Academy of Management Learning and Education*, 11(3), 349–370.
- Nagel, H., Rosendahl Huber, L., Van Praag, M., & Goslinga, S. (2019). The effect of a tax training program on tax compliance and business outcomes of starting entrepreneurs: Evidence from a field experiment, *Journal of Business Venturing*, 34(2), 261–283.
- Nga, J.K.H., & Shamuganathan, G. (2010). The influence of personality traits and demographic factors on social entrepreneurship start up intentions. *Journal of Business Ethics*, 95(2), 259–282.
- Nielsen, K., & Miraglia, M. (2017). What works for whom in which circumstances? On the need to move beyond the “what works?” question in organizational intervention research. *Human Relations*, 70(1), 40–62.
- OECD (2011). Fostering innovation to address social challenges. Retrieved from <https://www.oecd.org/sti/inno/47861327.pdf>
- Oosterbeek, H., Van Praag, M., & Ijsselstein, A. (2010). The impact of entrepreneurship education on entrepreneurship competencies and intentions: An evaluation of the junior achievement student mini-company program. *European Economic Review*, 54(3), 442–454.
- Pache, A.-C., & Chowdhury, I. (2012). Social entrepreneurs as institutionally embedded entrepreneurs: Toward a new model of social entrepreneurship education. *Academy of*

- Management Learning and Education*, 11(3), 494–510.
- Peredo, A.M., & McLean, M. (2006). Social entrepreneurship: A critical review of the concept. *Journal of World Business*, 41(1), 56–65.
- Peterman, N.E., & Kennedy, J. (2003). Enterprise education: influencing students' perceptions of entrepreneurship. *Entrepreneurship: Theory and Practice*, 28(2), 129–144.
- Plaskoff, J. (2012). Building the heart and the mind: An interview with leading social entrepreneur Sarah Harris. *Academy of Management Learning & Education*, 11(3), 432–441.
- Podsakoff, P.M., MacKenzie, S.B., Morrman, R.H., & Fetter, R. 1990. Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors. *Leadership Quarterly*, 1(2), 107–142.
- Rosendahl-Huber, L., Sloof, R., & van Praag, C.M. (2014). The effect of early entrepreneurship education: Evidence from a field experiment. *European Economic Review*, 72(11), 76–97.
- Short, J.C., Moss, T.W., & Lumpkin, G.T. (2009). Research in social entrepreneurship: Past contributions and future opportunities. *Strategic Entrepreneurship Journal*, 3(2), 161–194.
- Smith, B.R., Kitruck, G.M., & Cannatelli, B. (2016). The impact of moral intensity and desire for control on scaling decisions in social entrepreneurship. *Journal of Business Ethics*, 133(4), 677–689.
- Smith, I.H., & Woodworth, W.P. (2012). Developing social entrepreneurs and social innovators: A social identity and self-efficacy approach. *Academy of Management Learning and Education*, 11(3), 390–407.
- Smith, W.K., Besharov, M.L., Wessels, A.K., & Chertok, M. (2012). A paradoxical leadership model for social entrepreneurs: Challenges, leadership skills, and pedagogical tools for

- managing social and commercial demands. *Academy of Management Learning & Education*, 11(3), 463–478.
- Souitaris, V., Zerbinati, S., & Al-Laham, A. (2007). Do entrepreneurship programmes raise entrepreneurial intention of science and engineering students? The effect of learning, inspiration and resources. *Journal of Business Venturing*, 22(4), 566–591.
- Szucs, D., & Ioannidis, J.P.A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biol.* 15(3), e2000797, <https://doi.org/10.1371/journal.pbio.2000797/>.
- Tan, W.L., Williams, J., & Tan, T.M. (2005). Defining the “social” in “social entrepreneurship”: Altruism and entrepreneurship. *International Entrepreneurship and Management Journal*, 1(3), 353–365.
- Thompson, T.A., Purdy, J.M., & Ventresca, M.J. (2018). How entrepreneurial ecosystems take form: Evidence from social impact initiatives in Seattle. *Strategic Entrepreneurship Journal*, 12(1), 96–116.
- Tracey, P., Phillips, N. (2007). The distinctive challenge of educating social entrepreneurs: A postscript and rejoinder to the special issue on entrepreneurship education. *Academy of Management Learning and Education*, 6(2), 264–271.
- Universum (2009). *Ideal employers: Pan-European student survey*. Cologne: Universum.
- Van der Sluis, J., van Praag, M., & Vijverberg, W. (2005). Entrepreneurship selection and performance: A meta-analysis of the impact of education in developing economies. *World Bank Economic Review*, 19(2), 225–261.
- von Graevenitz, G., Harhoff, D., & Weber, R. (2010). The effects of entrepreneurship education. *Journal of Economic Behavior & Organization*, 76(1), 90–112.
- Yu, S. (2019). How do accelerators impact the performance of high-technology ventures? *Management Science*, 66(2), 503–1004.

Zietsma, C., & Tuck, R. (2012). First, do no harm: Evaluating resources for teaching social entrepreneurship. *Academy of Management Learning & Education*, 11(3), 512–517.

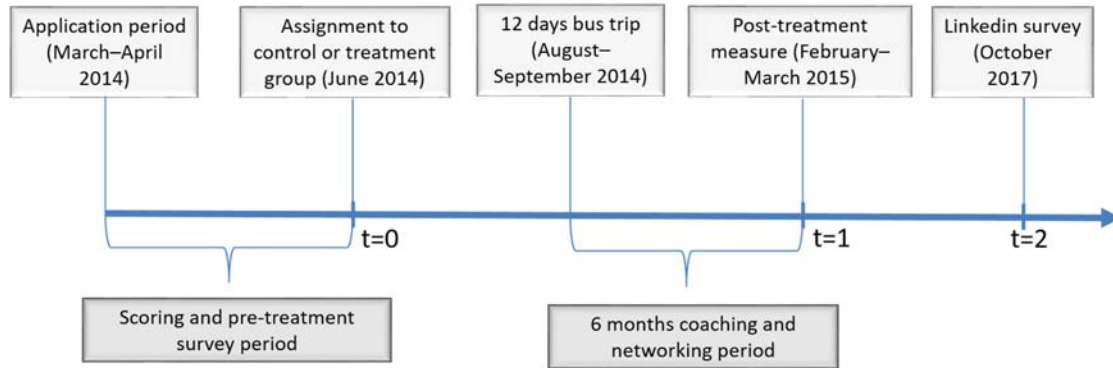


FIGURE 1 Timeline of Data Collection

TABLE 1 Main changes in the program between 2014 and 2015

Program phase	2014 program	2015 program
<i>Selection</i>		
Participants' profiles	Successful applicants had to show high motivation to positively change society.	Successful applicants had to show high motivation to positively change society and have at least some previous entrepreneurship-relevant experience.
Participants' commitment	Applicants were not asked to make a commitment to attend the follow-up coaching program.	Applicants had to commit to attending the follow-up coaching program.
<i>Bus trip</i>		
General program design	4 days of inspiration, 2 days of introspection, and 6 days of taking action.	3 days of inspiration, 3 days of introspection, and 6 days of taking action.
Pedagogy	Pedagogical focus on meeting role models, leadership training, and getting inspired.	Pedagogical focus on learning and applying analytical skills and engaging in entrepreneurial activities.
Content	Focus on leadership training and personal development exercises.	Focus on lean startup method and business modeling in combination with other analytical skills.
Content delivery	Long lectures and in-depth post-lecture general discussions.	Lecture time were reduced in favor of time for the participants to apply concepts to their concrete entrepreneurial projects.
Mentoring	Mentors were inspiring role models. Mentoring focused on getting to know the mentor's development path and to learn from it.	Mentors were inspiring role models. Mentoring was meant to help participants to progress with their projects. The amount of mentoring was higher in 2015.
<i>Follow-up coaching</i>		
	Six-month follow-up consisted mainly of (1) informal lunches every Monday with exchanges of information about progress on projects and provision of contacts for pressing needs (two to three hours), (2) a formal event on October 29 with about 20 participants, (3) two weekends (in December and February) for about 20–30 participants, (4) a one-day coaching session in collaboration with a French university for six participants, and (5) two meetings with the organization's person in charge with 20–30 participants.	Ten-month follow-up program consisted of five weekends (one weekend every six weeks) were participants received training on team management, digital communication strategies, personal development and general leadership training. Participants also tested their products and services through design thinking methods. The attendance rate was approximately 70–80 percent among the participants who went on the bus trip.

TABLE 2 Entrepreneurial behavior, leadership skills, and social identity, 2014 program

Measures	WLS, no controls		WLS, with controls	
	(1) δ	(2) β	(3) δ	(4) β
Traditional Entrepreneurial Behavior				
Entrepreneurial activities	0.18 (0.21)	0.37 (0.11)**	0.16 (0.22)	0.36 (0.11)**
New business creation	0.02 (0.16)	n.a.	-0.00 (0.17)	n.a.
Social Leadership Skills				
MTL affective identity	0.02 (0.08)	-0.23 (0.08)*	0.01 (0.08)	-0.23 (0.09)*
MTL noncalculative	0.06 (0.10)	-0.31 (0.06)***	0.08 (0.09)	-0.31 (0.06)***
MTL social-normative	-0.02 (0.08)	-0.43 (0.13)**	-0.03 (0.08)	-0.43 (0.13)**
TLS vision	-0.06 (0.14)	-0.41 (0.10)**	-0.07 (0.13)	-0.42 (0.10)**
TLS role model	0.08 (0.12)	-0.62 (0.15)**	0.10 (0.13)	-0.63 (0.15)***
TLS group goals	0.11 (0.16)	-0.44 (0.12)**	0.10 (0.17)	-0.49 (0.15)**
Pro-Social Activities and Identity				
Social entr. activity	0.19 (0.31)	0.45 (0.17)*	0.16 (0.33)	0.44 (0.18)*
Sustainable behavior	-0.13 (0.18)	0.54 (0.10)***	-0.16 (0.20)	0.54 (0.10)***
Social career intentions	0.19 (0.10)	-0.73 (0.20)**	0.15 (0.10)	-0.73 (0.19)**
Traditional career intentions	0.02 (0.08)	-0.42 (0.19)*	0.02 (0.09)	-0.42 (0.20)
<i>N</i>	61		61	

Note: *** <0.001, ** <0.01, * <0.05, two-sided. Standard errors are clustered by stratum and reported in parentheses. Variables standardized. Individual without repeated values at $t = 0$ and $t = 1$ omitted. Estimates are from separate regressions for each row. Coefficient estimates for the control variables and constants are suppressed, but available from the authors on request. Treatment effects reported in cols. 1 and 3. Coefficient for lagged outcome (y_0) reported in cols. 2 and 4. Significant controls included: gender and student. Regression weighted with inverse of sampling and nonresponse probability. MTL = motivation to lead, TLS = transformational leadership style. n.a. = not available.

TABLE 3 Entrepreneurial behavior, leadership skills, and social identity, 2015 program

Measures	WLS, no controls		WLS, with controls	
	(1) δ	(2) β	(3) δ	(4) β
Traditional Entrepreneurial Behavior				
Entrepreneurial activity	0.62 (0.22)*	0.52 (0.14)**	0.65 (0.21)**	0.60 (0.13)***
New business creation	0.29 (0.09)**	n.a.	0.30 (0.09)**	n.a.
Social Leadership Skills				
MTL affective identity	-0.02 (0.12)	-0.27 (0.15)	-0.03 (0.12)	-0.25 (0.18)
MTL noncalculative	0.02 (0.17)	-0.43 (0.25)	0.01 (0.18)	-0.41 (0.21)
MTL social-normative	0.00 (0.15)	-0.53 (0.10)***	-0.00 (0.16)	-0.51 (0.09)***
TLS vision	0.06 (0.19)	-0.77 (0.22)**	0.06 (0.19)	-0.78 (0.22)**
TLS role model	0.03 (0.15)	-0.81 (0.19)**	0.02 (0.14)	-0.79 (0.19)**
TLS group goals	0.30 (0.17)	-0.51 (0.23)*	0.29 (0.18)	-0.52 (0.24)*
Pro-Social Activities and Identity				
Social entr. activity	0.21 (0.20)	0.31 (0.12)*	0.23 (0.20)	0.31 (0.12)*
Sustainable behavior	0.02 (0.289)	0.61 (0.14)**	0.05 (0.301)	0.68 (0.140)***
Social career intentions	-0.09 (0.10)	-0.79 (0.06)***	-0.08 (0.10)	-0.79 (0.06)***
Traditional career intentions	0.18 (0.23)	-0.60 (0.18)**	0.19 (0.24)	-0.59 (0.16)**
p-value diff. between 2015 and 2014 programs				
Entrepreneurial activity	0.06		0.02	
New business creation	0.13		0.10	
<i>N</i>	41		41	

Note: *** <0.001, ** <0.01, * <0.05, † <0.10 two-sided. Standard errors are clustered by stratum and reported in parentheses. Variables standardized. Individual without repeated values at $t = 0$ and $t = 1$ omitted. Estimates are from separate regressions for each row. Coefficient estimates for the control variables and constants are suppressed, but available from the authors on request. Treatment effects reported in cols. 1 and 3. Coefficient for lagged outcome (y_0) reported in cols. 2 and 4. Significant controls included: gender and student. Regression weighted with inverse of sampling and nonresponse probability. MTL = motivation to lead, TLS = transformational leadership style. n.a. = not available.

TABLE 4 Career, business, and personal network development, 2014 and 2015

	2014				2015				
	Treatment Group (std. err.) (1)	Control Group (std. err.) (2)	Difference (std. err.) (3)	No. of observations (4)	Treatment Group (std. err.) (5)	Control Group (std. err.) (6)	Difference (std. err.) (7)	No. of observations (8)	<i>p</i> -value diff. between 2014 and 2015
Founder or co-founder of a startup? (Yes/No)	0.34 (0.08)	0.34 (0.08)	0.00 (0.06)	85	0.37 (0.08)	0.10 (0.05)	0.27*** (0.05)	93	0.03
Number of startups founded or co-founded?	0.41 (0.12)	0.39 (0.12)	0.03 (0.07)	85	0.40 (0.09)	0.12 (0.05)	0.28*** (0.05)	93	0.07
Number of socially responsible startups founded or co-founded?	0.32 (0.10)	0.27 (0.10)	0.04 (0.06)	85	0.35 (0.08)	0.04 (0.02)	0.31*** (0.05)	93	0.03
Number of full-time employees	1.27 (0.71)	1.21 (0.61)	0.05 (0.93)	25	1.29 (0.54)	1.80 (1.56)	-0.50 (1.28)	22	
Company closure (Closed = 1, Survived = 0)	0.18 (0.12)	0.07 (0.00)	0.11 (0.13)	25	0.06 (0.06)	0.20 (0.20)	-0.14 (0.15)	22	
External funding? (Yes = 1, No = 0)	0.64 (0.15)	0.64 (0.13)	0.00 (0.20)	25	0.06 (0.05)	0.00 (0.00)	0.06 (0.11)	22	
Worked at someone else's startup?	0.39 (0.08)	0.34 (0.09)	0.05 (0.06)	85	0.22 (0.08)	0.25 (0.08)	-0.03 (0.05)	93	
Number of socially responsible activities	2.21 (0.22)	1.59 (0.25)	0.63 (0.21)**	85	1.84 (0.26)	1.89 (0.23)	-0.05 (0.17)	93	
LinkedIn connections	396.37 (31.50)	375.42 (30.76)	20.94 (29.25)	83	432.38 (14.95)	397.53 (23.55)	34.84* (14.31)	85	
Number of skills endorsed	16.08 (1.97)	18.14 (2.21)	-2.05 (2.01)	75	16.07 (1.40)	17.60 (1.72)	-1.53 (1.17)	85	
Number of endorsements	91.38 (26.36)	80.46 (33.96)	10.91 (39.35)	75	83.07 (14.94)	71.02 (16.46)	12.04 (19.91)	85	
Sector pursuing career in? (Private sector = 1, Public sector/Nonprofit = 0)	0.35 (0.09)	0.39 (0.10)	-0.04 (0.06)	80	0.69 (0.09)	0.62 (0.09)	0.06 (0.06)	93	
Pursued graduate school? (Yes/No)	0.30 (0.08)	0.23 (0.08)	0.06 (0.06)	80	0.11 (0.05)	0.10 (0.05)	0.01 (0.05)	93	

Note: *** <0.001, ** <0.01, * <0.05, † <0.10. Standard errors were calculated using Athey and Imbens's (2016) stratum-level analysis. *P*-values calculated using two-sided tests. Data reflects activities since the end of the tour (either September 1, 2014, or September 1, 2015) and were collected in October 2017. *N* = 41 (45) treatment, 44 (48) control for 2014 (2015). If the participant founded more than one company, we display information on full-time employees and external funding for the largest surviving company in October 2017, if any.

TABLE 5 Using Allcott's reweighting procedure to make the 2015 participants reflect the composition of the 2014 participants

Measures	WLS, no controls		WLS, with controls	
	(1) δ	(2) β	(3) δ	(4) β
Traditional Entrepreneurial Behavior				
Entrepreneurial activities	0.58 (0.21)*	0.36 (0.254)	0.58 (0.17)**	0.41 (0.309)
New business creation	0.20 (0.07)*	n.a.	0.28 (0.05)***	n.a.
Social Leadership Skills				
MTL affective identity	0.11 (0.24)	0.94 (0.15)***	-0.03 (0.24)	0.96 (0.17)***
MTL noncalculative	0.01 (0.30)	0.81 (0.21)**	0.04 (0.32)	0.80 (0.22)**
MTL social-normative	0.49 (0.49)	0.51 (0.07)***	0.48 (0.54)	0.53 (0.07)***
TLS vision	0.08 (0.47)	0.27 (0.29)	0.02 (0.50)	0.30 (0.30)
TLS role model	-0.06 (0.33)	0.03 (0.18)	-0.07 (0.31)	0.09 (0.19)
TLS group goals	0.86 (0.36)*	0.34 (0.23)	0.78 (0.37)*	0.29 (0.24)
Pro-Social Activities and Identity				
Social entr. activities	0.01 (0.19)	0.09 (0.19)	0.07 (0.18)	0.10 (0.19)
Sustainable behavior	0.09 (0.26)	0.35 (0.21)	0.02 (0.30)	0.37 (0.23)
Social career intentions	-0.49 (0.31)	0.33 (0.12)*	-0.49 (0.31)	0.33 (0.10)**
Traditional career intentions	0.53 (0.30)	0.31 (0.13)*	0.49 (0.38)	0.30 (0.12)*
<i>N</i>	41		41	

Note: *** < 0.001, ** < 0.01, * < 0.05, two-sided. Standard errors are clustered by stratum and reported in parentheses. Variables standardized except for new business creation. Individual without repeated values at $t = 0$ and $t = 1$ omitted. Significant controls included: gender and student. Coefficient estimates for the control variables and constants are suppressed, but available from the authors on request. Each column is a separate regression. Regressions weighted with inverse of sampling, nonresponse probability, and Allcott's (2015) reweighting procedure to make the 2015 sample look like the 2014 sample. n.a. = not available.

APPENDIX

NOT FOR PUBLICATION

**Impact Measurement based on Repeated Randomized Control Trials:
The Case of a Training Program to Encourage Social Entrepreneurship**

APPENDIX A. Field Experiments on the Effects of Entrepreneurship Education and Training in Developed Countries.

Study	Peterman and Kennedy (2003)	Souitaris et al. (2007)	Oosterbeek et al. (2008)	Rosendahl-Huber et al. (2014)	Fairlie et al. (2015)
Sample	Grade 11 or 12, predominantly aged 16 in Australia.	Science and engineering students in London and Grenoble universities.	Bachelor students in administration, management, economics and law at two different locations in the Netherlands.	Children aged 11 or 12 from 63 primary schools (118 classes, 2,751 pupils) in the Netherlands.	Marketed to any individual interested in entrepreneurship in the U.S. Response: 19% self-employed and 39% on unemployment insurance.
Program	Junior Achievement Company Program (JACP). Students sell stock, elect officers, produce and market products or services; keep records and conduct shareholders' meetings.	Compulsory or elective module within entrepreneurship program.	JACP.	BizWorld. Taught by local entrepreneur and high-school teacher. Students sell stock, elect officers, produce and market products; keep records and conduct shareholders' meetings.	Growing America through Entrepreneurship (GATE) offered across seven cities 2003-2005. Free of charge. Classroom courses and one-on-one coaching.
Control group	Students from the same schools and the same class who had declined to enrol.	Non-entrepreneurship program students.	Students at another close location of the same university where JACP was not offered.	Random assignment to treatment or control group takes place at the class level.	Random assignment to treatment or control group.
Sample size: treatment / control	109 / 111	124 / 126	104 / 146	1,729 / 684	2,094 / 2,103. Effective sample by 3 rd wave survey 1,273 / 1,173.
Randomized allocation totreatment	No	No	No	Yes	Yes
Instrumented treatment	No	No	Yes, distance to school location from parents home.	N.A.	N.A.
Length of Treatment	5 months calendar time, teamwork, after school hours coaching by mentor.	One course or module.	One calendar year, teamwork, 5-10 h per week, lectures plus coaching by mentor. Students earn 10 ECTS.	5 days within a time span of 2 to 4 weeks. Teamwork.	15.6 h (13.8 h training, 1.8 h counseling). Cost per treated approx. \$850-\$1,300. Control group could seek training and did so for 6.9 hrs.
Outcome variable	Attitudes.	Attitudes, intentions, and actions.	Non-cognitive skills, entrepreneurial knowledge and intentions.	Non-cognitive skills, entrepreneurial knowledge and intentions.	Business plan writing, business start-up, employment, sales, household income, work satisfaction.
Outcome measurement time	End of program.	Approximately at end of course.	One to three months after treatment.	One month after treatment.	Follow-up at 6, 18, and 60 months after treatment.
Results	Positive effects on desirability and feasibility.	Positive effects on desirability and feasibility, zero effect on intentions and actions.	No effect on skills, sign. negative effect on intentions.	Significant positive effect on some non-cognitive skills, no effect on knowledge, sign negative effect on intentions.	Treatment group 11-13 percentage points more likely to create business plan and 2-6 percentage points more likely to start a business. No effect beyond 6 months for any outcome.

APPENDIX B1. Schedule of social entrepreneurship program (12 days) 2014.

Phase 1: "Inspiration"				
Place	Paris August 26	Paris August 27	Paris August 28	Marseille August 29
Pedagogical objective for the day	Start the Tour and establish a group dynamic	Foster group dynamic and get a sense of social entrepreneurship	Be inspired and unlearn prejudices about social entrepreneurship	Gain an understanding of social entrepreneurial solutions
Morning events		Team Building <i>The three coaches facilitate several team building exercises, physical challenges, and role plays. Subsequently, the group defines its rules of the trip, formulates expectations and fears.</i>	Site-immersion <i>Visit of two different social entrepreneurs and exchange about their experiences</i>	Treasure Map <i>Participants meet 10 social entrepreneurs on 10 different themes. Brainstorming and other creative methods are employed to help those social entrepreneurs.</i>
Afternoon events	First meeting among coaches and participants <i>The three coaches of the 2014 edition present the program of the following days. Subsequently, one famous French entrepreneur and one politician discuss with the participants about social entrepreneurship in France.</i>	France: needs and solutions <i>Several invited speakers from different fields (finance, IT, politics, etc.) present societal challenges of the 21st century. Based on these interventions, participants define the general issue that they want to work on (e.g., education, pollution, etc.) based on different working methods and tools.</i>	Meet social entrepreneurs <i>Informal meetings with four different famous French social entrepreneurs</i>	
Evening events	Launch reception <i>Public event to present and celebrate the 50 participants. Several famous people give short speeches or their video messages are projected. (450 persons)</i>	Forecast reception <i>Creativity workshop on innovative solutions for a better society facilitated by an external party through a gamification approach</i>	Unlearning Reception <i>Public event with invited speakers around the topics "unlearning" and "stereotypes" (200 persons)</i>	Inspiration Reception <i>Public event with invited speakers (social entrepreneurs) and first presentations from participants about their learnings. (200 persons)</i>

Phase 2: "Introspection"		Phase 3: "Take Action"		
Place	Les Amanins August 30	Les Amanins August 31	Lyon September 1	Strasbourg September 2
Pedagogical objective for the day	Discover individual talents and passions	Clarify personal vision and project mission	Learn how to prototype ideas	Develop sustainable business models; demystify entrepreneurial failure
Morning events	Discover agro-ecology <i>Visit of an agriculture project that aims at inventing new solutions for agriculture but also in terms of life styles.</i>	Searching for a personal project <i>Workshops and individual coaching sessions, group exercises, and role plays to find own project idea.</i>	Design Thinking <i>Workshop to find innovative solutions to identified societal issues</i>	Site-immersion <i>Visit of a successful social entrepreneurial organization and reflection on sustainable business models</i>
Afternoon events	Reflection <i>The participants meet and discuss with people (one CEO, one philosopher, one teacher) who follow the philosophy "Change oneself to change the world"; afterwards, they work on their own personal aspirations</i>	"My mission, my vision" <i>Workshop about generated output from the morning session and feedback rounds.</i>	Feedback session 1 <i>Collection of feedback from 25 local entrepreneurs and social entrepreneurs, and interviews with potential beneficiaries of project ideas</i>	Workshop <i>Participants learn about different business models and apply tools to further their own idea</i>

Evening events	In search of meaning <i>Participants watch a documentary and discuss with the producer about the film's idea and learnings</i>	Formation of Groups <i>Formation of groups to develop entrepreneurial projects</i>	Feedback session 2 <i>Feedback from 50 young persons</i>	Fail Night: demystifying failure <i>Public event: Testimonials of entrepreneurs who failed and rebounded (300 persons)</i>
-----------------------	--	--	--	--

Phase 3: "Take Action" (continued)				
Place	Lille September 3	Lille September 4	Paris September 5	Paris September 6
Pedagogical objective for the day	Learn about and experience financing and management issues	Learn about and experience leadership and communication	Improve presentation skills, create a network	Assess the Tour, feedback, and planning of the future
Morning events	The hive of financing <i>Three invited speaker teach participants about financing social entrepreneurial projects</i>	Leadership Workshop <i>Three experts on leadership work with participants on improving their personal leadership style</i>	Pitch preparation <i>Participants prepare their pitches based on feedback received</i>	Assessment of the Tour <i>Collective feedback rounds and exchange of strengths and weaknesses (facilitated by three coaches)</i>
Afternoon events	Communication workshop and pitches to investors <i>Participants receive a workshop on communication and afterwards pitch their projects to potential investors (banks, impact investors, philanthropists, foundations) in a "speed dating" event</i>	"Pitch your project" Workshop <i>Participants work on project presentations with the help of professionals</i>	Young entrepreneurs' pitches <i>Public event where participants' present their projects to a jury of partners (300 persons)</i>	Expectations for the future <i>Participants express their expectations for the follow-up program; milestones and future collaborations are defined</i>
Evening events	Live Reception <i>Live concert and festival (400 persons)</i>	Project presentations <i>Participants pitch their project ideas to experts for the first time and receive feedback</i>	Post-Tour Support Forum <i>Exchanges between the partners and participants about future support</i> Closing evening <i>Celebration of the first edition of the program in presence of the Minister of Youth, partners, jury members, etc. (600 persons)</i>	Closing of the program

The six-month follow-up program in 2014 consisted mainly of (1) informal lunches every Monday with exchanges of information about progress on projects and provision of contacts for pressing needs (two to three hours), (2) a formal event on October 29 with about 20 participants, (3) two weekends (in December and February) for about 20-30 participants, (4) a one-day coaching session in collaboration with a French university for six participants, and (5) two meetings with the organization's person in charge with 20-30 participants. Participation in all events was voluntary. The organization did not keep track of which participants attended lunches and events.

APPENDIX B2. Major differences between the 2014 and the 2015 program.

All questions in Appendix C showing the online document's questions remained either qualitatively or exactly the same, except for questions 10 to 14, which were replaced by questions asking applicants more about their role as a change maker and how they would cope with difficulties during their entrepreneurial adventure. In 2015 potential participants were required to upload a video on YouTube. This procedure replaced the 2014 bonus question for which uploading a video was only one of multiple options (see Appendix C). Participants' free-form answers were independently examined by up to three external judges (external jury) and two members from the organization (internal jury). Both juries received the same instructions, and scores were weighted together in a slightly more elaborate way than in 2014. Applicants were prescreened on three characteristics: capacity to dream, willingness to change society, and leadership potential. The fourth characteristic was included during in-person interviews described below. Each of the three characteristics was assessed based on a comparable number of items, which remained qualitatively the same as those shown in Appendix D, with slightly different weightings (capacity to dream 30%, willingness to change society 40%, leadership potential 30%). Weights of constructs and items were predetermined by the organizers based on their preferences. The organization (not the judges) formed a "preliminary suitability" score, which was based on a weighted average of the external jury's (one-third) and internal jury's (two-thirds) evaluation. Based on the preliminary suitability score ranking of applicants' online application, the organization interviewed 106 applicants in person (with some exceptions via Skype), and reassessed applicants' performance based on a second preliminary suitability score. The weighted average of the interview's preliminary suitability score (70%) and the online application's preliminary suitability score (30%) formed the final suitability score.

APPENDIX C. Application form for the social entrepreneurship program (extract).

A. General and Administrative Information

Last Name:
First Name:
Email address:
Phone number:
Sex:
Age:
Current city of residence:

What is your current situation? (Mark and respond on the corresponding line)

- Student. Specify your school level:
- Recent Graduate. Specify your diploma:
- Employed. Specify our position, your employer and contract type:
- Job Seeker.
- Other. Specify:

How did you hear of [the program]? (Mark and respond on the corresponding line)

- Social Network. Which one:
- The [the program] website.
- A partner site. Which one:
- A friend or a parent.
- Media. Which one:
- A conference. Which one:
- Other. Specify:

B. You and your motivation

As a reminder, here are the 4 criteria by which you will be selected: your ability to dream a more just society, your determination to push your boundaries, your leadership potential, and your ability to communicate your enthusiasm. All the questions in this section are mandatory.

We are looking for personalities, not CVs...so be you!

1. Who are you: tell us your story. (300 words max)

We want to know more about your journey, your stages in life and the most important experiences that have led you this far.

2. What are the 3 qualities that characterize you the most? Illustrate each in a concrete example. (200 words max)

3. What societal problem concerns you the most (discrimination, disabilities, exclusion, diseases, homelessness, environment, etc.)? Why? (200 words max)

4. Imagine a world where anything is possible. Propose to us an idea – even the craziest – to solve this problem. (200 words max)

5. In 20 years, what will the world look like as a result of this idea? Describe to us how this idea has helped change things. (200 words max)

6. Think of the last time you did something “crazy”: out of your routine, explored the unknown, out of line, dared to swim against the current, etc. Describe to us this experience. (200 words max)

7. Tell us about a moment in your life when you showed tenacity and perseverance. (200 words max)

Arriving at this question is in itself a great achievement, but it’s not enough. Don’t give up!

8. Tell us an achievement, a project, or an initiative that you led. What role did you play? Did you mobilize other people around the project? (200 words max)

9. How is [the program] an adventure for you? Why do you insist on riding the train? What do you expect from this experience? (300 words max)

10. In three words, what does [the program] represent for you? (3 words, a bit of a break)

11. To what extent is it important for you to share with others what you will have learned and lived during the tour? (200 words max)

12. What skills, knowledge, and passions are you going to bring to the 49 other participants and to the [the program] team? (100 words max)

13. What would you like us to remember of you in 100 years? (200 words max)

14. If you want to add any other information that you think is useful to communicate to us, now is the time! (100 words max)

C. Bonus Questions: Creative project

Communicate to us your enthusiasm for the idea of participating in the [the program] adventure!

Drawing, video, poem, painting, song, dance, viral buzz, model train of matches, ticket costume ...Let loose, you can express yourself in any way you want! This question will give bonus points to those who answer it, but it will not be discriminatory for those who do not respond.

APPENDIX D. Judges' criteria and items.

Weighting	Construct	Items ²⁶ (5-point Likert scale)
85 %	Capacity to dream (big)	<i>Item 1 (5.0 %): The candidate anticipates the needs of others, loves to help others, and is concerned about others.</i>
		<i>Item 2 (5.0 %): The candidate anticipates the needs of others, senses other's wishes, feels other's emotions. S/he is concerned about others and takes time out for others.</i>
		<i>Item 3 (7.5 %): The candidate asks questions that nobody else does. The candidate has a vivid imagination.</i>
		<i>Item 4 (7.5 %): The candidate has a broad outlook on what is going on and has an excellent view of the world. The candidate has an exciting and optimistic vision of the future.</i>
	Willingness to change society	<i>Item 5 (7.5 %): The candidate has taken frequent stands in the face of strong opposition. S/he does not hesitate to express an unpopular opinion.</i>
		<i>Item 6 (7.5 %): The candidate loves dangerous situation. S/he takes risks and knows how to get around the rules.</i>
		<i>Item 7 (5.0 %): The candidate works hard to turn plans into action. The candidate does more than what's expected of him/her and sets high standards for herself/himself.</i>
		<i>Item 8 (5.0 %): The candidate does not quit a task before it is finished, is a goal-oriented person, and finishes things despite obstacles in the way.</i>
		<i>Item 9 (5.0 %): The candidate likes to begin new things. S/he is, open to change.</i>
		<i>Item 10 (5.0 %): The candidate loves excitement, loves action, and seeks adventure.</i>
Leadership potential	<i>Item 11 (5.0 %): The candidate does not care what others think. S/he sails his/her own course.</i>	
	<i>Item 12 (5.0 %): The candidate knows that his/her ideas sometimes surprise people. S/he swims against the current.</i>	
Ability to communicate	<i>Item 13 (7.5 %): The candidate takes charge and knows how to captivate people.</i>	
	<i>Item 14 (7.5 %): The candidate feels comfortable around people. S/he does not mind being the center of attention.</i>	
15%	Rater's individual impression of candidate outside of four criteria above	<i>Item 15 (7.5 %): The candidate radiates joy, has a lot of fun, and amuses his/her friends.</i>
		<i>Item 16 (7.5 %): The candidate makes friends easily and is skilled in handling social situations, knows how to captivate people.</i>
		<i>Four items developed by the organization.</i>
		<i>Item 17: Do you think that the candidate's path of life is unique and a reason to admire him/her?</i>
		<i>Item 18: In your view, does the candidate show qualities to be a change agent that s/he supports by convincing examples? Do you think that s/he is a mature candidate?</i>
		<i>Item 19: Do you think that the candidate will play a significant role among the 50 participants that will be selected?</i>
		<i>Item 20: Do you think that the candidate has an inspiring understanding of what life is about?</i>

²⁶ Items except 17-20 are sourced from the *International Personality Item Pool (IPIP)* (<http://ipip.ori.org/>).

APPENDIX E. List of 22 entrepreneurial actions taken to commercialize a venture.

1. Do you have a concrete idea for a product or service that you would like to sell?
2. Have you already begun the preparation of a business plan for a new business? (Note that a business plan usually outlines the markets to be served, the products or services to be provided, the resources required – including money – and the expected growth and profit for the new business)
3. Have you already sent a formally written business plan to other people?
4. Have you tested a product or service that your new business will be selling with potential customers?
5. Have marketing or promotional efforts been started for the product or service that your new business will be selling?
6. Have you developed any proprietary technology, processes, or procedures that no other company can use?
7. Have you submitted an application for a patent, copyright, or trademark relevant to your new business?
8. Have any major items like equipment, facilities, or property been purchased, leased, or rented specifically for your new business?
9. Have you made an effort to talk with potential customers about the product or service of your new business?
10. Have you made an effort to collect information about the competitors of your new business?
11. Have you made an effort to define the market opportunities for your new business?
12. Have you developed financial projections, such as income or cash flow statements or break-even analyses?
13. Have you made an effort to determine the regulatory requirements for your new business, such as operating licenses, permits, or health and safety regulations?
14. Have you asked financial institutions or other people for funds for your new business?
15. Have you received the first outside funding from financial institutions or other people for your new business?
16. Did you hire any managers or employees, including exclusive subcontractors, now working for pay (not people who share ownership)?
17. Have you already opened a bank account to use exclusively for your new business?
18. Has your new business already received any money, income, or fees from the sale of goods or services?
19. Has monthly revenue ever exceeded monthly expenses for your new business?
20. Can potential customers contact your new business by phone, through e-mail or a website on the internet, or by both phone and through the internet?
21. For your new business, have any payments been made to the federal social security system?
22. Has a federal income tax return ever been filed for your new business, whether or not it reported a profit and tax payments?

APPENDIX F. Various scales and measures.

Scale for social entrepreneurial actions. (Yes/No answers for each item)

1. Have you acted as founding member in the development of a social business?
2. Have you written down the social mission of your business idea?
3. Have you developed a tool to measure and communicate the social impact of your business idea?
4. Do you have a concrete idea to solve a specific social or environmental problem?
5. Have you received an award for a social business idea?

Scale for Sustainable Behavior. (Yes/No answers for each item)

1. I have worked for a social or non-profit organization (WITH pay).
2. I have volunteered for a social or non-profit organization (WITHOUT pay).
3. I have donated money to charitable organizations.
4. I have actively supported social causes (activism).
5. I have convinced others to change their behavior towards a more sustainable lifestyle.
6. I have systematically recycled waste in my daily life.
7. I have used bicycle or public transportation instead of mine or someone's car.
8. I have bought local and/or fair-trade products to replace what I normally buy.
9. I have published an item to make people aware of today's global challenges (e.g., posted picture, wrote blog, published article).
10. I have systematically avoided products with too much packaging.
11. I have reduced consumption of animal products that stem from mass production.

Scales for Traditional Career Intentions (items 1-6) and Social Career Intentions (items 7-10)

When you think about your first job or the next job that you would like to have, how important is the following for you? Please tick the appropriate number in the scale below.

1. Profitability of the company
2. Economic growth of the company
3. To have prestige
4. To have a high salary
5. Long term career perspective within the company
6. Corporate brand image/company reputation
7. To be dedicated to a cause or to feel that I am serving a greater good
8. To serve a social mission
9. To make a positive contribution to society
10. Social mission of the company

APPENDIX G. Supplementary Statistical Analysis

Table G1 *Sampling strata, number of applicants, top 100 and sampling numbers, 2014*

		No Adversity		Adversity		Total
		Not elite	Elite	Not elite	Elite	
Non Business	Male	84 / 13 / 16	15 / 5 / 8	21 / 6 / 6	1 / 0 / 0	121 / 24 / 30
	Female	86 / 24 / 20	17 / 8 / 4	17 / 6 / 4	6 / 3 / 4	126 / 41 / 32
Business	Male	44 / 6 / 10	6 / 2 / 2	11 / 1 / 2	3 / 2 / 2	64 / 11 / 16
	Female	53 / 13 / 10	14 / 4 / 4	15 / 4 / 4	4 / 3 / 4	86 / 24 / 22
Total		267 / 56 / 56	52 / 19 / 18	64 / 17 / 16	14 / 8 / 10	397 / 100 / 100

The numbers are given as: number of applicants / number of individuals in the original Top-100 ranked / number of individuals sampled. There can be either oversampling or undersampling in a stratum compared to Top 100 applicants based on the organization's preference for, for example, equal gender balance. If oversampling, additional applicants within that stratum ranked below the Top 100 were added. Elite = elite school in France. Adversity = people that had struggled with adversities (race, gender or other types of discrimination or adversity.) Business = graduate with a business-related degree.

Table G2. *Sampling strata, mid50 ranked, and sampling numbers, 2015*

		Adversity	No Adversity	Total
Non Business	Male	6 / 5	6 / 6	12 / 11
	Female	11 / 11	5 / 4	16 / 15
Business	Male	9 / 9	1 / 0	10 / 9
	Female	8 / 10	4 / 5	12 / 15
Total		34 / 35	16 / 15	50 / 50

The numbers are given as: the number of people in the stratum that were in the original Mid-50 by test score / the number of people in the stratum that were sampled in the final Mid-50. There can be either oversampling or undersampling in a stratum compared to Mid-50 applicants based on the organization's preference for, for example, equal gender balance. If oversampling, additional applicants within that stratum ranked below the Mid-50 were added. Adversity = people that had struggled with adversities (race, gender or other types of discrimination or adversity.) Business = graduate with a business-related degree. In the case of uneven cell counts, allocation to treatment and control could not be exactly matched. By chance, 27 were allocated to treatment and 23 to control. One of the 27 in the treatment group did not answer at t=1, and in some regressions one therefore observes 26 observations for the treatment group instead of 25.

Table G3 Pre-treatment differences, 2014 and 2015 cohorts

Measures	2014			2015			Mean	Mean	Difference
	Treatment Group	Control Group	Difference	Treatment Group	Control Group	Difference	2014	2015	2014-2015
	(std. err.)	(std. err.)	[p-value]	(std. err.)	(std. err.)	[p-value]	(std. err.)	(std. err.)	[p-value]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Entrepreneurial actions									
Social Entrepreneurial Actions	0.89 (0.14)	0.82 (0.21)	0.07 [0.79]	1.07 (0.18)	1.60 (0.23)	-0.52 [0.08]	0.86 (0.12)	1.26 (0.15)	-0.39 [0.04]*
Traditional Entrepreneurial Actions	2.36 (0.53)	3.74 (0.96)	-1.38 [0.22]	3.65 (0.69)	4.26 (1.03)	-0.61 [0.61]	2.88 (0.49)	3.87 (0.57)	-0.99 [0.19]
Sustainable behavior	7.36 (0.38)	7.47 (0.46)	-0.11 [0.85]	6.84 (0.33)	7.86 (0.58)	-1.02 [0.11]	7.96 (0.25)	7.21 (0.31)	0.75 [0.06]
Leadership skills									
MTL affective identity	3.56 (0.08)	3.56 (0.12)	0.00 [0.99]	3.60 (0.07)	3.43 (0.09)	0.16 [0.17]	3.56 (0.07)	3.54 (0.06)	0.02 [0.84]
MTL non-calculative	4.07 (0.07)	3.99 (0.15)	0.08 [0.63]	4.22 (0.09)	4.08 (0.11)	0.14 [0.35]	4.04 (0.07)	4.17 (0.07)	-0.13 [0.24]
MTL social-normative	3.33 (0.09)	3.35 (0.13)	-0.02 [0.95]	3.34 (0.10)	3.05 (0.12)	0.28 [0.09]	3.34 (0.07)	3.24 (0.08)	0.10 [0.37]
PLS vision	4.11 (0.07)	4.01 (0.09)	0.10 [0.45]	4.21 (0.09)	3.98 (0.10)	0.23 [0.11]	4.07 (0.06)	4.13 (0.07)	-0.05 [0.54]
PLS role model	3.79 (0.08)	3.72 (0.12)	0.07 [0.62]	3.75 (0.12)	3.88 (0.12)	-0.13 [0.48]	3.77 (0.07)	3.80 (0.09)	-0.03 [0.76]
PLS group goals	4.35 (0.08)	4.42 (0.10)	-0.07 [0.61]	4.55 (0.09)	4.45 (0.13)	0.10 [0.52]	4.38 (0.06)	4.52 (0.08)	-0.14 [0.18]
Intentions									
Traditional career intentions	5.19 (0.19)	5.14 (0.24)	0.05 [0.86]	5.82 (0.22)	5.94 (0.31)	-0.12 [0.75]	5.17 (0.15)	5.86 (0.18)	-0.69 [0.00]**
Social career intentions	8.12 (0.19)	7.51 (0.38)	0.61 [0.16]	8.27 (0.14)	8.08 (0.44)	0.19 [0.61]	7.89 (0.19)	8.20 (0.18)	-0.31 [0.26]
Entrepreneurial intentions	5.76 (0.24)	5.82 (0.29)	-0.06 [0.87]	6.23 (0.32)	6.00 (0.54)	0.23 [0.69]	5.78 (0.18)	6.14 (0.28)	-0.36 [0.27]
Sampling variables									
Elite business school (elite=1)	0.29 (0.07)	0.21 (0.09)	0.08 [0.53]	n.a.	n.a.	n.a.	0.26 (0.05)	n.a.	n.a.
Gender (male=1)	0.47 (0.08)	0.47 (0.10)	0.00 [0.13]	0.46 (0.09)	0.40 (0.13)	0.06 [0.71]	0.47 (0.06)	0.44 (0.08)	0.04 [0.72]
Adversity=1	0.24 (0.07)	0.17 (0.08)	0.07 [0.55]	0.31 (0.09)	0.26 (0.12)	0.04 [0.78]	0.21 (0.05)	0.29 (0.07)	-0.08 [0.36]
Education=business	0.37 (0.08)	0.39 (0.10)	-0.02 [0.86]	0.46 (0.09)	0.60 (0.13)	-0.14 [0.40]	0.37 (0.06)	0.51 (0.08)	-0.13 [0.18]
Suitability score	4.09 (0.04)	4.09 (0.05)	-0.00 [0.97]	73.15 (1.36)	71.91(1.98)	1.23 [0.60]	4.09 (0.03)	72.70 (1.12)	n.a.
Demographics									
Age (years)	26.78 (0.45)	26.30 (0.46)	0.48 [0.48]	25.23 (0.60)	25.06(0.86)	0.16 [0.88]	26.60 (0.33)	25.17 (0.48)	1.43 [0.02]*
Marital status (single=1)	0.73 (0.07)	0.84 (0.06)	0.10 [0.34]	0.77 (0.08)	0.66 (0.12)	0.10 [0.49]	0.80 (0.05)	0.73 (0.07)	0.07 [0.40]
Student=1	0.24 (0.07)	0.39 (0.10)	-0.15 [0.22]	0.11 (0.06)	0.20 (0.10)	0.08 [0.47]	0.29 (0.06)	0.14 (0.05)	0.15 [0.08]
Full-time employed	0.29 (0.07)	0.17 (0.08)	0.12 [0.29]	0.11 (0.06)	0.33 (0.12)	-0.21 [0.09]	0.24 (0.05)	0.19 (0.06)	0.05 [0.55]
Ever self-employed	0.31 (0.07)	0.26 (0.09)	0.05 [0.65]	0.11 (0.06)	0.20 (0.10)	-0.08 [0.47]	0.29 (0.06)	0.14 (0.07)	0.15 [0.08]
Ever started business with employees	0.05 (0.04)	0.04 (0.04)	0.01 [0.87]	0.08 (0.05)	0.00 (0.00)	0.08 [0.28]	0.05 (0.03)	0.05 (0.03)	0.00 [0.99]
Have had different kinds of jobs	0.37 (0.07)	0.56 (0.11)	-0.19 [0.14]	0.69 (0.09)	0.52 (0.13)	0.15 [0.32]	0.44 (0.06)	0.63 (0.07)	-0.19 [0.06]
Father had different kinds of jobs	0.16 (0.06)	0.34 (0.10)	-0.19 [0.09]	0.15 (0.07)	0.66 (0.12)	-0.51 [0.00]***	0.23 (0.05)	0.34 (0.07)	-0.11 [0.22]
Mother had different kinds of jobs	0.21 (0.07)	0.13 (0.07)	0.08 [0.44]	0.23 (0.08)	0.33 (0.12)	-0.10 [0.48]	0.18 (0.05)	0.27 (0.07)	-0.08 [0.29]
Parents total income last year (More than €150K=1)	0.05 (0.04)	0.17 (0.08)	-0.12 [0.18]	0.04 (0.04)	0.00 (0.00)	0.04 [0.45]	0.09 (0.04)	0.02 (0.02)	0.07 [0.15]
Number of observations	38	23	61	26	15	41	61	41	102

Note. The number of observations in columns 1, 2, 4, 5, 7, and 8 are total number of responses with repeated data at t=0 and t=1. p-values in columns 3, 6 and 9 are italicized and put within brackets to visually separate them better from the standard errors which are within parenthesis in columns 1, 2, 4, 5, 7, and 8. Note. *** <0.001, ** <0.01, * <0.05, two-sided. New business creation was not measured at t=0. Instead a slightly different questions was asked regarding whether the individual had ever been self-employed and/or started a new business. Data for this variable reported under the heading "sampling variables" in this table. Elite business school status was not reported by the organization for 2015 as it was not used for sampling. The suitability score was also scaled differently in 2015.

Table G4 *Probit Regression Results on Probability of Response*

Measures	2014	2015
Treatment	0.44 (0.15)**	0.78 (0.30)*
Gender	-0.18 (0.16)	0.28 (0.28)
Age	0.70 (0.19)***	-0.18 (0.27)
Business	-0.06 (0.15)	-
Diversity	-0.03 (0.15)	0.48 (0.28)
Suitability score	-0.61 (0.18)**	0.12 (0.28)
Constant	0.49 (0.16)**	1.38 (0.33)***
Pseudo R2	0.29	0.28
<i>N</i>	100	50

Note. *** <0.001, ** <0.01, * <0.05, two-sided.
Standard errors are reported in parentheses. Variables standardized.

Table G5 Entrepreneurial Behaviour, Leadership skills and Social Identity, 2014 program

Measures	OLS, no controls		OLS, with controls	
	(1) δ	(2) β	(3) δ	(4) β
Traditional Entrepreneurial Behavior				
Entrepreneurial Actions	0.18 (0.22)	0.36 (0.10)**	0.16 (0.22)	0.34 (0.11)**
New Business Creation	-0.02 (0.148)	n.a.	-0.05 (0.153)	n.a.
Social Leadership Skills				
MTL affective identity	0.01 (0.08)	-0.24 (0.08)*	-0.00 (0.08)	-0.24 (0.09)*
MTL non-calculative	0.03 (0.09)	-0.31 (0.07)***	0.05 (0.09)	-0.31 (0.06)***
MTL social-normative	-0.01 (0.08)	-0.43 (0.14)**	-0.02 (0.08)	-0.43 (0.14)**
TLS vision	-0.07 (0.15)	-0.41 (0.10)**	-0.09 (0.14)	-0.42 (0.10)**
TLS role model	0.04 (0.11)	-0.59 (0.15)**	0.07 (0.12)	-0.60 (0.15)**
TLS group goals	0.06 (0.14)	-0.46 (0.10)***	0.05 (0.14)	-0.52 (0.13)**
Pro-Social Actions and Identity				
Social Entr. Actions	0.17 (0.31)	0.44 (0.17)*	0.13 (0.33)	0.42 (0.18)*
Sustainable Behavior	-0.12 (0.19)	0.53 (0.09)***	-0.15 (0.21)	0.54 (0.10)***
Social career intentions	0.16 (0.09)	-0.76 (0.18)***	0.11 (0.10)	-0.79 (0.17)***
Traditional career intentions	0.01 (0.08)	-0.42 (0.20)*	0.01 (0.08)	-0.42 (0.21)*
<i>N</i>	61		61	

Note. *** <0.001, ** <0.01, * <0.05, two-sided. Standard errors are clustered by stratum and reported in parentheses. Variables standardized. Individual without repeated values at t=0 and t=1 omitted. N=61 in all regressions. Estimates are from separate regressions for each row. Coefficient estimates for the control variables and constants are suppressed, but available from the authors on request. Treatment effects reported in cols 1 and 3. Coefficient for lagged outcome (y_0) reported in cols 2 and 4. Significant controls included: gender and student. Regression NOT weighted with inverse of sampling and non-response probability. MTL=Motivation To Lead, TLS=Transformational Leadership Style. n.a. = Not Available.

Table G6 *Entrepreneurial Behaviour, Leadership skills and Social Identity, 2015 program*

Measures	OLS, no controls		OLS, with controls	
	(1) δ	(2) β	(3) δ	(4) β
Traditional Entrepreneurial Behavior				
Entrepreneurial Actions	0.68 (0.20)**	0.52 (0.13)**	0.70 (0.20)**	0.59 (0.13)**
New Business Creation	0.33 (0.09)**	n.a.	0.34 (0.09)**	n.a.
Social Leadership Skills				
MTL affective identity	0.01 (0.10)	-0.24 (0.14)	-0.04 (0.19)	-0.20 (0.16)
MTL non-calculative	-0.03 (0.18)	-0.39 (0.21)	-0.04 (0.19)	-0.37 (0.20)
MTL social-normative	0.06 (0.15)	-0.55 (0.09)**	0.04 (0.16)	-0.52 (0.09)**
TLS vision	0.08 (0.18)	-0.72 (0.21)**	0.08 (0.19)	-0.73 (0.22)**
TLS role model	0.06 (0.14)	-0.76 (0.17)**	0.08 (0.14)	-0.73 (0.17)**
TLS group goals	0.30 (0.19)	-0.43 (0.26)	0.32 (0.19)	-0.46 (0.27)
Pro-Social Actions and Identity				
Social Entr. Actions	0.23 (0.19)	0.27 (0.11)*	0.26 (0.19)	0.27 (0.11)*
Sustainable Behavior	0.00 (0.26)	0.58 (0.16)**	0.02 (0.27)	0.62 (0.16)**
Social career intentions	-0.08 (0.10)	-0.78 (0.08)**	-0.08 (0.10)	-0.77 (0.08)**
Traditional career intentions	0.20 (0.21)	-0.61 (0.16)**	0.22 (0.20)	-0.61 (0.15)**
<i>N</i>	61		61	

Note. *** <0.001, ** <0.01, * <0.05, two-sided. Standard errors are clustered by stratum and reported in parentheses. Variables standardized. Individual without repeated values at t=0 and t=1 omitted. N=61 in all regressions. Estimates are from separate regressions for each row. Coefficient estimates for the control variables and constants are suppressed, but available from the authors on request. Treatment effects reported in cols 1 and 3. Coefficient for lagged outcome (y_0) reported in cols 2 and 4. Significant controls included: gender and student. Regression NOT weighted with inverse of sampling and non-response probability. MTL=Motivation To Lead, TLS=Transformational Leadership Style. n.a. Not Available.

Note on computing stratum-level treatment effects for Table G7 and G8 (Athey and Imbens, 2016, page 31)

To compute the treatment effect, define the first difference in each outcome variable by each sampling stratum g Δy_g . Δy_g is measured by the average change in the score of each construct between $t=0$ and $t=1$ for strata g ($\Delta \bar{y}_g = (\sum_i (y_{i1} - y_{i0}) / N_g$). Add notation t for treatment and c for control group. Within each stratum we can then estimate the average treatment effect as the difference in average outcomes for treated and control units, $\hat{t}_g = \Delta \bar{y}_{t,g} - \Delta \bar{y}_{c,g}$ and the within-stratum variance can be estimated as $\hat{V}(\hat{t}_g) = s^2_{t,g} / N_{t,g} + s^2_{c,g} / N_{c,g}$. According to Athey and Imbens (2016, page 31), we can then most efficiently estimate the average treatment effect given our sampling design by averaging the within-stratum estimates weighted by the stratum share N_g/N , such that $\hat{t} = \sum_{g=1}^G \hat{t}_g \frac{N_g}{N}$ with estimated variance $\hat{V}(\hat{t}) = \sum_{g=1}^G \hat{V}(\hat{t}_g) (\frac{N_g}{N})^2$.

We report weighted averages of the response at $t=1$: $\sum_g (\sum (y_{i1}) / N_g) (N_g / N)$, and at $t=0$: $\sum_g (\sum (y_{i0}) / N_g) (N_g / N)$, and then the averages of the first differences for the treatment and control groups: $\Delta \bar{y}_t$ and $\Delta \bar{y}_c$. Columns 1-6 reports these sample data computed at the cluster level and then weighted as above. The overall treatment effect \hat{t} is reported in column 7. Statistically significant p-values for the treatment effect are indicated using asterisks.

Note that we form the difference $y_{i1} - y_{i0}$ assuming that the actions taken between $t=1$ and $t=0$ represents a level reached at $t=1$, and thus that $\hat{t}_g = \Delta \bar{y}_{t,g} - \Delta \bar{y}_{c,g}$. However, New Business creation was not measured at $t=0$ and so the difference-in-difference between treatment and control reflects $\hat{t}_g = \bar{y}_{1,t,g} - \bar{y}_{1,c,g}$.

Table G7 Stratum-level computed Means, Differences in Means Between Treatment and Control Groups and Treatment Effect, 2014

Measures	Treatment		Diff	Control		Diff	Treatment Effect
	(1) t=0	(2) t=1	(3) Δy_t	(4) t=0	(5) t=1	(6) Δy_c	(7) $\hat{\tau}$
Traditional Entrepreneurial Actions							
Entrepreneurial Actions	2.36 (0.73)	3.81 (0.97)	1.44 (0.60)*	3.74 (1.37)	3.56 (0.92)	-0.17 (0.64)	1.62 (0.49)**
New Business	n.a.	0.50 (0.07)	n.a.	n.a.	0.52 (0.08)	n.a.	-0.02 (0.07)
Social Leadership skills							
MTL affective identity	3.56 (0.11)	3.43 (0.11)	-0.13 (0.13)	3.56 (0.15)	3.42 (0.12)	-0.14 (0.10)	0.01 (0.07)
MTL non-calculative	4.07 (0.10)	4.02 (0.10)	-0.05 (0.10)	3.99 (0.18)	3.93 (0.15)	-0.06 (0.21)	0.01 (0.06)
MTL social-normative	3.34 (0.11)	3.22 (0.14)	-0.11 (0.09)	3.35 (0.12)	3.24 (0.13)	-0.10 (0.11)	-0.01 (0.08)
TLS vision	4.11 (0.11)	3.91 (0.13)	-0.20 (0.09)*	4.01 (0.12)	3.93 (0.06)	-0.08 (0.06)	-0.12 (0.06)
TLS role model	3.79 (0.09)	3.85 (0.11)	0.05 (0.06)	3.72 (0.13)	3.76 (0.16)	0.04 (0.09)	0.01 (0.13)
TLS group goals	4.35 (0.11)	4.37 (0.09)	0.01 (0.06)	4.42 (0.13)	4.33 (0.09)	-0.09 (0.08)	0.10 (0.08)
Pro-Social Actions and Identity							
Social Entrepreneurial Actions	0.89 (0.19)	1.71 (0.32)	0.81 (0.19)***	0.82 (0.31)	1.43 (0.32)	0.60 (0.18)**	0.21 (0.25)
Sustainable Behavior	8.05 (0.49)	8.07 (0.36)	0.03 (0.31)	7.82 (0.35)	8.21 (0.37)	0.40 (0.33)	-0.36 (0.23)
Social career intentions	8.12 (0.25)	8.07 (0.17)	-0.05 (0.12)	7.51 (0.43)	7.48 (0.35)	-0.02 (0.38)	-0.02 (0.22)
Traditional career intentions	5.19 (0.27)	5.17 (0.33)	-0.02 (0.17)	5.14 (0.39)	5.11 (0.46)	-0.03 (0.20)	0.01 (0.14)
N	38	38		23	23		61

Note. *** <0.001, ** <0.01, * <0.05. Standard errors in parentheses. p-values calculated using two-sided tests. Individual without repeated values at t=0 and t=1 removed. The treatment effect is measured as $\hat{\tau}_g = \Delta \bar{y}_{t,g} - \Delta \bar{y}_{c,g}$ for each sampling stratum g using Athey and Imbens' method (2016, page 31). However, for New Business creation the treatment effect is measured as $\hat{\tau}_g = \bar{y}_{1,t,g} - \bar{y}_{1,c,g}$.

Table G8 Stratum-level computed Means, Differences in Means Between Treatment and Control Groups and Treatment Effect, 2015

Measures	Treatment		Diff	Control		Diff	Treatment Effect
	(1) t=0	(2) t=1	(3) Δ_{yt}	(4) t=0	(5) t=1	(6) Δ_{yc}	(7) $\hat{\tau}$
Traditional Entrepreneurial actions							
Entrepreneurial Actions	3.65 (0.93)	7.46 (1.17)	3.81 (0.59)***	4.26 (0.99)	4.33 (1.85)	0.06 (1.61)	3.74 (1.06)**
New Business	n.a.	0.73 (0.11)	n.a.	n.a.	0.40 (0.16)	n.a.	0.33 (0.12)*
Leadership skills							
MTL affective identity	3.61 (0.07)	3.39 (0.15)	-0.21 (0.07)**	3.43 (0.05)	3.25 (0.13)	-0.18 (0.06)**	-0.03 (0.08)
MTL non-calculative	4.22 (0.10)	3.84 (0.16)	-0.38 (0.11)**	4.08 (0.12)	3.80 (0.10)	-0.28 (0.07)***	-0.09 (0.12)
MTL social-normative	3.34 (0.07)	3.21 (0.11)	-0.13 (0.07)	3.05 (0.16)	3.00 (0.13)	-0.05 (0.08)	-0.08 (0.12)
TLS vision	4.21 (0.12)	3.94 (0.12)	-0.27 (0.06)***	3.98 (0.08)	3.80 (0.14)	-0.18 (0.08)*	-0.08 (0.11)
TLS role model	3.75 (0.11)	3.76 (0.12)	0.01 (0.08)	3.88 (0.13)	3.73 (0.14)	-0.15 (0.10)	0.16 (0.14)
TLS group goals	4.55 (0.13)	4.35 (0.14)	-0.20 (0.08)*	4.45 (0.16)	3.90 (0.21)	-0.55 (0.11)***	0.35 (0.11)**
Pro-Social Actions and Identity							
Social Entrepreneurial Actions	1.07 (0.25)	1.38 (0.27)	0.31 (0.16)	1.60 (0.25)	1.26 (0.34)	-0.33 (0.23)	0.64 (0.19)**
Sustainable Behavior	6.84 (0.42)	6.96 (0.51)	0.11 (0.28)	7.86 (0.76)	7.60 (0.79)	-0.27 (0.51)	0.38 (0.35)
Social career intentions	8.27 (0.19)	8.01 (0.17)	-0.27 (0.11)*	8.08 (0.56)	8.15 (0.45)	0.06 (0.36)	-0.33 (0.27)
Traditional career intentions	5.82 (0.24)	5.72 (0.32)	-0.11 (0.18)	5.94 (0.35)	5.23 (0.31)	-0.71 (0.27)*	0.60 (0.33)
N	26	26		15	15		41

Note. *** <0.001, ** <0.01, * <0.05. Standard errors in parentheses. p-values calculated using two-sided tests. Individual without repeated values at t=0 and t=1 removed. The treatment effect is measured as $\hat{\tau}_g = \Delta \bar{y}_{t,g} - \Delta \bar{y}_{c,g}$ for each sampling stratum g using Athey and Imbens' method (2016, page 31). However, for New Business creation the treatment effect is measured as $\hat{\tau}_g = \bar{y}_{1,t,g} - \bar{y}_{1,c,g}$.

Table G9 Means of entrepreneurial actions of responding treatment and control groups, 2015 program

Measures	Treatment Group	Control Group	Difference (1)-(2)
	(std. err.)	(std. err.)	(std. err.)
	(1)	(2)	(3)
Traditional Entrepreneurial Actions			
Have a concrete idea for a product or service to sell	0.65 (0.11)	0.53 (0.16)	0.12 (0.08)
Have begun the preparation of a business plan	0.57 (0.11)	0.33 (0.15)	0.24 (0.08)**
Have sent a formally written business plan to other people	0.35 (0.11)	0.13 (0.07)	0.21 (0.08)*
Have tested your product or service that you want to sell	0.50 (0.11)	0.20 (0.14)	0.30 (0.07)***
Have started marketing or promotional efforts	0.35 (0.09)	0.20 (0.11)	0.15 (0.07)*
Have developed any proprietary technology or processes	0.23 (0.09)	0.20 (0.09)	0.03 (0.07)
Have submitted an application for a patent or copyright	0.50 (0.10)	0.20 (0.15)	0.30 (0.07)***
Have purchased or rented equipment, facilities, or property	0.42 (0.10)	0.20 (0.15)	0.22 (0.08)**
Have talked with potential customers	0.11 (0.04)	0.00 (0.00)	0.11 (0.03)*
Have collected information about competition	0.15 (0.06)	0.20 (0.11)	-0.05 (0.06)
Have defined the market opportunities	0.69 (0.11)	0.40 (0.16)	0.29 (0.08)**
Have developed financial projections or break-even analyses	0.73 (0.11)	0.26 (0.14)	0.46 (0.08)***
Have determined the regulatory requirements for the new business	0.07 (0.06)	0.13 (0.07)	-0.05 (0.08)
Have asked financial institutions or other people for funds	0.15 (0.09)	0.13 (0.11)	0.02 (0.07)
Have received the first outside funding from financial institutions or other people	0.11 (0.04)	0.06 (0.07)	0.05 (0.06)
Hired any managers or employees, or exclusive subcontractors, working for pay	0.15 (0.08)	0.06 (0.07)	0.09 (0.07)
Have opened a bank account to use exclusively for your new business	0.19 (0.09)	0.13 (0.11)	0.06 (0.07)
Have received any income from the sale of goods or services	0.04 (0.05)	0.13 (0.11)	-0.09 (0.06)
Monthly revenue ever exceeded monthly expenses for the new business	0.57 (0.09)	0.33 (0.16)	0.24 (0.08)**
Potential customers can contact you by phone, through e-mail or a website	0.53 (0.12)	0.20 (0.13)	0.33 (0.08)***
Made payments to the federal social security system for the new business	0.07 (0.05)	0.13 (0.11)	-0.06 (0.06)
Filed income tax return whether or not it reported a profit and tax payments	0.04 (0.00)	0.13 (0.11)	-0.09 (0.06)
Number of observations	26	15	41

Note. *** <0.001, ** <0.01, * <0.05. Standard errors are calculated using Athey and Imbens's (2016) stratum-level method. P-values are calculated using two-sided tests. Only candidates in the Mid50 group were included. Individual without repeated values at t=0 and t=1 are omitted.

Table G10 *Seemingly Unrelated Regression, 2014 cohort*

Measures	With controls, unweighted			
	(1) δ	(2) β	F-stat	R2
Traditional Entrepreneurial Behavior				
Entrepreneurial Actions	0.11 (0.16)	0.19** (0.07)	2.65**	0.74
New Business Creation	0.03 (0.12)	n.a.	0.20	0.57
Social Leadership Skills				
MTL affective identity	0.11 (0.09)	-0.27* (0.11)	1.33	0.42
MTL non-calculative	-0.01 (0.08)	-0.33*** (0.07)	4.98***	0.36
MTL social-normative	-0.06 (0.11)	-0.62*** (0.07)	9.84***	0.49
TLS vision	0.00 (0.08)	-0.59*** (0.08)	7.05***	0.35
TLS role model	0.07 (0.10)	-0.94*** (0.13)	6.67***	0.48
TLS group goals	0.03 (0.08)	-0.33* (0.16)	1.04	0.37
Pro-Social Actions and Identity				
Social Entr. Actions	0.10 (0.20)	0.41*** (0.06)	6.85***	0.95
Sustainable Behavior	0.01 (0.18)	-0.35*** (0.10)	2.52***	0.83
Traditional career intentions	-0.01 (0.17)	-0.06 (0.17)	0.36	0.82
Social career intentions	0.37* (0.17)	-0.64*** (0.06)	16.11***	0.78

Note. *** <0.001, ** <0.01, * <0.05, two-sided. Standard errors in parentheses. Variables standardized. Individual without repeated values at t=0 and t=1 removed. N=61 for each outcome. Number of parameters per outcome: 10. Significance based on t-statistic. Iterated ML estimation. As a divisor in computing the covariance matrix for the equation residuals we use a small-sample adjustment $\sqrt{((n - k_i) * (n - k_j))}$, where k_i and k_j are the number of parameters in equations i and j , respectively. Controls include gender, elite school, business school, suitability score, student, father had different kinds of jobs and parental total income.

Table G11 *Seemingly Unrelated Regression, 2015 cohort*

Measures	With controls, unweighted			
	(1) δ	(2) β	F-stat	R2
Traditional Entrepreneurial Behavior				
Entrepreneurial Actions	1.11 (0.40)**	-0.18 (0.11)	2.65**	0.74
New Business Creation	0.39 (0.21)+	n.a.	0.20	0.57
Social Leadership Skills				
MTL affective identity	0.21 (0.20)	-0.29 (0.21)	1.33	0.42
MTL non-calculative	0.05 (0.27)	-0.26 (0.19)	4.98***	0.36
MTL social-normative	0.09 (0.18)	-0.55 (0.12)***	9.84***	0.49
TLS vision	0.20 (0.19)	-0.59 (0.11)***	7.05***	0.35
TLS role model	-0.01 (0.25)	-0.82 (0.15)***	6.67***	0.48
TLS group goals	0.43 (0.24)	-0.35 (0.10)***	1.04	0.37
Pro-Social Actions and Identity				
Social Entr. Actions	0.61 (0.47)	-0.59 (0.12)***	6.85***	0.95
Sustainable Behavior	-0.50 (0.90)	-0.50 (0.14)***	2.52***	0.83
Traditional career intentions	0.46 (0.57)	-0.68 (0.17)***	0.36	0.82
Social career intentions	-0.17 (0.40)	-0.74 (0.09)***	16.11***	0.78

Note. *** <0.001, ** <0.01, * <0.05, two-sided. Standard errors in parentheses. Variables standardized. Individual without repeated values at t=0 and t=1 removed. N=61 for each outcome. Number of parameters per outcome: 10. Significance based on t-statistic. Iterated ML estimation. As a divisor in computing the covariance matrix for the equation residuals we use a small-sample adjustment $\sqrt{((n - k_i) * (n - k_j))}$, where k_i and k_j are the number of parameters in equations i and j , respectively. Controls include gender, elite school, business school, suitability score, student, father had different kinds of jobs and parental total income.

APPENDIX H

Robustness Analysis

We examine several threats to the validity of our inferences. These are: multiple hypothesis testing and p-fishing (p-hacking); and truncation bias. In summary we unearth low to moderate threats to the inferences made. We also discuss heterogeneous treatment effects and whether inclusion of the top-25 non-randomly selected in the 2015 cohort may have had an effect on the mid-25 randomly selected to treatment. Details of these examinations are found in this Appendix.

1. Heterogeneous treatment effects

We experiment with analysing heterogeneous treatment effects, some of which have already been mentioned. The program might be able to accelerate the efforts by those that are already more socially entrepreneurially oriented as they may be more receptive. Alternatively, the program has more of an effect on those who previously were less entrepreneurially oriented as the program could be ineffective in making much change for those who are already committed. However, few such detectable interaction effects are found, most likely because of the low power in these tests. We also looked at potential distributional effects using quantile regressions at the 25th, 50th, and 75th percentiles, but found no remarkable patterns worth reporting. Results are robust to using various estimation schemes. Results are available from the authors on request.

2. Multiple hypothesis testing and p-fishing (p-hacking)

An additional concern is that we are conducting multiple tests of the null hypothesis and that, by chance, we will obtain a false positive if we perform enough tests. We can correct for this using at least three approaches. First, in the paper we report results for three different methods of computing the treatment effect and call our results robust only

if a construct has a significant treatment effect across all three methods. Second, we could Bonferroni-adjust the cut-off value for significance with the number of tests conducted. The number of constructs tested in Table 3 is 12, and we would therefore require significance at $0.05/12 = 0.004$. Of course, that would render our already insignificant results even more insignificant, and the conclusion of no robust treatment effects for these variables would remain the same. Our results on the two traditional entrepreneurial activity measures are most often (in 3 out of 4 cases) strong enough to survive such Bonferroni-adjustments. A third way is to perform seemingly unrelated regressions (SUR). This takes into account potential common correlation across the outcome variables. If the common correlation is positive and substantial, then the regressions in Tables 2 and 3 would be upwardly biased, because, with correlated errors, we could find multiple significant results from independent tests that stem from just one or a few main effects. Appendix Table G10 and G11 provide examples of results running SUR instead. Results remain in the direction already indicated.

3. Truncation bias

A technical argument may be that the results are due to measurement truncation at the lowest (or highest) scale levels, rather than true behavioral patterns. However, most constructs are composite weighted measures of several items, which makes them regress toward the mean and are less likely to take extreme values close to the upper or lower scale limit. Finally, all scales except the dummy for new business are retransformed by subtracting the mean and dividing by the standard deviation, which means that natural truncation values are absent.

4. Non-random selection of top 25 to the 2015 cohort

Recall that for the 2015 cohort in the past we have only analyzed the Mid50 sample

which were those randomly allocated to treatment and control. Because we collect data from LinkedIn on (close to) all 100 top applicants, we need to make sure that the treatment effects for the 2015 cohort are not confounded by the organization's selection of the top 25 applicants into treatment. Using two dependent variables; founder or co-founder of a start-up since the end of the tour, and number of entrepreneurial actions, we perform WLS regressions on the 2015 cohort including a dummy for control+mid50, a dummy for treatment+mid50 and a dummy for selected into the top 25 treatment group. The baseline group are those in the bottom 25 control group. Results are presented in Table H1 below. The table reveals that there are no differences in the effect of the treatment on being a founder/co-founder of a start-up between those selected for treatment (top25), and those randomly allocated to treatment (half of those in the mid50 group). Regressions for the number of start-ups, and number of start-ups which are socially responsible produce similar results. Specifically, a t-test fails to reject the null of a difference between the coefficients for the two treatment groups in column 1 ($\beta_2 - \beta_3 = -0.06$, $t = -0.36$, $p < 0.73$) and in column 2 ($\beta_2 - \beta_3 = -2.01$, $t = -1.02$, $p < 0.32$). We therefore conclude that the analysis performed on the 2015 sample does not suffer from bias due to the non-random selection of 25 participants into the treatment group.

Table H1 *WLS Regression: Co-founder status and entrepreneurial actions, individual level, 2015 cohort*

	(1)	(2)
	Founder or co-founder of start-up	Number of entre- preneurial actions
Control+Mid50 (Yes=1)	0.04 (0.04)	4.78 (1.31)**
Treatment+Mid50 (Yes=1)	0.37 (0.15)*	7.09 (1.26)***
Treatment+Top25 selected (Yes=1)	0.43 (0.10)***	9.10 (1.72)***
R-Squared	0.30	0.53
N	93	81

Note. *** <0.001, ** <0.01, * <0.05, two-sided. Standard errors are clustered by stratum and reported in parentheses. Outcome in column 1 collected from LinkedIn and reflects founding a start-up within two years of end of program. Outcome in column 2 collected from post-program survey. The omitted category is bottom25 and control group. Observations weighted by inverse of sampling and response frequency.

APPENDIX I

Case Studies in Treatment Group

Sen Women Up

The project originally had two goals—women empowerment and sustainable cereal farming. In Senegal, over the past decade, gold mining has increased dramatically. People quit jobs and abandoned farming in the hopes of finding gold. A French and one local Senegalese founder wanted to resurrect farming at least on a small scale and decided to focus on sustainable cereal farming by women. The French founder worked full-time unpaid on location for the first two years. He then returned to France, obtained a full-time job and now supervises the project remotely. All else involved in the project are Senegalese. To build a factory they obtained a donation of 80,000 euros from a foundation. Ten women work and live at the factory. Their customers include luxury hotels in Dakar and local goldmines. One goldmine became their first customer, providing cereal for the lunches of the workers. In 2014, the startup won the Price Springboard for Young Solidarity by the NGO Santé Sud, Eureka and Simon Martin Consulting. It won 2nd place in Unilever Sustainable Living Young Entrepreneurs Awards 2015.

Ecole Hotel Solar

Launched as an extension of a hotel that has been active for 20 years. Two full-time partners run the existing hotel with seven employees. Ecole Hotel Solar aims to open new ecofriendly hotel locations in the coming years and has already expanded its activities into providing training on sustainable development practices in the tourism industry within the context of its new school. The operations of the project are currently funded with income from hotel and training activities as well as 20,000 euros in funding obtained by the projects' originator from Association France Active, which seeks to promote entrepreneurial activities in France. They provide training to hotel staff and tour operators and are also launching a pilot summer project where they will be training 15 teenagers who dropped out of school. Since the launch of the project, Ecole Hotel Solar applied for tenders to develop social and innovative programs or businesses in unused urban sites from governmental organisations, such as the City of Paris and Reinventer la Seine. Five hotel projects have already been accepted and construction is planned to start in 2020. Additional funding is being applied for.

Un monde reenchante

An association with the objective to conduct travel expeditions abroad. They are also considering teaching seminars to businesses. Their first expedition took place in October 2017 and lasted 15 days. 62 candidates applied and 10 were selected to travel to Morocco accompanied by 5 staff members. They used the same selection principles as the organization and also adapted the same process—in particular, the expedition was divided into three parts—inspiration (meeting with social entrepreneurs), introspection (working with coaches) and action (volunteering for local permaculture projects). The after-expedition stage included 2 weekends of workshops. A special evening took place where the participants pitched their transformation/journey to an audience of about 200 people. Their next expedition was planned for October 2018. They planned to go to the UK to see several pioneer villages and “garden towns”. The two founders are working on making their jobs full-time paid positions. Funding comes from 300 individuals who donate 10 euros

a month. In addition, they have engaged in “street fund-raising” where they wear t-shirts and tell people about their cause. They were able to raise 6,200 euros this way in order to subsidize the cost for the participants.

Ridy

The startup provides bicycle repair services at the location of the client and was launched in order to facilitate commuting by bicycle between home and work. Its main objectives include decreasing cars' carbon dioxide emissions in Paris and encouraging the city's residents to be more healthy. It is led by two co-founders, employs four bicycle mechanics and is in the process of hiring two additional team members. Having already performed repairs on more than 2,400 individual clients' bicycles, Ridy has also secured contracts to maintain the fleets of bicycles of 15 firms based in Paris, including Schneider Electric, Google and Ubisoft. It has also entered into a partnership with Paris' public transport authorities to offer its services in local train stations. In 2017, the co-founders won the Marathon Pitch Award granted by France's main business news journal, Les Echos. Ridy is seeking additional funding in order to expand its activities.