

# Skills development and training interventions in Africa: Findings, challenges, and opportunities<sup>†</sup>

*Bernd Beber, Regina Dworschak, Tabea Lakemann, Jann Lay, Jan Priebe\**

*June 2020*

## Introduction

### Motivation

Skills trainings that target individuals or small businesses feature prominently in the development assistance landscape. They are an important vehicle for both bilateral and multilateral economic support. One estimate puts the volume of skills training interventions facilitated by the World Bank alone at about a billion U.S. dollars per year (Blattman and Ralston 2015).

This briefing serves to summarize key insights that have so far been generated in the related research literature, identify study gaps that remain, discuss core challenges that compelling impact evaluations must grapple with, and outline one such set of evaluations that is planned to accompany Invest for Jobs.

### Focus of this review

The substantive focus of this briefing is on training and related support programs, in which funding is deployed to implementing organizations to directly assist beneficiaries. We exclude development cooperation efforts that are directed at meso- or macro-level conditions. Such conditions and related interventions, for example support to Foreign Direct Investment, are to some extent addressed in WP2 as part of this project.

Our main objective is to understand the effects of skills development and training programs on two groups of outcomes, which roughly correspond to primary core performance indicators of Invest for Jobs: First, employment, measured in terms of job retention, acquisition and/or lengths of employment spells; and second, job quality, including earnings and workplace conditions.<sup>1</sup>

Broadly speaking, trainings aim to achieve improvements in these outcomes by boosting one or more of the following (Kluve *et al.* 2017, p. 29):

1. Fundamental skills like reading, writing, and basic math;
2. Technical, specialized skills, such as carpentry;

---

<sup>†</sup> The authors gratefully acknowledge funding from Germany's Federal Ministry for Economic Cooperation and Development (BMZ). This briefing paper was written as part of the [RéUsSITE project](#), which provides research support to the Ministry's Special Initiative on Training and Job Creation (Invest for Jobs).

\* Beber: RWI - Leibniz Institute for Economic Research; Dworschak, Lakemann, Lay: German Institute for Global and Area Studies (GIGA) and University of Goettingen; Priebe: German Institute for Global and Area Studies (GIGA). Corresponding author: Bernd Beber, [bernd.beber@rwi-essen.de](mailto:bernd.beber@rwi-essen.de).

<sup>1</sup> These two broad types of outcomes—employment and job quality—are relevant for both person- and firm-level interventions. Firm-level studies frequently also consider other target outcomes, depending on firm size and training contents. Studies of microenterprises often consider productivity, investment, and sales growth as downstream outcomes, while employment growth becomes more relevant for larger firms. Studies of trainings that promote cost-saving strategies place more emphasis on efficiency, whereas studies of marketing trainings focus on sales growth.

3. Management techniques and business skills, such as principles of bookkeeping;
4. Soft skills, including problem-solving and teamwork techniques, etc.

Frequently, trainings are also combined with other measures such as job placements or capital injections (for small businesses or individuals, e.g. by way of a cash grant).

Methodologically, the briefing highlights rigorous impact evaluations, in particular randomized controlled trials (RCTs). Other types of studies, including qualitative work, are also occasionally referenced. Still, the high internal validity of RCTs means we place them center stage in our analysis.<sup>2</sup>

This briefing paper can build on a number of literature reviews on the impacts of training and skill development interventions (see Betcherman *et al.* 2007, Blattman and Ralston 2015, Card *et al.* 2015, Kluge *et al.* 2017, 2019).

## Principal insights

The evidence concerning the efficacy of training interventions, broadly conceived, is mixed. While some interventions appear to produce increases in earnings or the probability of employment, many do not. Integrated, multipronged interventions that might combine vocational training with a capital infusion generally appear to be more efficacious, as are those that involve private firms and those of superior program quality (e.g. high training intensity, responsiveness to beneficiary needs, consistent and dependable delivery of activities). Soft skills training alone does not appear to improve economic livelihoods. Overall effects are heterogeneous across contexts and beneficiary characteristics. Not surprisingly, there is no silver bullet that addresses training needs everywhere.

A key reason for this is that program effects hinge on market demand for the intervention's target skills, which complicates theories of change for training interventions. An improvement in the laborer's economic livelihood requires not only an effective transfer of skills, but also a market for upskilled labor. In contexts where formal employment opportunities are scarce and labor markets segmented this cannot be taken for granted. In turn this means that programs need to correctly anticipate market needs or perhaps combine a training intervention with a meso- or macro-level market-creating intervention. Either way, these are no easy tasks, especially in generally data-poor developing economies (see e.g. Blattman and Ralston 2015). The importance of labor demand may partly explain why integrated, multifaceted interventions appear to perform better, and why the available empirical evidence suggests significant effect heterogeneity and context dependency.

This points to key research gaps identified in this briefing. We note that comparatively few training programs have been rigorously evaluated in developing countries, especially low- and lower-middle income countries including in Africa, which means it is unclear how well training programs are able to meet market-specific needs in these contexts. The role of private sector involvement, often considered a driver of training efficacy, remains understudied in less developed countries, as are the effects of trainings on comprehensive and contextually sensitive sets of job quality indicators. Given the central role that program funders sometimes ascribe to well-run private firms as training implementation partners, firm-level interventions are surprisingly rare not only in Africa, but in general. We also note that while the literature on training programs is sufficiently mature to have generated a number of well-identified causal effects, some important aspects call for further study, including the causal pathways undergirding these effects, their endurance or dissipation over time, the cost effectiveness of training programs, and how processes of participant recruitment and selection affect program outcomes.

Contemporary research on these questions should strive to meet leading-edge methodological standards. Most importantly this means that effect estimates reflect well-justified counterfactual

---

<sup>2</sup> Not all rigorous impact evaluations are RCTs. Other techniques, such as difference-in-differences estimations or instrumental variable approaches, also find usage and we include important results from such studies in our review. Still, RCTs are widely viewed as the backbone of high-standard impact evaluations.

comparisons. The most compelling and straightforward way to accomplish such comparisons is to randomly assign potential beneficiaries into different groups, where one group receives a particular intervention and the other does not. Methods that try to recover an equivalent counterfactual comparison without random assignment exist and constitute valuable contributions to the literature but are not as persuasive in general. Studies should also strive to work with large samples, minimize and assess dropout and attrition rates, and estimate spillovers to untreated study participants or non-study populations.

The rigorous impact evaluations that we are planning to implement as part of Invest for Jobs aim to apply state-of-the-art methodologies and study designs to help fill some of the research gaps identified in this briefing. They are intended to be large RCTs in low-income contexts; they are planned to include a firm-level intervention; measured outcomes are designed to reflect a comprehensive understanding of job quality and economic improvement; and selection and recruitment strategies will themselves be the subject of rigorous evaluation.

## Plan for briefing

We first lay out the principal theory of change associated with training programs in the literature, in which the creation of human capital serves as the link between interventions and outcomes. We note the importance of the economic context for this narrative, and also outline alternative mechanisms that emphasize the signaling and network effects of training programs. We then turn to brief descriptions of core insights from the empirical literature on training interventions, and identify a set of research gaps that remain. We next discuss methodological and practical questions, including the advantages and challenges of randomization. We highlight several issues related to the current COVID-19-induced economic crisis, and end the briefing with a description of two candidate projects for impact evaluations, the Professionalization of Artisans initiative in Ghana and SME Trainings in Côte d'Ivoire.

## Theories of change

### Skills acquisition

A number of mechanisms that link training programs to improvements in employment and job quality outcomes have been proposed in the literature.

At the level of individual workers and potential employees, many programs aim to improve technical (“hard”) skills, which can be either general (e.g. literacy, basic quantitative or financial reasoning)<sup>3</sup> or specialized (e.g. trade-specific competencies or knowledge concerning particular production processes). Many programs additionally or alternatively target life (“soft”) skills such as the ability to communicate effectively and work well with others.

At the management (or firm) level, programs similarly strive to improve technical capacity (e.g. accounting techniques, product development, or marketing) as well as process and soft skills (e.g. decision-making and effective team leadership).<sup>4</sup> Several studies have lamented that “managerial capital” is often missing in developing countries and that suboptimal business practices can help explain the productivity gap between poor and rich countries (Bloom and Van Reenen 2010, Bloom *et al.* 2010), so this is where management-level training interventions try to make a difference.<sup>5</sup>

---

<sup>3</sup> Drexler *et al.* (2014) propose conveying simplified rules of thumb for financial management as opposed to standard financial literacy curricula.

<sup>4</sup> See McKenzie and Woodruff (2014) for an overview.

<sup>5</sup> Bruhn, Karlan and Schoar (2010) argue that low levels of managerial capital in developing countries affect firm growth both directly and indirectly through impeding the productivity of other inputs. McKenzie and Woodruff

Training programs vary widely in terms of target groups, contents, methods, length and intensity. For example, interventions that enroll owner-managers of micro firms tend to provide instruction on standard business practices, thereby aiming to increase labor and/or capital productivity, generate higher incomes for owners, and enable hiring of additional staff. Interventions targeting medium-sized or large firms of greater organizational complexity are more likely to work with consulting services in order to suggest bespoke improvements.<sup>6</sup>

Most trainings have traditionally been classroom-based, but a growing awareness of behavioral decision drivers has given rise to innovation and pluralism in techniques. Training programs now frequently include mentoring, visits by role models (Lafortune *et al.* 2018), peer learning (Dalton *et al.* 2019a), combinations of classroom instruction and individual consulting sessions, and/or a focus on personal initiative (Campos *et al.* 2017). But they all have in common that they claim to increase valuable human capital, which in turn affects employment and job quality indicators.

## Context matters

Linking training programs to employment outcomes by way of human capital accumulation seems straightforward, but outcomes are influenced by numerous factors beyond the direct individual (or firm-specific) accumulation effect. Most importantly, the value of newly created human capital depends on its context-specific market. In order for interventions to work the program and the participants must perform well at identifying value-generating training opportunities. Market knowledge is a crucial and underappreciated driver of program efficacy.

In fact, programs can have negative or no effects on economic outcomes even if they successfully and comprehensively transfer target skills to beneficiaries (e.g. Cho *et al.* 2013). This can be the case if the value of newly acquired skills fails to offset any earnings and career opportunities lost during the duration of the course itself, or if a course induces beneficiaries to switch into a sector that turns out to yield inferior earnings. The latter can easily produce a situation in which a training program can be a success in terms of teaching skills while actually failing to create pecuniary value for the individual participant.<sup>7</sup>

Long-term effects are especially difficult to anticipate and can be counterintuitive, given the complex ways in which trainings interact with their economic contexts. For example, Kugler *et al.* (2015) argue that a vocational training opportunity in Colombia led beneficiaries to try to pursue tertiary education, because the training improved individuals' knowledge about sector-specific returns to a university degree.

## Signaling and networks

Training programs can also affect employment outcomes through channels other than skills acquisition and the creation of human capital. Perhaps the most prominent alternative mechanisms concern how program beneficiaries access job opportunities, whether or not program participation leads to improved job-related skills.

First, training programs can provide references and certifications that employers may value and rely on when screening applicants (Abel *et al.* forthcoming). Depending on program enrollment and certification criteria, successful program completion can serve as a signal of quality which increases the chances of an

---

(2017) validate this point empirically, showing that variation in business practices predicts a large share of differences in firm performance for a sample of micro-enterprises in developing countries.

<sup>6</sup> Notable examples from developing economies are Bruhn *et al.* (2018) for micro, small and medium enterprises in Mexico, and Bloom *et al.* (2013, 2020) for large textile firms in India.

<sup>7</sup> For example, Novella *et al.* (2018) analyze the Chilean Bono Trabajador Activo (BTA) voucher program, which funded training courses for formally employed individuals. They find that employment probabilities and formal-sector income are lower for those who enrolled in a training opportunity, even years later and in particular for those who were willing to change sectors.

applicant getting a job. For instance, prospective employers might infer from a reference or certificate that applicants are relatively smarter and more motivated than other applicants. Therefore, an applicant might not get the job because of skills acquired in training but because of other characteristics associated with a certificate and reference letter. In this narrative, training programs usefully facilitate sorting and the allocation of skills profiles in the local economy, but the actual training provided can be of secondary importance.

Second, training programs can create formal and informal networks that can later help beneficiaries to be successful on the job market. Note that training programs can have this effect and give participants a leg up in their search for employment even if they do not impact human capital.<sup>8</sup>

As we note below, the causal pathways described here remain largely speculative in many studies, and the mechanisms that link training programs to job acquisition, retention, and quality are in need of continued empirical examination.

## Empirical evidence

Training programs and related entrepreneurship interventions have received considerable attention in the empirical literature in development economics, notwithstanding the relative dearth of evidence concerning causal mechanisms. A recent meta-analysis drew on more than a hundred impact evaluations (Kluve *et al.* 2017, 2019) and is one of several reviews that have attempted to synthesize the voluminous literature (Betcherman *et al.* 2004, Cho and Honorati 2014, Blattman and Ralston 2015, Grimm and Paffhausen 2015, J-PAL 2017). The scholarly interest in this topic seems appropriate, given that at this point “skills training programs are ... the most widely used labor market intervention for young people worldwide” (Kluve *et al.* 2017).

Overall, the majority of stand-alone training programs appear to have weak or no average effects on employment and job quality outcomes (Betcherman *et al.* 2004, Card *et al.* 2011, McKenzie and Woodruff 2014, Grimm and Paffhausen 2015, J-PAL 2017). Blattman and Ralston (2015) conclude that perhaps development initiatives should shift from a focus on trainings to a focus on capital-intensive programs, whether in the form of cash transfers or small business loans.

In contrast, reviews also find evidence of positive effects in a sizable minority of studies. Kluve *et al.* (2017, 2019) find that more than one in three youth employment programs positively affected employment rates or earnings, and Cho and Honorati (2014) identify impacts on business knowledge but not revenue or earnings in a review of programs in developing countries. McKenzie and Woodruff (2014) caution that well-powered studies (with large enough samples and a sufficient number of participants) remain relatively rare.

Meta-analytic summaries aside, well-designed evaluations have found some training programs to be efficacious, including some in African countries. While trainings are clearly not universally beneficial, they can work under certain conditions for some populations.<sup>9</sup>

---

<sup>8</sup> In the case of training courses that are implemented jointly with private firms, it is not uncommon for these firms to fill open positions or apprenticeships with course graduates. Such transfers may even be an advertised feature of integrated, multipronged programs. Here the program may not create any additional jobs, but still provides training participants with access to employment they would not otherwise have had.

<sup>9</sup> For example, Adoho *et al.* (2014) show large increases in employment and earnings among beneficiaries of a program for women in Liberia, relative to a control group, with income gains far exceeding budgetary program costs. Alcid (2014) report higher rates of employment for participants of a training program in Rwanda. De Mel *et al.* (2014) find that trainings improve expected profitability among individuals opening a new business, although the same does not hold for existing businesses.

## Multipronged programs fare better

Comprehensive, multipronged training and support programs appear to outperform stand-alone classes. Multicomponent interventions usually combine a vocational training element with job placement or financing support. For example, Chakravarty (2016) link an intervention that provided skills training and employment placement services to more than 40,000 youths in Nepal to a 15 percentage point increase in non-farm employment a year later. Attanasio *et al.* (2011) report on similarly positive effects of a vocational training program in Colombia that combined three months of classroom instruction with another three months of an in-company apprenticeship. Others have stressed the importance of financial support, given the ubiquitous credit constraints that innovators and entrepreneurs face across Africa (Cho and Honorati 2014, Blattman and Ralston 2015, Blattman and Annan 2016, Blattman *et al.* 2016). Skills trainings have also been successfully combined with formalization support, including e.g. opening a bank account and registering small informal businesses (Benhassine *et al.* 2018). These successes have led some best-practice guides to recommend that training interventions should in general be multipronged if at all possible (Datta *et al.* 2018).

## Private-sector involvement helps

A number of studies have found private-sector involvement to contribute to training program efficacy (J-PAL 2017). There are primarily three related ways in which companies tend to become involved. First, they can assist in curriculum development and help to ensure that training content is contextually suitable and responsive to market demand. A retail sector program in Indonesia, for example, found that a learning module that included information on best practices gathered from local shop owners performed particularly well (Dalton *et al.* 2019b).

Second, private actors can provide mentors and points of contact for beneficiaries seeking information or opportunities beyond the confines of the classroom. This can have desirable effects when local know-how is superior to generic information delivered in a formal class setting. For example, a program for female entrepreneurs in Kenya that evaluated formal business classes against mentorships by local business owners found no changes in profit for in-class participants, but a (short-term) jump in profits among mentees (Brooks *et al.* 2018).

Third, private companies can play an important role in directly absorbing newly trained beneficiaries, whether temporarily through internships and apprenticeships or by way of long-term job offers. Studies in Colombia (Attanasio *et al.* 2011), Argentina (Alzúa *et al.* 2016), and Yemen (McKenzie *et al.* 2016) have found that the provision of internship opportunities by private-sector employers in combination with a training course is associated with higher rates of subsequent employment. Note that trainings in these cases provide an advantage to beneficiaries even if they do not lead to additional jobs being created.

## Program quality matters

Implementation quality and design details of training programs matter and might even be more important than the choice of a specific program type with respect to improvements in employment outcomes (Kluve *et al.* 2019). Effects may be enhanced through increased training intensity and duration, responsiveness to beneficiary needs and abilities, sustained monitoring of service providers and beneficiary progress, dependable and consistent program delivery, and appropriately incentivized contracts with service providers.<sup>10</sup>

---

<sup>10</sup> Program quality also concerns the tools used to effectively signal the value of skills training to employers. Research in this area is limited, and more evaluations of training program certifications are needed (J-PAL 2017).

Short-duration trainings, for example, frequently fail to affect employee or business performance, while high-duration trainings that can cover more content have been associated with increased long-term employment and earnings (McKenzie and Woodruff 2014, Escudero *et al.* 2019).<sup>11</sup>

It is not always straightforward to anticipate which combination of program features will be optimal in order to improve beneficiaries' economic lives. Programs also have budget constraints, and they have to decide if it is better to deliver high-quality training to a small group or lower-quality training to more people (J-PAL 2017). Blattman und Ralston (2015) suggest an iterative, incremental process of improvement in these circumstances, which prescribes piloting various implementation changes on a small scale before rolling them out to large groups of beneficiaries.

### Soft skills training is usually not enough

Employers frequently identify soft skills such as teamwork competency, time management, work ethic, honesty and integrity, and an ability to communicate effectively as an essential requirement for recruitment. However, results on the efficacy of trainings with a soft skill component in a developing context have been lackluster (Kluve *et al.* 2017). For example, Groh *et al.* (2016) conducted an RCT of a soft skills program for female college graduates in Jordan. The study found no significant employment effects, even though program quality and length were above the national average and Jordan employers had confirmed their demand for soft skills in graduates.

One avenue for raising the impact of a soft skills intervention may be to combine it with cognitive skills training. A J-PAL review (2017) contends that this interaction can improve labor market and educational outcomes such as attendance and graduation rates, especially among the most disadvantaged. Blattman and Ralston (2015), by contrast, summarize that the evidence for any effect of soft skills training on employment outcomes is quite limited, even when combined with technical or cognitive skills training.<sup>12</sup>

One challenge is that some evaluations of mixed trainings find positive effects on employment outcomes, but are unable to separate out the impacts of each program component. Blattman und Annan (2016), for example, analyze the effects of a program that provided agricultural training and inputs as well as significant psychosocial support to ex-fighters in Liberia. The program was able to increase farm employment and earnings, but the partial contribution of its "life skills" component remains unknown.<sup>13</sup> The same applies with respect to life skill modules included in a youth training program in the Dominican Republic studied by Ibararán *et al.* (2014).<sup>14</sup>

At this point skepticism seems warranted with respect to the efficacy of stand-alone soft skills programming. Perhaps soft skills cannot readily be taught via (typically short-term) interventions, or there may be an inconsistency between employers' self-declared and their actual needs and recruitment priorities, which future research could explore.

---

<sup>11</sup> The effect of program duration can vary across contexts. For example, Hirshleifer *et al.* (2014) suggest that for vocational training courses in Turkey longer programs have less impact on employment than shorter ones, perhaps because longer courses leave less time for job searches.

<sup>12</sup> More encouragingly, they also find that soft skills training can reduce anti-social, criminal, and violent behaviors.

<sup>13</sup> More precisely, the study could not identify separate effects for different training components in the absence of a factorial experimental design.

<sup>14</sup> The study does identify positive program effects on non-cognitive outcomes such as social skills, but cannot link these results to improved employment outcomes.

## Effects are highly heterogeneous

Many studies report effect heterogeneity, both in terms of the regional or country-level economic context and in terms of beneficiary characteristics. We broadly highlight three factors: Macro-level economic conditions; poverty and disadvantage at the beneficiary level; and gender.<sup>15</sup>

### Regional or country-specific economic conditions

Program effects vary with the development level of a country. Despite the generally moderate impacts Kluve *et al.* (2017) find systematic evidence of larger training effects on employment and earnings outcomes in low- and middle-income countries than in high-income countries. In advanced economies, demand for skilled labor is high and cohorts are overall relatively well-educated to start with, so a limited-duration skills intervention targeting disadvantaged workers might not be enough to help them catch up with their peers.

Both within and between developing countries, the performance of training programs should vary with labor demand conditions and the extent to which programs' target skills match local market demand. Indeed, Ibarrarán *et al.* (2019) document greater returns to training investments in regions with higher demand for skilled labor in the case of a training program in the Dominican Republic. More generally, Escudero *et al.* (2019) find more pronounced positive effects of training programs on labor market outcomes in times of economic expansion in a meta-analysis of 296 impact evaluations of active labor market programs in Latin America and the Caribbean. An expanding, innovating economy often sees demand for skilled workers rise (Vivarelli 2014).<sup>16</sup>

### Beneficiary-level disadvantage and poverty

Program activities targeted at particularly disadvantaged, poor beneficiaries tend to produce the largest effects. Training interventions are often narrow in scope and duration, so they might not move the needle much for better-off individuals, but they can make a difference for those starting out with more limited skill sets. Abebe *et al.* (2017), for example, detect the most substantial effects of an intervention including a job application workshop and transport subsidies in Ethiopia among the most disadvantaged participants.

Reviews have come to similar conclusions. In their meta-analysis of interventions in Latin America and the Caribbean, Escudero *et al.* (2019) find that training programs increase employment rates and earnings in particular when they explicitly target poor individuals. Kluve *et al.* (2017, 2019) also find that disadvantaged, low-income individuals benefit most in their meta-analysis of youth employment interventions.

### Gender

Training programs frequently produce heterogeneous gender effects. Reviews indicate that, if anything and on average, women benefit slightly more from skills training than men. Escudero *et al.* (2019) find this divergence in their meta-analytic study of programs in Latin America and the Caribbean. Blattman and Ralston (2015) note that youth technical and vocational trainings seem to positively affect women but not men, although business skill programs appear to have similar effects for both. And Kluve *et al.* (2017) observe that effect sizes with respect to employment and earnings may be larger for young

---

<sup>15</sup> These are not the only characteristics associated with effect heterogeneity in the literature. For business trainings, for example, Anderson *et al.* (2016) show with an RCT conducted in South Africa that learning financial skills enhances established firms' profitability, but marketing skills improve the profits of small, new market entrants. A finance focus helps reduce costs and improves profits through efficiency, while a marketing focus achieves greater profits through growth.

<sup>16</sup> Hirshleifer *et al.* (2014) note that the links between contextual economic indicators—in their case regional unemployment rates—and demand for skilled labor can be ambiguous.



women than young men, but caution that this should not be taken to imply that programs targeting women will generally hold greater promise.

There is in fact tremendous variation in how gender interacts with interventions across studies, and the extent to which beneficiaries of one gender outperform others may be heavily context-dependent. In a number of studies, women appear to be better served by a given intervention. For example, Chakravarty *et al.* (2016) show that women gained employment (both overall and non-farm) to a greater extent than men following a youth skills training and job placement program in Nepal, and Attanasio *et al.* (2011) find large effects for women only on employment and earnings for a training program in Colombia consisting of classroom skills training and internships for poor youths.<sup>17</sup>

In other studies, meaningful effects are observed for male but not female participants. Ibararán *et al.* (2019) find an effect of youth trainings on formal employment for men only in the Dominican Republic. So do Alzúa *et al.* (2016) for a vocational training program including skills training and private sector internships in Argentina as well as Hirshleifer *et al.* (2014) in the context of short-term vocational skills programs in Turkey, with the latter also reporting an effect for men only on hours worked. In Yemen, McKenzie *et al.* (2016) too find an effect of a short skills training and subsequent internship placement program on hours worked for men only. Cho *et al.* (2013) conducted an RCT on a youth vocational training and entrepreneurship program in Malawi, in which participants received an apprenticeship by a master craftsman, and identify significant post-training investments in human capital among male beneficiaries only.

How and why gender interacts with programming as it does remains largely speculative at this point. The reason why women benefit more from training interventions in some contexts could be that they have more to gain, as they usually start out with worse outcomes than men. In other contexts, constraints may still be binding. Cho *et al.* (2013), for instance, suggest that female participants suffer disproportionately from credit constraints and external shocks (pregnancy, illness or injury in the household, other family obligations), which make sustained investments in their own human capital difficult. Even if an effect is present for both men and women, different mechanisms could be underpinning each. Take the Kugler *et al.* (2015) study in Colombia, which found that participants in a training course were more likely to be enrolled in tertiary education programs more than a decade later. For women, vocational training appears to have relaxed credit constraints. For men, training improved their field-specific understanding of the returns to tertiary education. Untangling gender-specific mechanisms more systematically across contexts remains a task for future research.

## Key research gaps

### Rigorous impact evaluations in low-income countries

We have summarized some central insights from the extensive empirical literature on training programs above, but note that the bulk of this literature studies interventions in developed or upper middle-income countries. Far fewer training programs in low-income or lower middle-income countries are rigorously evaluated, in particular in Africa. Among the reviewed studies there are hardly any German initiatives that would have undergone a rigorous impact assessment in excess of routine monitoring and evaluation.

This evidence gap matters, because context matters for program efficacy (Quinn and Woodruff 2019). Take the “persistent informality” and ubiquity of self-employment that characterize many least

---

<sup>17</sup> Note that Chakravarty *et al.* (2016) observe that effects on employment type, hours worked, and earnings did not differ across gender, and that the employment effect for women could be due to the fact that men started at a much higher level of employment and it was easier for women to make large marginal gains, or could stem from women’s course choices as employment effects were primarily observed in traditionally female-dominated training fields. Attanasio *et al.* (2011) report no gender difference in program impact on rates of formal employment, but warn that the effect for men could be attributable to attrition bias.

developed economies where typically two-thirds of the labor force are informal (World Bank 2019). In these settings, people usually have a “portfolio of work” rather than one place of employment (Blattman and Ralston 2015), and a large share of the poor act as entrepreneurs, in both agricultural and non-agricultural sectors, as Banerjee and Duflo (2007) noted in their classic description of the lives of the poor across countries.

It seems reasonable to expect that training interventions that “work” in a richer economy might have very different effects in this kind of setting. We see a need for additional impact evaluations in lower-middle and low-income countries to expand the relevant evidence base.

### Comprehensive measures of job quality

The development of comprehensive and context-sensitive measures of job quality is of particular concern for understanding the impacts of training interventions in developing countries, including those in Africa. It may not be sensible to apply a default model of what constitutes a good job in a rich-country context. Employment, for example, is not a useful binary indicator in a place where many people are underemployed, lack sufficient cash income, must work for subsistence and cannot afford the luxury of unemployment (Blattman and Ralston 2015). Earnings from formal and informal work, hours worked and hours spent seeking work, types and variety of tasks completed, reported job satisfaction, workplace safety, health indicators and work-related injuries, time spent with family or household members, other family and household considerations such as food security and relatives’ ability to access medical care or educational opportunities could all feature in a discussion of employment quality, and future studies could benefit from conceiving of job quality in broad, contextually appropriate ways.

### Private-sector engagement in context

Private-sector involvement in training interventions appears to be a winning strategy. Private firms may be better suited to provide technical training, both in terms of content development and in terms of their incentives to induce learning and effectively transfer skills. They may be better able to mold training programs to sector-specific needs and to ensure industry acceptance of any program certifications. They can absorb trainees directly or rely on industry networks for placement assistance. Funders can also incentivize private training providers to achieve benchmarked labor market outcomes in a way that they might not be able to with a public sector entity (Chakravarty *et al.* 2016).

Reviews have indeed found that trainings with private sector involvement outperform sole public sector initiatives (Card *et al.* 2015). However, a preponderance of studies evaluate trainings in developed countries. This is an issue in general, as noted before, but particularly so in this context, because developing countries in which self-employment dominates the labor market and informality rules may well lack the private sector capacity needed to make trainings that hinge on private sector contributions a success.

Studies of private firm engagement have also tended to focus on outcomes such as employability and earnings, which are of course important outcomes for any study of a training intervention, but the recent work by Blattman and Dercon (2018) has shown that private sector involvement (in their case through the allocation of factory jobs to study participants in Ethiopia) may have important and countervailing implications for job quality, life satisfaction, and health indicators. This is an underexplored issue and likely an especially pertinent observation for programs in low-income economies, including in many African countries, where labor rights and social protection are often weak and/or not enforced.

### Firm-level interventions and outcomes

While there is a vast body of literature on the impacts of business and management trainings for microenterprises in developing countries, the evidence base becomes progressively thinner for larger firm sizes. Even studies that—according to their title—focus on small and medium-sized enterprises (SMEs) often have mainly micro-sized firms in their sample. This partly reflects the heavily skewed firm size distribution in developing countries, but the paucity of evidence on medium or even large enterprises

also constitutes a genuine research gap. The appropriate type of intervention is closely linked to firm size, with larger, more complex firms requiring individual consulting and tailor-made strategies that take the whole firm or plant into account. As the findings of McKenzie and Woodruff (2017) illustrate, there is considerable scope for improving firm-level outcomes, with business practices having large effects on the productivity and profits in their sample of medium and large firms in developing economies. At the same time, there are hardly any studies of consulting interventions in medium or large enterprises, with Bruhn *et al.* (2018) for Mexico and Bloom *et al.* (2013, 2020) for India being notable exceptions.

This omission is particularly problematic from an employment perspective, as one might expect wage job creation to happen especially in medium and large enterprises. Studies focusing on microenterprises often do not consider employment creation explicitly or find no or very small effects (McKenzie and Woodruff 2014, Grimm and Paffhausen 2015), as the scope of expansion is limited. The expected employment effects of business trainings/consulting in larger firms are not only positive, as the limited existing evidence illustrates. On the one hand, Bruhn *et al.* (2018), find positive, large and significant effects of consulting services on employment creation in their sample of micro, small and medium-sized enterprises. Bloom *et al.* (2013), on the other hand, find that while productivity rose, employment dropped in treatment plants due to the adoption of labor-saving technology. Anderson *et al.* (2016) provide some evidence on the link between training contents and employment effects, finding that a marketing curriculum leads to increases in sales and employment, whereas an accounting curriculum leads to cost savings. Their results underscore the importance of balancing the objectives of firm performance and job creation when designing interventions, which still requires a much improved understanding of the mechanisms driving firm-level effects.

### Causal pathways

The existing literature contains relatively little systematic evidence about the mechanisms that connect training programs with employment and job quality outcomes (Kluve *et al.* 2017). Many studies offer qualitative accounts that helpfully lay out the landscape of possible pathways, but few collect data on intermediate outcomes or try to quantitatively estimate the importance of these pathways. This also means that meta-analytic work on mechanisms is scarce.

Intermediate outcomes are often difficult to measure and mechanisms thus challenging to estimate. Some insights can be gleaned from thinking creatively about testable corollaries of different proposed mechanisms. For example, Alzúa (2016) suggests that the temporal fragility of effects could hint at the underlying mechanism. Training that increases human capital, for example by transferring vocational skills and technical expertise, should put beneficiaries on a long-term trajectory of increased earnings and heightened employability. Conversely, effects might decay more rapidly for trainings—especially those administered in partnership with private companies—that affect employment by providing beneficiaries with informal or formal contacts that help them get a job.<sup>18</sup>

### Dissipation over time

The effects of training interventions diminish over time. Few programs have sustained, long term effects (J-PAL 2017). But the rates at which effects decay is highly uncertain and difficult to anticipate. We still lack a systematic understanding of the extent to which and the processes by way of which different types of effects wane over time (Grimm and Paffhausen 2015).

The time horizon is narrow for the effects of many training programs. For example, an evaluation of labor market interventions that targeted young women in poor areas of Nairobi identified income effects that dissipated after a year (Brudevold-Newman *et al.* 2017). A study of a training program in Turkey estimated a three-year window for improvements in rates of formal employment (Hirshleifer *et al.* 2014). A

---

<sup>18</sup> Depending on the context, one could also argue that network effects should outlast content effects. Our point is that researchers should try to develop testable implications of potential mechanisms.

vocational training program in Argentina substantially improved formal employment rates and earnings—carefully measured using administrative data—after 18 months, but no main effects could be detected after 33 months and after four years (Alzúa *et al.* 2016). In an analysis of thirteen different programs that recorded subjects’ employment situation on an annual basis and of which twelve showed a positive effect after a year, Cummings and Bloom (2020) report that six increased earnings beyond the first and only four beyond the second year.

However, several other studies have identified long-run effects (Attanasio *et al.* 2015, Kugler *et al.* 2015, Ibarrarán *et al.* 2019). A Colombian voucher program enabling training access appears to have increased formal sector earnings by 35,000 Colombian pesos, or 14 percent, the probability of working in the formal sector by four percentage points and for a large firm by three percentage points, all up to ten years later (Attanasio *et al.* 2015), and a randomly assigned training program in the Dominican Republic continued to raise earnings six years after the fact (Ibarrarán *et al.* 2019). Why and how some effects do not dissipate quickly over time, as most do, remains uncertain at this point.

### Cost effectiveness

Rigorous impact evaluations have made strides in the last few years estimating the causal effects of training programs, but rarely do they address the cost effectiveness of interventions. One review found that only one in ten evaluations discussed cost effectiveness (Betcherman *et al.* 2007).<sup>19</sup> More studies are needed that compare the economic value generated by an intervention to the intervention’s budgetary cost as well as its associated opportunity costs. In some cases, programs may yield returns in excess of their budget, but fail to outperform alternative interventions (e.g. cash grants to beneficiaries).<sup>20</sup> At this time our understanding of how training programs perform in these comparisons remains limited.

### Recruitment and selection processes

Documenting and analyzing the selection and recruitment of trainees and applicants should be a part of rigorous impact evaluations whenever possible. Learning more about the application and selection process could help explain the mixed record of success of training interventions. RCTs like the ones discussed above typically ensure comparable treatment and control groups, but frequently suffer from selection problems that manifest themselves at the program enrollment stage. Nobody is forced to participate in a training program. Often, entrepreneurs and training applicants self-select into training participation. Likewise, certain types of applicants are screened out by the recruitment process. Self-selection and screening processes, however, imply that these entrepreneurs, firms and applicants are possibly more motivated, possess denser networks, and higher cognitive skills than those who did not apply and those who got rejected. Extrapolating the measured effect of an RCT to other sectors or the entire population of, say, micro-entrepreneurs or youths is therefore highly problematic, unless recruitment processes and selection effects are well-understood and incorporated into the analysis.<sup>21</sup>

Understanding the recruitment process would also appear to be a promising area for research because relatively light-touch interventions have been shown to increase applications from disadvantaged target groups. For example, Gee (2018) shows that providing information on the number of applicants for a position can increase application rates, particularly among women. Likewise, Subramanian (2019) demonstrates for online platforms in urban Pakistan that adding information on instructors’ gender and

---

<sup>19</sup> See Kluge *et al.* (2017) for a more recent review discussing this persistent issue.

<sup>20</sup> This is one reason why compelling studies often have multiple treatment arms, which permit comparisons not only between one treatment and a control group, but also between different treatment alternatives (or combinations of those alternatives in a factorial design).

<sup>21</sup> Selection also occurs by way of non-compliance with the experimental assignment and sample attrition, which we address in the following section on methodological challenges.

reminders on family responsibilities (social norms) can influence application rates, particularly among women.

Another area with very little published evidence, despite various national and international policy initiatives, concerns the question of how to increase women's application and enrollment rates in technical education and training opportunities (Del Carpio and Guadalupe 2018). Possible and relatively low-interference interventions could include quotas (Ibanez and Riener 2018), emphasizing post-training gains (Aloud *et al.* 2020, Ashraf *et al.* 2020), or offering female mentorship (Blau *et al.* 2010, Carrell and Sacerdote 2017).

## Methodological and practical challenges

Our discussion of key methodological challenges that arise in studying the impacts of skills development programs is deliberately kept brief. For a more extensive methodological discussion of RCTs and associated challenges in general, we refer interested readers to Glennerster and Takavarasha (2013).

### Non-randomized study designs

For any given skills development program, those who sign up to participate or select into participating are likely to differ in observable and unobservable ways from those who do not. Applicants may be more motivated, driven, career-oriented, or have a better idea of the skills they need to acquire than non-applicants, all of which affects their potential success in the labor market. Comparing key labor market outcomes such as income only between participants and non-participants is thus likely to overstate the true impact of the program. Impact evaluation techniques rely on constructing a credible control group that is comparable to the participants of an intervention, in order to illustrate how the participants would have performed without the intervention.

Many studies in the past have relied on techniques other than random assignment to construct such credible control groups, in line with methodological guidance at the time (Kluve 2011, GIZ 2016). Difference-in-differences estimation, in which change in outcomes over time in one group (e.g. training participants) is compared to changes in outcomes over time in another (non-participants), was and remains a popular choice whenever random assignment is not possible. It is a good idea to combine cross-sectional with before-after comparisons, as this method does. The problem, however, is that it is not guaranteed to address selection. For example, individuals that are particularly motivated to do well in a training program might be more likely to enroll. Indeed, Grimm and Paffhausen (2015) find in their review of entrepreneurship interventions in low and middle income countries that non-randomized, quasi-experimental studies systematically (and probably misleadingly) report larger effects than RCTs. For this reason, RCTs are considered the gold standard for impact evaluations and have become increasingly common, although German-funded initiatives still lag behind those of other large donors.

### Randomized controlled trials

We can avoid selection problems and identify the causal impact of an intervention by drawing treatment and control groups randomly from the same pool. Such RCTs are a powerful tool to advance our understanding and the practice of development assistance, and they are the benchmark for rigorous impact evaluations for economists around the world. They are particularly useful and the resources needed to conduct them justified if (a) there is insufficient knowledge about the impact of a program or (b) there is uncertainty whether the impact of an intervention corresponds to previously established impacts of similar interventions in other contexts. However, RCTs can be challenging to implement and conditions are not always right for them. Among other things, they usually require:

- Minimal or well-justified risks of treatment and non-treatment to subjects and staff, and a sufficiently large sample of potential beneficiaries;

- Flexibility and commitment on the implementer side and openness to researcher involvement at the pre-intervention design stage and an ability to commit to key procedures (for example a recruitment/sampling strategy);
- A data collection strategy, usually including a pre-intervention baseline (and commonly including researcher access to all actual as well as potential program participants), sharing of administrative data, and a commitment to the publication of results and data usage and publication rights for the researchers.

Practitioners are frequently concerned about the ethical, political, or practical implications of randomly assigning program benefits. Ethically, random assignment may be a particularly fair approach when the number of applicants exceeds the available training opportunities. But what about instances of undersubscription, or cases in which particular individuals in need must be treated? Implementers and researchers have in fact developed tools to address many of the most common practical objections to randomization. For example, if some units must be included in the treatment group, e.g. for political reasons, one could proceed with randomly assigning all other units. This has implications for the scope of the analysis (i.e. its applicability beyond the sample at hand), but allows researchers to cleanly identify a causal effect. In cases where all relevant units must be treated (e.g. if a program is undersubscribed), one could randomize program rollout. If it is necessary to permit self-selection into treatment or control (e.g. in vulnerable populations), one can use a so-called encouragement design that randomizes exposure to different recruitment strategies.<sup>22</sup>

Even though such technical solutions to potential concerns are available, there may be (perceived) trade-offs between evaluation needs and program objectives. In our example (discussed below) of a potential impact evaluation of an SME training program in Côte d'Ivoire, the implementing organization has the objective of creating additional employment with its training program and has developed a non-random selection and assignment mechanism that will optimally achieve this result, as far as the organization can tell.

In line with what was described above, a staggered, randomized rollout could permit compelling counterfactual comparisons. One could also try to construct an additional sample of firms that would meet the implementing organization's selection criteria. But perhaps most importantly, funders, implementing partners, and researchers in situations like this need to agree on which effects actually remain so uncertain that they require a rigorous impact evaluation. If a funder and implementing organization are convinced that they are already able to identify an optimal set of activities for a given beneficiary, there is little need for a rigorous impact evaluation. In most instances, of course, effects of programs or particular program components do remain uncertain, and the challenge lies in funders, implementing partners, and researchers finding ways to embrace this uncertainty (and potential failure) for the sake of learning.

## Sample sizes

Many impact evaluations of skills development programs are underpowered, meaning that their sample sizes are too small to find statistically significant effects. McKenzie and Woodruff (2014) report this to be the case for most existing evaluations of business training programs. In other words, it could be that some evaluations that found programs to be ineffective just lacked the necessary sample size to detect the true impact. Larger studies are clearly needed to fill this knowledge gap, but solving the sample size problem is less straightforward than it seems. First of all, the sample size necessary to detect an impact is not an absolute number, but depends on the impact of a program (with smaller impacts needing larger samples), statistical properties such as the means and variances of key variables, as well as take-up and dropout rates, all of which need to be estimated in advance. Secondly, while large samples with thousands of

---

<sup>22</sup> Encouragement designs can suffer from low program take-up rates, which translate into low statistical power, so careful planning, piloting, and monitoring of take-up rates is crucial.

beneficiaries are thus clearly desirable, their size is limited by program and research budgets as well as cost-benefit considerations.

Such trade-offs became clearly apparent when screening Invest for Jobs training programs for interventions suitable for an impact evaluation. Training programs involving many beneficiaries are typically short courses with small expected impacts. These require very large sample sizes and can be costly to evaluate, especially relative to program costs per participant and impacts. More intensive training programs with larger expected impacts tend to involve fewer beneficiaries and are often too small. As a result, only few training programs could be identified as suitable in terms of the expected number of beneficiaries.

Potential remedies to the sample size problem on the research side include letting impact evaluations run over several cohorts of a program, or evaluating multiple programs jointly, but these remedies may not always solve the problem. From a policy perspective, a key take-away for program design is that in order for a rigorous impact evaluation to robustly identify effects, the number of program participants needs to be larger than it typically is. As long as this is not the case, the true impact of many skills development programs remains unknown.

### Dropouts and survey attrition

Participants dropping out of training programs or not responding to follow-up surveys pose a serious threat to the power and validity of an impact evaluation. 10-50% of participants are estimated to drop out of training programs in developing countries (Choe *et al.* 2011). Apart from reducing sample size and statistical power, dropouts can complicate impact estimation considerably if they are not random, i.e. rates differ between treatment and control group or along other dimensions relevant to the analysis. For instance, women and persons from disadvantaged socioeconomic backgrounds are frequently found to be more likely to quit a training program. Take Cho *et al.* (2013), who find that women drop out of a vocational training program in Malawi at higher rates than men, and that their participation decisions are more strongly affected by external constraints, whereas men react more to opportunities. Being among the most disadvantaged could also correlate with living in a rural area, which increases travel distances and the likelihood of dropouts. Maitra and Mani (2017), for example, find a correlation between distance to training site and attrition in an RCT evaluating a vocational education program for Indian women. In situations like these, the estimated program impacts are no longer representative of the general population or the initial sample and can over- or understate differences in treatment impacts across genders or other factors.

Dropouts are of course primarily a concern for the implementation side, as they hamper the effectiveness of a training program and resources are spent on training spots that remain empty. Also, dropouts may be concentrated among vulnerable groups who stand to gain the most from participating in a program. Tackling dropouts requires an improved understanding of the underlying causes, meaning more tracking studies like Cho *et al.* (2013) are required. Some studies also incentivize continued participation, for example by requiring regular deposits to be repaid upon program completion (Maitra and Mani 2017) or by enrolling participants in an end-of-program raffle for valuable goods (Scacco and Warren 2018).

### Spillovers and general-equilibrium effects

Apart from their direct impact on beneficiaries, skills development programs may also affect non-beneficiaries indirectly, through spillover effects. Spillover effects can be positive, when skills learned during training are passed on to coworkers or family members benefit from income gains, or negative, for example when a business training leads entrepreneurs to expand their business and cause losses for their competitors. If a program is large and impactful enough, it may have general equilibrium effects and change outcomes for whole markets, for example when the availability of skilled employees attracts foreign investment.

Spillovers and general equilibrium effects pose several analytical challenges: spillovers can sometimes be captured in an experimental setting but identifying program impacts becomes more complicated if

control group outcomes are also affected by the program. Spillover effects on individuals outside the control group may go undetected. Ways to deal with possible spillover effects need to be considered explicitly in the experimental design. One way of doing so is to cluster treatments and ensure distance between treatment and control groups, another is to incorporate the possibility of measuring individual-level spillovers when planning an RCT. An example from Colombia is given by Kugler *et al.* (2015), who are able to show that vocational training boosts not only female participants' own rates of formal employment, but also those of their relatives.

The importance of studying spillovers and general equilibrium effects is a direct function of their policy relevance and how they change the balance between costs and benefits. Positive spillovers on non-participants may often be intentional and are side benefits that should be seen as part of the overall program impact and deserve to be studied further. Negative spillovers can defeat a training program's original objectives, theoretically up to the point where income gains and losses become a zero-sum game at the community level. For example, McKenzie and Puerto (2017) find that training female market entrepreneurs in Kenya grows the size of the markets without harming competitors, but the evidence base on this question overall is still thin (McKenzie and Woodruff 2014).

General equilibrium effects are hard to tackle empirically, as the RCT methodology is not designed to study them. The omission of general equilibrium effects becomes particularly relevant if a program is to be scaled up: consider a training program for tailors in Ethiopia, which was found in an impact evaluation to increase incomes for participants. Expanding this program to the whole province might lead to a surge in the supply of clothing made in Ethiopia, a saturated market, depressed prices, and thus much smaller income gains than before. General equilibrium effects can be positive as well, if e.g. intervention-induced economic activity sparks investment in complementary sectors. Overall, general equilibrium effects do not seem too likely in the case of Invest for Jobs skills development interventions given the small size of most programs. There may be cases, however, where such effects occur in local markets.

## COVID-19 considerations

The ongoing coronavirus pandemic has profound impacts on economies and daily lives around the world. We see at least six key implications for impact evaluations of training programs, which will complicate near-term studies, above and beyond the methodological challenges we described above.

First, the steep decline in economic activity due to lock-downs as a response to the pandemic, unprecedented in its speed and geographic reach, means that interventions will take place against a backdrop of sudden and severe economic hardship. This context of course matters for the efficacy of training programs. It is difficult to see trainees improve their economic lot if the relevant demand for upskilled labor has collapsed. This might lead some to hypothesize a lack of efficacy under current conditions, but the real challenge (and perhaps opportunity) is that effects in a situation as it is now have not been previously assessed. The body of literature on training programs during economic hard times is small, even though it is important to ask what their effects are during periods of economic contraction.<sup>23</sup>

Second, this means that we need to reconceptualize some hypotheses. Given the current economic crisis, an evaluation would assess program impact on economic resilience rather than on performance improvements under normal conditions. The relevant questions are whether an intervention can mitigate the negative consequences of the crisis, e.g. by smoothing consumption at the individual level or company survival at the firm level. Can training programs help businesses cope with a potential recession, retain customers and prioritize core activities? Do beneficiary firms face less drastic contractions in sales and profits, and do they lose fewer employees and less capital? And do interventions at the individual level make it less likely that participants lose their job or report drops in job satisfaction and income?

---

<sup>23</sup> See for example Hirshleifer *et al.* (2014) on this point.



Third, it is perhaps especially important to ensure proper counterfactual comparisons when overall economies are contracting. Many individuals have lost or will lose work, which means that a typical program participant could easily be worse off after a training than before. In such a situation, a simple pre-post comparison would be misleading. Alcid's (2014) evaluation of a youth employment program in Rwanda, for example, showed that participants were more likely to be employed after graduation than non-participants in a context where overall levels of employment were declining over the course of the intervention.

Fourth, the pandemic has engendered a tremendous amount of innovation in terms of content delivery during trainings. Many programs, even in developing countries, have switched or are considering switching to remote learning and instruction via video. This changed landscape is in urgent need of study. If online systems prove effective, this has important implications for how to achieve cost-efficient training delivery, even beyond the current pandemic.

Fifth, to the extent that trainings do not become virtual, we must now also worry about disease spread as a possible negative externality of programming. Interventions should routinely collect data on this issue, and ethics reviews need to consider the potential effects of training courses on disease proliferation.

Sixth, the current situation is changing quickly. Individuals, firms, training providers, and funders all face tremendous uncertainty about what economic life will look like in six months or a year. This also means that development assistance will need to proceed flexibly, remain responsive to changing circumstances, embrace adaptive programming, and allow adaptive programming itself to be assessed by way of rigorous impact evaluations.

## Invest for Jobs as an example

The Special Initiative on Training and Job Creation (now branded *Invest for Jobs*) was launched by Germany's Federal Ministry for Economic Cooperation and Development (BMZ) in mid-2018. Invest for Jobs aims to contribute to the implementation of recent policy initiatives geared towards fostering private investment, such as the Marshall Plan with Africa and the G20 Compact with Africa. Complementing these initiatives with its focus on employment, Invest for Jobs is active in eight African countries, namely Côte d'Ivoire, Egypt, Ethiopia, Ghana, Morocco, Rwanda, Senegal and Tunisia. The objective is to create up to 100,000 permanent jobs, train up to 30,000 persons, and to improve incomes and working conditions in these countries. Focal areas of Invest for Jobs are (i) *Business and Invest*, concentrated on overcoming investment barriers, (ii) *Cluster*, meaning industrial clusters, and (iii) *African Mittelstand*, comprising support for African SMEs. Training activities may be conducted in all these fields.

While focusing on employment and skills development is nothing new in German development cooperation, Invest for Jobs sets itself apart mainly in terms of its demand-driven approach to skills development and its close cooperation with the private sector. The latter applies both to German or European firms willing to invest in African partner countries and to African SMEs. What is noteworthy as well is the comprehensive nature of the outcomes targeted. Employment is targeted both in terms of conventional metrics (e.g. jobs acquired and held, hours worked) as well as in terms of job quality (e.g. job satisfaction, workplace safety).

In the remainder of this section, we briefly present two training programs implemented under Invest for Jobs, along with potential evaluation designs.

### Professionalization of Artisans in Ghana

The Professionalization of Artisans project (ProfArts) is part of Invest for Job's SME (Mittelstand) focal area in Ghana and targets craftsmen in areas such as masonry, plumbing, electrical installation, carpentry and roofing. Its aim is to improve beneficiaries' employment situation by helping them tap into the rapid technological progress that characterizes the construction sector. Beneficiaries will enroll in an artisan

directory and access skills training, modern tools, and licensing under the master craftsmen register of the Ghana Institution of Engineering (GhIE). The project will also develop a mobile application to improve artisans' market access and create linkages to potential customers, and it aims to place craftsmen in formal employment relationships with companies in the construction sector.

A rigorous impact evaluation would intersect with this project as follows. In a first step, the GhIE mobilizes 20,000 artisans across Ghana and registers them in a database. The researchers assist in developing strategies for this recruitment drive and randomly assign strategies to different target groups and geographies in order to assess their effectiveness. At the time of registration, artisans fill out a baseline survey module, if they consent to participate in the study. GhIE prepares a list of registered individuals that are eligible for program activities (top-up training, retooling, etc.), and a random subset of 10,000 of these individuals is then selected for participation. The researchers conduct post-training endline interviews with these treated individuals as well as the other registered artisans in the control group, provided they have agreed to be contacted again. Furthermore GhIE shares administrative data concerning the program participation of any potential beneficiaries that consented to participate in the research study (e.g. offers made to enroll in program activities, acceptance of such offers, enrollment, attendance, and completion of activities), which can help us understand to what extent results may be sensitive to take-up rates. The study then assesses impacts on a diverse set of employment and work quality indicators.

This evaluation would constitute a large, well-powered assessment of the impact of skills training and related support activities in a developing country in Africa, and thus help fill an important knowledge gap. The intervention would follow best practices in terms of the comprehensiveness of its multipronged approach and private firms' high level of engagement, although the study would not vary these attributes and would therefore not attempt to estimate their relative impact. Recruitment strategies, however, would be assigned so as to permit an analysis of their effects, another important area of interest to practitioners in which research has been limited.

### SME Training in Côte d'Ivoire

As part of Invest for Jobs, the Ivorian SME promotion agency, Agence Côte d'Ivoire PME, is offering a training program for managers and technical staff of SMEs aiming to create employment by boosting SME productivity. The program targets SMEs in all sectors and is adapted to each firm: after a diagnostic assessing the specific needs of each enterprise, an individual curriculum is developed. The program combines courses in small groups with individual-level training sessions at the site of the enterprise.

The chronology of an impact evaluation of the program could look as follows. After SME applications have been screened for eligibility in accordance with the existing process, a baseline survey would be conducted among all eligible SMEs. This baseline survey would cover firm-level and individual-level information and include employees who were nominated for the training program as well as those who were not. It would also collect information necessary for a diagnostic assessing training needs. Ideally, participating SMEs would then be selected randomly from the pool of eligible applicants. About six months after the training, a follow-up survey of the same firms and individuals would be conducted.

This set-up allows for a comparison of developments in treatment group and control group firms across a broad range of indicators. At the firm level, relevant outcomes are business practices, capital and labor productivity, investment in capital goods and employment creation, as well as firm performance indicators such as sales and profits. At the individual level, increases in relevant skills ranging from accounting to personnel management, marketing and logistics are of interest. The evaluation would also consider key employment outcomes, such as the quality of employment measured in terms of having a formal contract, job satisfaction, job mobility and earnings.

The program combines several features that make it very interesting from a research perspective and highly suitable for an impact evaluation: First, there is a significant evidence gap with respect to trainings offered at the firm level and for medium-sized enterprises. Second, the approach of targeting

management and technical staff at the same time is an innovative one that has hardly been studied so far. Third, it would be possible to capture individual-level outcomes of existing employees alongside firm-level outcomes such as job creation, which provides a rare opportunity to study firm-level determinants of employment outcomes. Fourth, the program is sufficiently large to allow for a well-powered study.

## Conclusion

This briefing has summarized six sets of insights concerning skills trainings in Africa and beyond. First, we described the dominant theory of change that links training interventions and related support programs to key economic outcomes, including employment and job quality, namely the creation of market-relevant human capital. We observed that market relevance of transferred skills is crucial and that we should not expect to observe program efficacy in the absence of local (or otherwise accessible) demand for upskilled labor. We also pointed to a set of alternative mechanisms, in which trainings leave graduates better off because of signaling effects and because they embed them in rent-generating networks, regardless of any effects on human capital as a productive input.

Second, we presented several core insights that emerge from the empirical literature as it stands. One common point that resonates across these findings is that program characteristics that support contextual relevance tend to work: Multicomponent programs can develop different types of human capital, some of which might prove valuable, in a way that a one-size-fits-all program (or a single-component soft skills program) cannot. Private-sector involvement and high-quality implementation can ensure responsiveness to local demands. And effects are often heterogeneous across contextual factors, such as geographic and beneficiary-level economic conditions.

Third, we outlined several avenues for contemporary and future research to build on and extend the previous work. We noted the relative dearth of research in Africa and with respect to private-sector involvement in the developing world, which strike us as important areas for continued research given the contextual contingencies of program effects. We also point out that much remains to be learned about training programs beyond employment and earnings effects, in particular about effects on job quality, firm-level interventions, causal mechanisms, the decay of effects over time, programs' cost effectiveness, and recruitment processes.

Fourth, we discussed various methodological challenges that rigorous impact evaluations must grapple with, such as attrition, sample size issues, and spillovers. The thrust of this section, however, concerns the fact that RCTs have become the global standard for rigorous impact evaluations of development assistance programs, despite the fact that they remain rare in the context of German initiatives. Non-randomized designs continue to have their place, but they are generally considered inferior in international best practice.

Fifth, we offered several implications of the ongoing COVID-19 pandemic for training program evaluations. We suggested that we know little about program effects during periods of severe economic contraction, and in this sense evaluations conducted now can perhaps help us better understand what trainings in general can achieve in the context of economic hardship. We also pointed out that RCTs with proper control groups can recover positive program effects even if everyone, including treated beneficiaries, is worse off at the end of an intervention due to economic upheaval.

Finally, we turned to two projects (in Ghana and Côte d'Ivoire) that are being implemented as part of the Special Initiative on Training and Job Creation and which are suitable for rigorous impact evaluations and, more specifically, RCTs. We note the opportunities they represent and how they connect with existing gaps in the research literature.

## References

- Abebe, G., Caria, S., Fafchamps, M., Falco, P., Franklin, S., and Quinn, S., 2017. Anonymity or Distance? Job Search and Labour Market Exclusion in a Growing African City, (34), 96.
- Abel, M., Burger, R., and Piraino, P., forthcoming. The Value of Reference Letters: Experimental Evidence from South Africa. *American Economic Journal: Applied Economics*.
- Adoho, F., Chakravarty, S., Korkoyah, D.T., Lundberg, M., and Tasneem, A., 2014. *The Impact of an Adolescent Girls Employment Program: The EPAG Project in Liberia*. The World Bank.
- Alcid, A., 2014. A Randomized Controlled Trial of Akazi Kanoze Youth in Rural Rwanda. *Education Development Center Report*.
- Aloud, M.E., Al-Rashood, S., Ganguli, I., and Zafar, B., 2020. Information and Social Norms: Experimental Evidence on the Labor Market Aspirations of Saudi Women. *National Bureau of Economic Research Working Paper Series*, Working Paper 26693.
- Alzúa, M. Laura, Cruces, G., and Lopez, C., 2016. Long-run effects of youth training programs: Experimental evidence from Argentina. *Economic Inquiry*, 54 (4), 1839–1859.
- Anderson, S.J., Chandy, R., and Zia, B., 2016. Pathways to Profits: Identifying Separate Channels of Small Firm Growth through Business Training. *World Bank Policy Research Working Paper 7774*.
- Ashraf, N., Bandiera, O., Lee, S.S., and Davenport, Edward, 2020. Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. *American Economic Review*, 110 (5), 59.
- Attanasio, O., Guarín, A., Medina, C., and Meghir, C., 2015. Long Term Impacts of Vouchers for Vocational Training: Experimental Evidence for Colombia. *NBER Working Paper Series*, Working Paper 21607, 1–38.
- Attanasio, O., Kugler, A., and Meghir, C., 2011. Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial. *American Economic Journal: Applied Economics*, 3 (3), 188–220.
- Banerjee, A.V. and Duflo, E., 2007. The Economic Lives of the Poor. *Journal of Economic Perspectives*, 21 (1), 141–167.
- Benhassine, N., McKenzie, D., Pouliquen, V., and Santini, M., 2018. Does inducing informal firms to formalize make sense? Experimental evidence from Benin. *Journal of Public Economics*, 157, 1–14.
- Betcherman, G., Godfrey, M., Puerto, S., Rother, F., and Stavreska, A., 2007. *A review of interventions to support young workers : findings of the youth employment inventory*. The World Bank, No. 41412.
- Betcherman, G., Olivas, K., and Dar, A., 2004. Impacts of active labor market programs : new evidence from evaluations with particular attention to developing and transition countries. The World Bank, No. 29142.
- Blattman, C. and Annan, J., 2016. Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State. *American Political Science Review*, 110 (1), 1–17.
- Blattman, C. and Dercon, S., 2018. The Impacts of Industrial and Entrepreneurial Work on Income and Health: Experimental Evidence from Ethiopia. *American Economic Journal: Applied Economics*, 10 (3), 1–38.
- Blattman, C., Green, E.P., Jamison, J., Lehmann, M.C., and Annan, J., 2016. The Returns to Microenterprise Support among the Ultrapoor: A Field Experiment in Postwar Uganda. *American Economic Journal: Applied Economics*, 8 (2), 35–64.

- Blattman, C. and Ralston, L., 2015. Generating Employment in Poor and Fragile States: Evidence from Labor Market and Entrepreneurship Programs. *SSRN Electronic Journal*.
- Blau, F.D., Currie, J.M., Croson, R.T.A., and Ginther, D.K., 2010. Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial. *American Economic Review*, 100 (2), 348–352.
- Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J., 2013. Does Management Matter? Evidence from India\*. *The Quarterly Journal of Economics*, 128 (1), 1–51.
- Bloom, N., Mahajan, A., McKenzie, D., and Roberts, J., 2010. Why Do Firms in Developing Countries Have Low Productivity? *American Economic Review*, 100 (2), 619–623.
- Bloom, N., Mahajan, A., McKenzie, D., and Roberts, J., 2020. Do Management Interventions Last? Evidence from India. *American Economic Journal: Applied Economics*, 12 (2), 198–219.
- Bloom, N. and Van Reenen, J., 2010. Why Do Management Practices Differ across Firms and Countries? *Journal of Economic Perspectives*, 24 (1), 203–224.
- Brooks, W., Donovan, K., and Johnson, T.R., 2018. Mentors or Teachers? Microenterprise Training in Kenya. *American Economic Journal: Applied Economics*, 10 (4), 196–221.
- Brudevold-Newman, A.P., Honorati, M., Jakiela, P., and Ozier, O., 2017. *A firm of one's own : experimental evidence on credit constraints and occupational choice*. The World Bank, No. WPS7977.
- Bruhn, M., Karlan, D., and Schoar, A., 2010. What Capital is Missing in Developing Countries? *American Economic Review*, 100 (2), 629–633.
- Bruhn, M., Karlan, D., and Schoar, A., 2018. The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico. *Journal of Political Economy*, 126 (2), 53.
- Campos, F., Frese, M., Goldstein, M., Iacovone, L., Johnson, H.C., McKenzie, D., and Mensmann, M., 2017. Teaching personal initiative beats traditional training in boosting small business in West Africa. *Science*, 357 (6357), 1287–1290.
- Card, D., Ibararán, P., Regalia, F., Rosas-Shady, D., and Soares, Y., 2011. The Labor Market Impacts of Youth Training in the Dominican Republic. *Journal of Labor Economics*, 29 (2), 267–300.
- Card, D., Kluve, J., and Weber, A., 2015. *What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations*. Cambridge, MA: National Bureau of Economic Research, No. w21431.
- Carrell, S. and Sacerdote, B., 2017. Why Do College-Going Interventions Work? *American Economic Journal: Applied Economics*, 9 (3), 124–151.
- Chakravarty, S., Lundberg, M.K.A., Nikolov, P.V., and Zenker, J., 2016. The role of training programs for youth employment in Nepal : impact evaluation report on the employment fund. The World Bank, No. WPS7656.
- Cho, Y. and Honorati, M., 2014. Entrepreneurship programs in developing countries: A meta regression analysis. *Labour Economics*, 28, 110–130.
- Cho, Y., Kalomba, D., Mobarak, A.M., and Orozco Olvera, V.H., 2013. Gender differences in the effects of vocational training : constraints on women and drop-out behavior. The World Bank, No. WPS6545.
- Choe, C., Flores-Lagunes, A., and Lee, S.-J., 2011. Do Dropouts Benefit from Training Programs? Korean Evidence Employing Methods for Continuous Treatments. *IZA Discussion Paper No. 6064*, 39.
- Cummings, D. and Bloom, D., 2020. Can Subsidized Employment Programs Help Disadvantaged Job Seekers? A Synthesis of Findings from Evaluations of 13 Programs. *OPRE Report 2020-23*, 79.

- Dalton, P.S., Rüschenpöhler, J., Uras, B., and Zia, B., 2019a. Learning to Grow from Peers: Experimental Evidence from Small Retailers in Indonesia. The World Bank.
- Dalton, P.S., Rüschenpöhler, J., Uras, B.R., and Zia, B., 2019b. Local Best Practices for Business Growth. *SSRN Electronic Journal*.
- Datta, N., Assy, A.E., Buba, J., and Watson, S.M., 2018. Integration : A New Approach to Youth Employment Programs - General Guidelines for Project Teams. The World Bank, No. 135323.
- Del Carpio, L. and Guadalupe, M., 2018. More Women in Tech? Evidence from a Field Experiment Addressing Social Identity. *CEPR Discussion Paper No. DP13234*.
- Drexler, A., Fischer, G., and Schoar, A., 2014. Keeping It Simple: Financial Literacy and Rules of Thumb. *American Economic Journal: Applied Economics*, 6 (2), 1–31.
- Escudero, V., Kluge, J., López Mourelo, E., and Pignatti, C., 2019. Active Labour Market Programmes in Latin America and the Caribbean: Evidence from a Meta-Analysis. *The Journal of Development Studies*, 55 (12), 2644–2661.
- Gee, L.K., 2018. The More You Know: Information Effects on Job Application Rates in a Large Field Experiment. *Management Science*, mns.2017.2994.
- GIZ, 2016. Full and productive employment and decent work for all.
- Glennerster, R. and Takavarasha, K., 2013. *Running randomized evaluations: a practical guide*. Princeton: Princeton University Press.
- Grimm, M. and Paffhausen, A.L., 2015. Do interventions targeted at micro-entrepreneurs and small and medium-sized firms create jobs? A systematic review of the evidence for low and middle income countries. *Labour Economics*, 32, 67–85.
- Groh, M., Krishnan, N., McKenzie, D., and Vishwanath, T., 2016. The impact of soft skills training on female youth employment: evidence from a randomized experiment in Jordan. *IZA Journal of Labor & Development*, 5 (1), 9.
- Hirshleifer, S., McKenzie, D., Almeida, R., and Ridao-Cano, C., 2014. The Impact of Vocational Training for the Unemployed: Experimental Evidence from Turkey. *The Economic Journal*, 126 (597), 2115–2146.
- Ibanez, M. and Riener, G., 2018. Sorting through Affirmative Action: Three Field Experiments in Colombia. *Journal of Labor Economics*, 36 (2), 437–478.
- Ibarrarán, P., Kluge, J., Ripani, L., and Rosas Shady, D., 2019. Experimental Evidence on the Long-Term Effects of a Youth Training Program. *ILR Review*, 72 (1), 185–222.
- Ibarraran, P., Ripani, L., Taboada, B., Villa, J.M., and Garcia, B., 2014. Life skills, employability and training for disadvantaged youth: Evidence from a randomized evaluation design. *IZA Journal of Labor & Development*, 3 (1), 10.
- J-PAL, 2017. *J-PAL Skills for Youth Program Review Paper*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab.
- Kluge, J., 2011. Measuring employment effects of technical cooperation interventions: Some methodological guidelines.
- Kluge, J., Puerto, S., Robalino, D., Romero, J.M., Rother, F., Stöterau, J., Weidenkaff, F., and Witte, M., 2017. Interventions to improve the labour market outcomes of youth: A systematic review of training, entrepreneurship promotion, employment services and subsidized employment interventions. *Campbell Systematic Reviews*, 13 (1), 1–288.

- Kluve, J., Puerto, S., Robalino, D., Romero, J.M., Rother, F., Stöterau, J., Weidenkaff, F., and Witte, M., 2019. Do youth employment programs improve labor market outcomes? A quantitative review. *World Development*, 114, 237–253.
- Kugler, A., Kugler, M., Saavedra, J., and Prada, L.O.H., 2015. *Long-Term Educational Consequences of Vocational Training in Colombia: Impacts on Young Trainees and Their Relatives*. Cambridge, MA: National Bureau of Economic Research, No. w21607.
- Lafortune, J., Riutort, J., and Tessada, J., 2018. Role Models or Individual Consulting: The Impact of Personalizing Micro-entrepreneurship Training. *American Economic Journal: Applied Economics*, 10 (4), 222–245.
- Maitra, P. and Mani, S., 2017. Learning and earning: Evidence from a randomized evaluation in India. *Labour Economics*, 45, 116–130.
- McKenzie, D., Assaf, N., and Cusolito, A.P., 2016. The demand for, and impact of, youth internships: evidence from a randomized experiment in Yemen. *IZA Journal of Labor & Development*, 5 (1), 1.
- McKenzie, D. and Puerto, S., 2017. Growing Markets through Business Trainings for Female Entrepreneurs. *World Bank Policy Research Working Paper 7993*.
- McKenzie, D. and Woodruff, C., 2014. What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World? *The World Bank Research Observer*, 29 (1), 48–82.
- McKenzie, D. and Woodruff, C., 2017. Business practices in small firms in developing countries.
- de Mel, S., McKenzie, D., and Woodruff, C., 2014. Business training and female enterprise start-up, growth, and dynamics: Experimental evidence from Sri Lanka. *Journal of Development Economics*, 106, 199–210.
- Novella, R., Rucci, G., Vazquez, C., and Kaplan, D.S., 2018. Training Vouchers and Labour Market Outcomes in Chile. *LABOUR*, 32 (2), 243–260.
- Quinn, S. and Woodruff, C., 2019. Experiments and Entrepreneurship in Developing Countries. *Annual Review of Economics*, 11 (1), 225–248.
- Scacco, A. and Warren, S.S., 2018. Can Social Contact Reduce Prejudice and Discrimination? Evidence from a Field Experiment in Nigeria. *American Political Science Review*, 112 (3), 654–677.
- Subramanian, N., 2019. Workplace Attributes and Women’s Labor Supply Decisions: Evidence from a Randomized Experiment, Job Market Paper, 87.
- Vivarelli, M., 2014. Innovation, Employment and Skills in Advanced and Developing Countries: A Survey of Economic Literature. *Journal of Economic Issues*, 48 (1), 123–154.
- World Bank, 2019. Creating Markets in Ethiopia.