Leveraging youth employment program beneficiary data

Issue 4 - April 2018

In this Knowledge Brief[1] we discuss ways in which youth employment programs can make strategic use of their data. This Brief is part of the Solutions for Youth Employment (S4YE)[2] Knowledge Brief series, which highlights the nuts and bolts of youth employment programs.

1. Introduction: Why is data so critical to youth employment?

Many solutions are being implemented to address the long list of challenges in youth employment. They take on different forms such as targeted skills and employment services programs, innovative technology applications, and using education and advocacy to raise awareness. Regardless of the method, one thing is constant: data is being generated and it is important to leverage it.

Below we discuss the importance of leveraging data in youth employment programs and explore innovative ways to collect and use beneficiary data to enhance the impact of youth employment programs. We especially highlight two projects to illustrate ways in which data can be used strategically. The first is Yes!Me, a recently developed data collection platform which allows for efficient collection of high-quality data into a central cloud-based repository and real-time analysis for youth employment program beneficiary data. The second is Tounes Ta3mal, an online youth employment platform that combines virtual and physical interactions and uses data to match youth with hiring employers (and vice versa).

2. Where is the data being created and how can we use it?

---

[1] This brief was co-authored by Jose Manuel Romero (S4YE), Danielle Carter (Accenture), and Ali Mnif (Silatech). It does not necessarily reflect the views of individual institutions or specific S4YE partners.

[2] S4YE is a multi-stakeholder coalition among public sector, private sector, and civil society actors that aims to provide leadership and resources for catalytic action to increase the number of young people engaged in productive work. For more information please go to https://s4ye.org/.

2.1 Data collection and stages of program

Youth employment programs collect data at various stages through their interactions with beneficiaries. From the initial applications that youth beneficiaries complete to their employment status several months (or years) after they exit the program, data is being generated. Its collection is usually guided by a program's monitoring and evaluation (M&E) framework. Additional data may be being 'created' but not collected. Examples include geo-location data and information on social network connections that LinkedIn gathers. Figure 1 lists three important types of data, and where they can be collected.

Basic beneficiary data: This is information that beneficiaries already have or create, and is available during a program's application process or at the start of the program (e.g. baseline). Examples include demographic characteristics, measures related to skills, and higher-level contextual data (beneficiary location, their economic context).

Intermediate outcomes and output data: During the program implementation and at its conclusion there are measures of beneficiary performance being generated. Most intermediate outcomes are individual results in areas perceived to increase youth's access to market, such as

Figure 1. Stages of beneficiary data collection

a youth's technical skills or proficiency to search for jobs. Output data refers to a beneficiary's information as it relates to a program's output and includes beneficiary attendance and exit dates from the program, variation of coursework completed, duration of attendance, number of courses taken, and potentially cost for servicing the individual beneficiary.

Job outcomes data: These are data relevant to improved access to good jobs by connecting people to opportunities and inclusion of disadvantaged segments of youth. They include a beneficiary's employment status; sector of employment; length of employment, duration of job search; wage and income levels; and number of jobs created (in the case of labor market demand-side programs). This data is usually collected once beneficiaries exit the program and thus is most difficult and costly to create. Nevertheless, it is still feasible to collect such data by taking measures such as asking for more than one phone number per beneficiary, including information for family and friends to track beneficiaries, or tracking employers and getting in touch with beneficiary through them.

Having the 'right' approach to data collection determines a project's ability to leverage it. While the 'right' approach will vary by program based on needs, data quality is an underlying factor. However, aside from its quality, the *quantity* of data matters too. It matters a lot. Programs may have an opportunity of collecting more data points and consolidating larger amounts of data (i.e. aggregating it), but often may not realize the value of any additional data, beyond the purpose of donor reporting, as discussed below. This is especially pertinent to scaled-up decentralized programs implemented through a network of partnerships or those that have outsourced services to many implementers. New opportunities are now also available to collect conventional and unconventional data, such as data from social media network analysis, administrative data of partner employers,
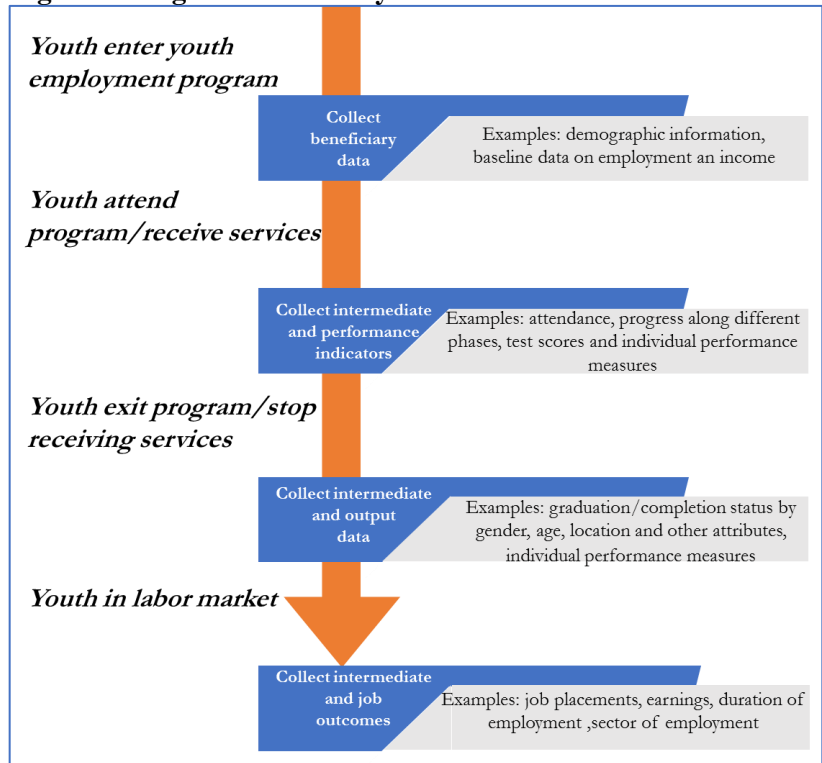
contextual data from ex ante labor market assessments, communication from peers or family, and data through the use of text messages[3]. Caseworkers offer a good opportunity to collect data. Beyond potentially helping to track beneficiaries (after they finish the program), they may be able to highlight the services they find best suited for candidates based on their firsthand experience[4]. These can be converted to data points which can be used to better document the consistency of services suggested for different segments of youth, and allow for a bottom-up approach to understanding program implementation through data. Some criteria to consider in the design of a data collection approach are cost, feasibility, and staff capacity to aggregate/combine data from data collection points.

2.2 Uses of beneficiary data

In this section we provide illustrative uses of beneficiary data to improve program outcomes, adjust program design, or provide a service directly to beneficiaries. The most advantageous uses of data may require mixing the different categories of data mentioned below (Table 1). These uses are above and beyond data collection for the purposes of donor reporting or to exclusively track progress. However, it is always necessary to weigh the advantage of using data, which often requires allocating limited resources, against making decisions based on the judgment of implementers who may have a clear idea of the implications that will result from data analysis.

Table 1. Examples of using combinations of different types of data[5]

	Basic	Intermediate	Output
Intermediate	**Data:** Performance by age group, gender, age, location, time period **Use:** Micro-adjustment of curriculum or other programmatic elements; cohort analysis.		
Output	**Data:** Graduation rates across segments. **Use:** Constraints analysis for participant groups; retention methods for specific groups.	**Data:** Test scores for those graduating. **Use:** Testing whether some services disparage beneficiaries from finishing.	
Jobs outcomes	**Data:** Graduation rate across segments. **Use:** Better projections of resource allocations needed to serve different youth segments; Determining most effective services for different youth segments.	**Data:** Performance levels of youth in different training courses and their success at increasing incomes or being placed in a job. **Use:** Determining whether there is a skills mismatch.	**Data:** Youth's cohort or region, whether they graduated, and whether they were successfully placed in job. **Use:** Overall performance of program; Determining if follow-up services (after exit) are adequate.

Profiling and screening: In the early phases of a program, profiling and screening are done for related and overlapping purposes. Profiling is the customization of services to individual participants while screening is a way to identify a target population and form the beneficiary group who will best take advantage of the services

---

[3] For a discussion of this see S4YE (2017).

[4] However, it should be noted that in many cases, ultimately this relies on caseworker's subjective assessment.

[5] Top row does not include jobs outcomes since the combinations with jobs outcomes and other data types covered by cells included in table.

provided. Basic beneficiary data can be used in profiling to segment beneficiaries and use in combination with historical output, intermediate, and job outcomes data to best tailor services. Effective profiling is important to program success as suggested by the results of a recent systematic review of youth employment programs, which found it was one of the strongest factors in predicting labor market outcomes (Kluve et al. 2016). Aside from segmenting beneficiaries to tailor services, profiling can also be used for resources and case management planning. Types of profiling include[6] (1) case-based profiling, (2) time-segmented rules-based profiling, (3) demographic-segmented rules-based profiling, and (4) data-based profiling[7]. Loxha and Morgandi (2014) point out that limited case management experience and high beneficiary to case manager ratios (results of tighter budget constraints) makes developing countries more likely to benefit from statistical profiling than high-income countries. While access to a large amount of data is necessary, it does not mean the programs have to generate them themselves. For example, connecting with those in the public sector, even in other countries, that have run statistical based profiling that can be done to learn how to adapt one's program.

Adjusting program design and course-correction: Data can allow real-time decisions when the project is mid-implementation. Course corrections do not need to be made to fundamental aspects of the program; leveraging data can allow small adjustments, such as small revisions of a curriculum. A cutting-edge approach to this is the use of experiential learning, developed by Pritchett, Samji, and Hammer (2012). It is a service prospecting approach that relies on combining standard M&E systems with experiential learning and thus referred to as MeE. It requires mainly combining intermediate and jobs outcomes data and having an ex ante implementation plan that crawls the 'design space' and sequentially adapts the program at the different implementation stages based on comparative indicators. The 'design space' refers to the possible best solutions to address a constraint, acknowledging that it is difficult to know which may perform best in the given scenario a priori. Save the Children applied this approach to tackle under-enrollment of girls in science, technology, engineering, and math (STEM) training (see Box 1).

Identifying entrepreneur potentials: In the past, youth entrepreneurship programs focused on livelihood outcomes and self-employment (Cho 2013, Grimm and Pauffhaussen 2014) and not job creation. With recent evidence on the performance of supply side programs (McKenzie 2017, Crepon and Van den Berg 2016, Fox and Kaul 2017, Kluve et al. 2016, S4YE 2018a) the potential for entrepreneurship programs to create jobs has garnered attention and, as a corollary, the use of baseline data ,which is often available, to accomplish this. Two recent studies in developing countries have considered the ability to predict business outcomes in the context of business plan competitions – both comparing human-based

> **Box 1: Save the Children's use of MeE approach**
>
> Save the Children, in partnership with Accenture, offers the "Skills to Succeed" program in Indonesia and analyzed the cause of the low enrollment of girls in STEM jobs. They diagnosed the problem using a combination of output and basic beneficiary data such average age and education (basic beneficiary data) and enrollment in each course (outputs). They identified possible root causes to be (1) courses not being interesting to girls, (2) lack of parent approval, (3) timing of course and transport available at time, (4) STEM work being unsafe for girls, and (5) girls seeing the STEM-based occupations as less feminine. After analysis based on feasibility, cost effectiveness, and time frame to implement, three of the root causes (1,4, and 5) were selected to be tested by implementing corresponding solutions at different locations. The solution tested for (1) was working with employers and trainers to develop a STEM campaign and garner girl interest; for (4) was to work with employers to improve workplace environments and share information at trainings; and for (5) to find woman mentors who currently work in STEM. After implementation, they collected participant data through phone calls asking them to rate the value, safety, and appeal of STEM courses at each location. Based on baseline, midline, and end-line data points from implementation over a few months they found that STEM mentorships made the largest difference.
>
> *Source: Save the Children (2015)*

---

[6] For an extended discussion of profiling in youth employment programs see the S4YE (2018b).
[7] See Loxha and Morgandi (2014) for detailed descriptions of each.

selection with data-based scoring. One is by Fafchamps and Woodruff (2016) who analyze a business plan competition in Ghana and find that both scoring from expert panels and survey responses have predictive power for employment, revenues, and profits, and that combining expert panel scores and baseline survey data generated the most accurate predictions. McKenzie and Sansone (2017) analyzed various prediction criteria for the YouWiN! business plan competition in Nigeria[8]. They tested predictions of employment, business survival, profits, and sales using various methods including the business plan judge scores, predictive regressions of outcomes on expert-selected characteristics, and machine learning algorithms. Thus far, the studies have shown that using data on its own might predict an entrepreneur's performance as well as, or even better than, experts doing case-by-case assessments.

Skills and labor intermediation: For job placement, an important tool is matching skills sets of beneficiaries with potential employers. This data is often available from CVs and beneficiaries can fill out simple forms or it can be collected automatically (budget permitting) as Tounes Ta3mal has done, further discussed below.

3. Data collecting through Accenture & Plan International's Yes!Me platform[9]

The Yes!Me system developed by Accenture and Plan International is a tool capable of generating high quality data that is comparable for aggregation, decreasing collection time, and allowing for real-time data monitoring. In this section we discuss its development and features, the team's experience in this process, and their next steps.

Summary: Yes!Me is an online platform that began development in 2015 and serves as a data collection and analysis solution for M&E systems for youth employment programs. It is currently in implementation in the Philippines and Indonesia, and will be launched in Tanzania, Myanmar, Tanzania, Nepal, and Vietnam in the near future. The data is entered into the system through its application and automatically consolidated in real-time with a central M&E database. There are 350 beneficiary-level variables available such as basic beneficiary information, employment, enterprise, livelihood, health, well-being, and finance information. These variables were developed using international standards set by World Bank and International Labour Organization standards (among others), and thus comparable internationally. Moreover, they were further validated and refined through workshops with implementers.

Overview of Yes!Me's development process: Human centered design (HCD) has been front and center through Yes!Me's development, by engaging end users in the earliest stages of development. After gathering information based on consultations and 'demonstration' country offices, the Yes!Me team found adapting a prebuilt and open source system (DHIS2[10]) to the youth employment context to be the best solution. The HCD approach allowed the team to make important adjustments to the platform. For example, recent feedback made it clear that (1) users specialized in different tasks, such as collecting data, creating reports, and conducting advanced analysis and (2)

---

[8] During the competition, winners were selected based on scoring of a possible 100 points composed of points given for several aspects. Job creation held the highest weight, 25 points.

[9] This section of the Knowledge Brief is based on consultations with Bo Percival, Regional Digital Development Manager, Plan International, Asia.

[10] DHIS is an open source system that was originally developed for health management and is used in 60 countries and 23 organizations.

users collecting data were not doing it through desktop computers. In response, Yes!Me has pivoted away from desktop version data collection and is now concentrating efforts on the mobile application.

Migrating the Yes!Me digital Ecosystem to Microsoft Azure cloud services was a key step in the platform's development since it permitted the efficient transfer of data into a central repository. Azure is a Microsoft Cloud system that includes a suite of cloud-based products. Nevertheless, the suitability of cloud-computing is always carefully assessed with each new implementation of the platform since legal data transfer restrictions can be rigid (See Box 2). For example, in the case of Tanzania, Yes!Me engaged the government on how and what data was being collected from the beginning.

Box 2: Data localization law and cloud computing

Data localization laws require that data about a country's residents are collected, stored, or processed in the country. Data localization laws may restrict or require meeting specific criteria before any cross-border data flows. In many instances this requires companies to store data locally, have it processed locally, and/or have government consent for data transfers. Cloud computing, the use of applications, storage, and virtual machines and other hosted services over the Internet, relies heavily on cross-border data flows.

Data quality as motivation: During the creation of Yes!Me, the aim was to unlock the potential of data analytics. However, in their initial assessments it became clear that data quality was the foremost necessity. The available data was low quality and would require high costs to standardize. The different data collection and storage methods were non-trivial issues towards its consolidation. They included collecting via paper questionnaires and transferring and saving into MS Excel tables in some instances or storage in MS Access databases in others, for example. Inconsistent data-collection protocols caused ambiguity on which were mandatory fields/questions to be filled out, which resulted in incomplete survey forms as well.

Cost and effort needed in adopting Yes!Me: The Yes!Me model is based on economies of scale. While the initial development required a large investment, being an open-source platform, the main cost of adopting the platform is in customizing it to the program's needs. Nevertheless, it is a cost and the decision to adopt Yes!Me is a strategic decision based on factors like: (1) Yes!Me's mobile app reducing data collection time by 50%[11] , (2) having built-in data quality assurance measures, (3) automating tasks such as donor reporting, and (4) the program's needs to upload existing data to the new system. The level of effort needed to upload existing data varies based on the existing M&E existing system, country capacity, and amount of data. This type of assessment forces an organization to form a clear understanding on where they stand on data quality and may reveal data inconsistencies, gaps, and opportunities.

 Challenges along the way: Internet connectivity is always an important factor for platforms of this type and it indeed presented some challenges during the early stages when data was largely being collected through desktop/laptop computers. However, since shifting to a mobile (Android) oriented collection approach, which also operates offline, it no longer plays a prominent role. Addressing local capacity was an unexpected challenge considering having taken the HCD approach. Despite Yes!Me's commitment to HCD, the capacity of local staff to adopt a new system was underestimated because there were different capacities across countries and thus concepts like 'ease of use' meant different things across them.

Next steps: The next phase of the platform will be to unpack the data's use toward decision making, especially as it related to advanced capabilities such as real-time data visual and pivot table style analysis. Another current

---

[11] In their study they found that on average it took 1 hour to collect and input the data for beneficiaries while it took around 30 minutes with Yes!Me's app.

effort is in improving successful beneficiary follow-up data collection (i.e. lowering attrition). The solutions are being considered around the concept of providing beneficiaries incentive to respond down the line. Potential outcomes include creation of a professional online network like LinkedIn for the informal sector and chat bots that operate on Facebook to reach out to friends.

4. Using data to inform youth what they need to improve and effectively match them with employers: Tounes Ta3mal in Tunisia.

Summary: Tounes Ta3mal is an online portal on employability that provides a platform for students and jobseekers, specifically those in Tunisia[12], who are transitioning into the world of work. Co-founded by Silatech and Microsoft, Tounes Ta3mal provides a wide range of career related resources and services. With 37,000 registered users and almost 800 registered employers, the platform serves as a prime example of one of the methods for creating and using youth beneficiary data to improve job-matching capacity.

Data source and collection: The main source of beneficiary data used is from the CVs of job seekers. This data is either automatically extracted from the CV that the job seeker manually uploads or via the platform's CV builder feature. The invitation to complete one's CV via this feature comes only after one subscribes to the overall platform. The CV builder plays a large role in collection and has two key aspects. First, the CV builder is meant to understand the needs of the job seeker. The information gathered in this step helps to have a clear understanding in terms of what the job seeker is expecting throughout the process. Among the data collected is the salary the job seeker expects to receive, how long the applicant expects the entire application process to take and how many applications the job seeker plans to complete before receiving an initial offer. The second key aspect is understanding their qualifications: professional experience, education, skills, and languages. The CV builder's two main components provide a 360-degree view of the beneficiary's capacity. This first component of the CV building focuses on information surrounding the job seeker's previous roles/positions. It is designed to give the employer an overview of the applicant's educational background and major/course of study. The process focuses upon the applicant's current skills/skillsets, both non-cognitive and technical. The second component focuses on information regarding the job seeker's language skills and associated proficiencies in each language listed. The final step in the process is where the user uploads their own original CV to ensure that all relevant information was provided.

Using data for segmentation reports: With the intake of thousands of users' CV data, much benefit comes from consolidation. The collected data from the CV uploads and the CV builder is consolidated and is used to help generate a talent segmentation report. This report displays all relevant data for each job seeker side by side for a more comprehensive analysis on the users in terms of the quality of the CVs, career levels and overall backgrounds to help optimize/customize the provided services. Additionally, the consolidated report enables the system to display recommended jobs for the applicants and suitable profiles for the prospective employers.

Using data for feedback and psychometric assessment: As users complete their CVs, they will receive a completeness score that directly correlates to their visibility to prospective employers to help incentivize users to complete all sections.

---

[12] However, it is a regional project also being implemented in different forms in various countries.

Using data in the Job Seekers/Employers Dashboard: One can view the overall platform through the lens of the job seeker or the employer. The job seekers dashboard is catered to the applicant, allowing them to see a variety of different pieces of information related to their job search based on their beneficiary-generated data. The dashboard automatically generates recommended jobs to the job seeker and provides a platform for the user to view their applications, how many times their CV appears in searches and how many times their CV is viewed. The job seeker can also edit or complete their CV and apply directly to jobs from this dashboard. In contrast, the employer's dashboard provides employers with numerous ways to adjust the way they interact with applicants. This dashboard enables employers to activate or deactivate a job offer, edit a job or applicant setting, customize alerts, attach questionnaires to job offers, export CVs as well as search and filter through applicants.

Using data to improve the program's existing framework: As with any digital solution in the youth employment space, there are challenges. Currently, the platform integrated into the field is the third iteration of the initial vision. It is necessary to continue to try and improve upon the existing framework and one way that is being accomplished is through the periodic surveying of users. For any solution to be successful, it is important to consider changes in the associated market and to adjust accordingly. To do this, Tounes Ta3mal has developed into a platform designed for users with no single kind of background/industry experience. Additionally, the platform is working to adapt to the market by including IT jobs in its overall agenda. Following the trends in the market is critical to success. For Tounes Ta3mal, the organization began as a demand-side driven platform but has since adapted by starting to seek out beneficiaries that met the required criteria. Further developing artificial intelligence to leverage the data within the platform and including more technologies is an area of interest for the program. Currently, best practices are being benchmarked around the world to help serve as a guide to incorporating more technologies while maintaining the service efficiently and free of charge for all users.

Looking ahead and obstacles along the way: A major persisting challenge of the platform is generating employer engagement and ensuring that employers are fully taking advantage of all the features and data the platform has to offer rather than solely using the platform for the searching and downloading of CVs. Another major challenge is the ability to consistently have clear data on job placement. Initially, job seekers were simply not uploading their CVs to the platform. However, the CV builder feature has had much success in incentivizing applicants to provide complete information as it helps them identify where their profile needs to be strengthened.

5. Conclusion

The value of being opportunistic with a mindset towards leveraging data, either already being created or available for collection, can have substantial benefits as highlighted above. As the two examples, Yes!Me and Tounes Ta3mal, illustrate, there are new and exciting opportunities for data to generate more and better services to address youth employment challenges. Technology is quickly advancing and new applications and opportunities will be available from such new tools as artificial intelligence and blockchain. Youth employment programs with an eye for such opportunities, as described in this Knowledge Brief, will be well placed to quickly adapt and find more effective paths to their goals.

References

Cho, Y., and M. Honorati, (2013). Entrepreneurship programs in developing countries: a meta regression analysis. Policy Research working paper; no. WPS 6402. Washington, DC: World Bank.

Crepon, B., and G. J. van den Berg (2016). "Active Labor Market Policies". Annual Review of Economics. Vol. 8, pp. 521-546.

Fafchamps, M. and Woodruff, C. (2016)."Identifying Gazelles : Expert Panels vs. Surveys as a Means to Identify Firms with Rapid Growth Potential." World Bank Policy Research Working Paper, No. 7647. World Bank, Washington, DC.

Fox, L. and K. Upaasna (2017)."The evidence in in: How should youth employment program in low income countries be designed?" USAID report, September,2017. Washington, DC.

Grimm, M., and A.L. Paffhausen (2014). "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. IZA DP No. 8193.

Kluve J., D. Puerto S., D. Robalino D., J.M. Romero, F. Rother, J. Stroterau, F. Weikenkaff , M. Witte (2016). "Do Youth Employment Programs Improve Labor Market Outcomes? A Systematic Review". IZA Discussion Paper No. 10263, October 2016.

Loxha, A., and M. Morgandi (2014). " Profiling the unemployed: a review of OECD experiences and implications for emerging economics". Social Protection and labor discussion paper No. SP 1424. Washington, DC: World Bank Group.

McKenzie, D. (2017). "How Effective Are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence". World Bank Policy Research Working Paper No. 8011. World Bank, Washington, DC.

McKenzie, D.(2017b). "Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition." American Economic Review, Vo.107(8): 2278-2307.

McKenzie, D. and D. Sansone (2017). "Man vs. Machine in Predicting Successful Entrepreneurs: Evidence from a Business Plan Competition in Nigeria". World Bank Policy Research Working Paper No. 8271. World Bank, Washington, DC.

Pritchett, L., S. Samji, and J. Hammer (2013). "It's All About MeE: Using Structured Experiential Learning ("e") to Crawl the Design Space." CGD Working Paper 322. Washington, DC: Center for Global Development.

S4YE (2017). "Promising Solutions to Youth Employment: S4YE's Impact Portfolio" S4YE report. Available at: https://www.s4ye.org/sites/default/files/2017-11/Impact%20Portfolio%20Report.pdf

S4YE (2018a). "Integrated Youth Employment Programs: A Stock-take of Evidence on what works in Youth Employment programs." S4YE report, *forthcoming*.

S4YE (2018a). "Integration: A New Approach to Youth Employment Programs." S4YE report, *forthcoming.*

Save the children (2015). Presentation "What's all the talk about data-driven decision-making?" given by Particial Langan, Andrea Lzano, and Arief Firdaus, Sep. 22, 2015. Available at: https://youtheconomicopportunities.org/resource/2638/webinar-resources-whats-all-talk-about-data-driven-decision-making

Wolff, J. (2017). "Borders I the Cloud." Slate Magazine article. Available at:
http://www.slate.com/articles/technology/future_tense/2017/11/countries_are_increasingly_imposing_borders_on_the_cloud.html