Skills Matter

Save the Children

Results from a Cross-Country Analysis of Skills Acquisition and Employment Outcomes

by Larry Dershem, December 2019

A Skills to Succeed Report

Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

I would like to acknowledge support and assistance from Eyerusalem Tessema (SCUS Hunger and Livelihoods Dept.), as well as the staff of S2S programs in Mexico, Indonesia, and Bangladesh in producing this analysis and report.

Skills to Succeed Program
Program Director: **Monica Caminiti**
Specialist, MEAL & Research: **Eyerusalem Tessema**
Specialist, ICT4D: **Steve DiPangrazio**
Program Coordinator: **Melissa Donaher**

Any questions or clarifications about data analysis or content of this report should be directed to Larry Dershem, Senior Advisor, Research & Evaluation, Program Quality & Impact Department, International Programs, Save the Children-US ldershem@gmail.com or Eyerusalem Tessema, Specialist, MEAL and Research, Skills to Succeed, Hunger and Livelihoods Department, Save the Children-US etessema@savechildren.org

EXECUTIVE SUMMARY

This report presents findings from the use of the Employability Assessment Tool (EA Tool) among youth participating in Save the Children's Skills to Succeed (S2S) program in three countries, which focuses on equipping disadvantaged young people to acquire skills, develop networks, and seek opportunities they need to get a job or build a business. The youth involved in SC's S2S Program were both in and out-of-school.

The EA Tool was developed internally, using a rigorous process of local language adaptation, item reliability testing, and test / retest reliability, in 2016 for the S2S Program (Dershem, 2016). The EA Tool was designed to measure changes in basic employability skills that program participants received training and practicums. Moreover, with the EA Tool being a standardized, tested instrument, this enabled the S2S Program to compare in-country and cross-country results, which allows S2S Program to learn and adapt its program model within and across countries where it implements.

The EA Tool measures the overall domain, "employability," based on six sum-domains: 1) Positive Self-Concept, 2) Self-Control, 3) Social Skills, 4) Communication Skills, 5) Higher Order Thinking Skills, and 6) Job Search Skills. Each sub-domain is measured by four questions, thus, the EA Tool is comprised of 24 questions and is used for a pre- and post-assessment to measure changes in S2S's program youth's employability skills. The EA Tool has been adapted to Bahasa (Indonesia), Tagalog (Philippines), Bengali (Bangladesh), Achoi (Uganda), Vietnamese, Khmer (Cambodia), Amharic (Ethiopia), Italian, with Mandarin (Chinese) in the near future.

This report answers the following three questions:

1. What is the internal reliability of the EA Tool at the pre- and post-assessment?
2. What predictors are statistically significantly correlated with increased EA scores?
3. What is the relationship between EA scores and Employment Outcomes?

Data analyses were conducted on data from three S2S countries and five programs: Mexico (Cohort 1 and Cohort 2), Indonesia (Cohort 2), and Bangladesh (Cohort 4). Baseline (Time 1) and Follow-up (Time 2) data were included in the analyses.

SUMMARY OF KEY FINDINGS:

1. The analyses shows that the reliability of the EA Tool overall, and for the six sub-domains, is good to excellent in Mexico. However, in Bangladesh and Indonesia only the overall EA Tool has adequate to good reliability with the most of the sub-domains having poor to very poor reliability.

2. On average, there was very little change from the baseline and follow-up EA Tool scores in all three countries. When change occurred, the follow-up EA Tool score represented a 7% - 9% increase. However, the standard measurement error of the EA Tool (Dershem, 2016) ranges from 7% - 10%, Thus, even though in some countries a 9% increase was statistically significant, nonetheless it is within the standard measurement, and therefore not a meaningful increase. When significant increase in the EA Tool score did occur, it was among certain groups (younger youth) and only for one sub-domain (Job Search Skills).

3. Due to little increase in EA Tool scores at the follow-up, no gender differences were found.

4. In assessing which youth characteristics best predictors of increased EA Tool scores at the follow-up, the only characteristic significantly related, though negatively, was a youth's baseline score. That is, youth with low baseline EA Tool scores had high EA Tool score at the follow-up and, visa versa, youth with high baseline EA Tool scores had low EA Tool scores at the follow-up.

5. With little change in the baseline and follow-up EA Tool scores, significant relationships between increased EA Tool scores and youth characteristics were difficult to find. Of the various youth characteristics examined (gender, age, education, at EA Tool scores) in relationship to gaining employment after completing the S2S program, for youth in Mexico and Indonesia the best predictor having been previously employed. Among Bangladesh youth cohort 4, the only statistically significant predictor of gaining employment after completing the S2S program, when controlling for multiple youth characteristics, was the EA Tool follow-up score, which was positively related; that is, youth with the highest EA Tool follow-up scores were more likely to be employed than youth with lower EA Tool follow-up scores. This finding supports S2S's Theory of Change that youth involved in S2S training are able to gain critical employability soft skills that lead to actual employment outcomes.

6. The amount of time between the baseline and follow-up assessments was almost one year. The general rule for training programs is that the follow-up assessment should occur not long after the training is completed due to a) recall error and b) relative assurance than an increase in the follow-up assessment score is due to the training rather than other factors. The S2S program needs to review the timing of the follow-up assessment to minimize recall error and to be confident that any increase in the EA Tool follow-up score is most likely due to the S2S training than other non-S2S program factors.

1. INTRODUCTION

SC's S2S Program attempts to address the critical challenge of youth employment in Asia and Latin America, especially at-risk youth who are from 16 to 24 years of age by helping them to acquire relevant skills, networks, and opportunities to obtain a decent job or create a business. S2S is designed to address the barriers and challenges that deprived youth face in urban and peri-urban labor markets, which include lack of relevant skills, networks of information and support, and decent job opportunities.

In 2016, the S2S Program conducted a performance evaluation to document its performance to date as well as evaluate it Theory of Change, and identify areas of the program in need of modification. As part of the evaluation, the program attempted to answer the question of whether increased employability skills are associated with S2S employability skill-training curricula. A component of this evaluation was whether increased employability skills are associated with S2S employability training as well as associated with getting a job.

There is a growing evidence base (Kautz and Zanoni, 2014) of the critical role of employability skills (i.e., soft skills) have on predicting employment and earnings, even rivaling academic and technical skills, With the transition to service oriented economies around the world, the demand for such skills has increased over the past 20 years. Nevertheless, a "soft" skills gap remains and is noted by many employers around the world (ILO, 2013).

Over the years, routine monitoring of S2S programs was that S2S employability training was associated with increased employability skills; however, S2S programs lacked a standardized tool and method to measure change in employability skills. In the past, different S2S Programs were using individual tools, which made cross-program and cross-country comparisons nearly impossible. Eventually, the S2S Program team decided to develop a standardized, self-report tool.

Extensive testing was conducted to validate the language adaption and reliability of the EA tool (Dershem, 2016). The EA Tool measures six skills: positive self-concept; self-control; social skills; communication skills; higher-order thinking skills and job search skills, which have an evidence-based relationship to employability, are malleable and can be improved through S2S program and its constituent interventions, and these skills had been measured by other validated tools, both in and outside Save the Children. Based on these extensive tests, the EA Tool was finalized and rolled-out for use in countries with S2S programming.

This report provides additional documentation of the use of the EA Tool and its use in youth employment based on the Skills Matter: Results from a Cross-Country Analysis of Skills Acquisition and Employment Outcomes (Ying, 2018).

The analysis in this report answers the following three questions:

1. What is the internal reliability of the EA Tool at the pre- and post-assessment?
2. What predictors are statistically significantly correlated with increased EA scores?
3. What is the relationship between EA scores and Employment Outcomes?

Analyses were conducted on data from three S2S countries and five programs: Mexico (Cohort 1 and Cohort 2), Indonesia (Cohort 2), and Bangladesh (Cohort 3 and Cohort 4). Baseline (Time 1) and Follow-up (Time 2) data were included in the analyses.

2. Description of EA Tool

The EA Tool consists of 24 questions/items that collectively measure the global domain of "employability", with four questions/items measuring six sub-domains of employability as mentioned above. Youth are provided a questionnaire asking them to assess the degree to which they currently have various employability skills. The EA Tool is administered in Bangladesh and Indonesia to all youth enrolled in S2S programming upon entering the program. The EA Tool takes, on average, 12-15 minutes to complete. The pre-assessment (Time 1) is administers once youth enter the S2S Program and the post-assessment (Time 2) at least 3-months after completion of the program, which means the time from baseline assessment to post-training assessment can be 1 year. In contrast, in Mexico, the S2S Program conducts retrospective administration of the EA Tool (Moore and Tannis, 2009). That is, youth only are administered the baseline (Time 1) and follow-up (Time 2) which is completed AFTER the training (when they have become more aware of aspects of employability). Thus, youth are given the EA Tool and asked to reflect and assess their skills level before entering the program. Once the complete this, they are asked to assess the skills level after completing the training. Both the baseline and follow-up are conducted at the same time, which is after the training has ended.

Table 1: Employability Assessment Tool

Domain	**Employability**	**All 24 skills below.**
6 Sub-domains	Positive Self - Concept	1. I feel valued and appreciated by others. 2. I feel good about my future.

		3. I anticipate my own needs ahead of time. 4. I can adapt to changes by learning new skills.
	Self-Control	5. I'm able to complete assignments in time. 6. I feel proud when I produce high quality work. 7. I go to work even when I feel like staying at home. 8. I follow workplace or school dress codes.
	Social Skills	9. I can understand and work with people of different backgrounds. 10. I accept people who are different than me. 11. I value the input and contributions of others. 12. I take responsibility for what I do.
	Communication Skills	13. I know how to express myself in proper ways. 14. I know how to articulate my own ideas clearly. 15. I read so I can comprehend and use new information. 16. I listen actively to understand and learn.
	Higher Order Skills	17. I collect, analyse, and organize information to find the best solution to a problem. 18. I seek many sources of information to solve a problem in school or at work. 19. I learn from my past successes and mistakes to make future decisions. 20. I can adapt to changing circumstances.
	Job Search Skills	21. I have the knowledge and skills needed to interview for jobs. 22. I know how to prepare a resume. 23. I know how to complete a job application 24. I have the skills and experience valued by employers.

As part of the EA Tool's local adaptation process, each translated version of the EA Tool goes through two pilot administrations to establish statistical reliability before the tool is rolled out to the larger S2S population.

3. Methodology & Analytical Approach

Data for this report were obtained from the baseline and follow-up administration of the EA Tool in three S2S countries (Mexico, Indonesia, and Bangladesh). However, not all youth attending the S2S program in each country was included; thus, this analysis only includes those youth who completed both the baseline (Time 1) and follow-up (Time 2) EA Tool administration.

To assess internal reliability of the overall domain of the EA Tool, employability, and the six sub-domains, Cronbach alpha was used. The degree of reliability was assessed based on the conventional levels of alpha coefficients less than 0.70 indicating low reliability, coefficients from 0.70 to 0.79 indicating adequate reliability, 0.80 to 0.89 indicating good reliability, and coefficients from 0.90 or higher indicating excellent reliability.[1]

To examine predictors of increased EA Tools scores, ordinary least squares (OLS) regression was used with the magnitude of change in the EA Tool score regressed on various youth characteristics.

To examine the relationship between EA Tool scores and employment, binomial regression was used in which employment/non-employment was regressed on various youth demographic, employment experiences, and the EA Tool scores.

4. Study Limitations

S2S implementing partners in each context specified conducted this data collection with minimal supervision from S2S staff. While the partners received training and guidance from SC, it is possible that differences in the way that the partners administered the tool may have led to issues in data quality.

[1] https://en.wikipedia.org/wiki/Cronbach%27s_alpha

Sampling for the follow-up survey in each S2S country was limited to youth who were reachable by S2S implementing partner staff. This may have led to sampling bias, which would have affected the overall analysis.

5. Description of Youth Samples

A description of the characteristics of each youth cohort is discussed below. The only common set of characteristics collected from each of the five cohorts was gender and age. Other characteristics collected for some cohorts were level of education at the follow-up, whether the youth had a child, and various employment experiences, such as having a job or creating their own business prior to attending S2S program.

Mexico (Cohort 1 & Cohort 2)

Cohort 1 – A total of 263 youth were in cohort 1, comprised of 173 (65.8%) girls and 90 (34.2%) boys who ranged from 16 to 20 years of age, with the average age being 17.3 years of age for both girls and boys. Most of these youth had completed a secondary level of education (90.5% or 238) with 22 (8.5%) having only completed a primary level of education. About one-half (53.2%) reported having paid work for more than one month upon entering S2S program, with 15.5% (40) of them having created a business by her/himself or in partnership with someone.

Cohort 2 – A total of 296 youth were in cohort 2, comprised of 144 (48.6%) girls and 152 (51.4%) boys who ranged from 17 to 27 years of age, with the average age being 18.1 years of age. The Mexico country office was unable to match baseline and follow-up data for this cohort, which means that only internal reliability analysis for the baseline and follow-up can be analyzed and reported.

Indonesia (Cohort 2)

Cohort 2 – A total of 215 youth were in cohort 1, which was comprised of 100 (46.5%) are girls and 115 (53.5%) boys who ranged from 16 to 21years of age, with an average being 17.3 years of age. The vast majority of these youth, 202 or 94.0% had completed a secondary level of education with a small number of youth (13 or 6.0%) completing only a primary level of education. Slightly less than one-half (89 or 41.4%) had ever had paid employment.

Bangladesh (Cohort 4)

Cohort 4 – A total of 149 youth were in cohort 4, which was comprised 74 (49.7%) girls and 75 (50.3%) boys, who ranged from 16 to 23 years of age with the average being 18 years of age. Youth's level of education was not collected. Slightly less than one-fifth (18.8% or 28) youth reported having had paid work for more than a month and 14.8% (or 22) had created a business themselves or in partnership with someone.

6. Internal Reliability of EA Tool

Cronbach's alpha is a statistical technique used to estimate the reliability, or internal consistency, of a composite score. Cronbach's alpha gives us a simple way to measure whether or not a score is reliable. It is used under the assumption that you have multiple items measuring the same underlying construct, which is "Employability" for the EA Tool. Cronbach alpha coefficients are generally interpreted as 0.70 – 0.79 as acceptable, 0.80 – 0.89 as good, and 0.90 or greater as excellent.

The data tables below present the Cronbach alpha coefficients for each youth cohort by country using a color-coding of poor reliability (<0.70) with yellow highlighting, acceptable reliability (0.70 – 0.79) with light green highlighting, good reliability (0.80 – 0.89) with green highlighting, and excellent reliability (>=0.90) with dark green highlighting.

Mexico (Cohort 1 & Cohort 2)

Cohort 1 – At the baseline and follow-up, the overall domain, "employability," had an alpha coefficient was greater than 0.90, which indicates excellent reliability. This suggests that the adaptation process of converting the English version of the EA Tool into Spanish language was effective in producing a reliable measurement for employability overall for both girls and boys.

Since Mexico used a retrospective approach when administrating the EA Tool, the baseline and follow-up were both administered to youth after they completed the S2S program.[2] The six employability sub-domains had alpha coefficients ranging from good to excellent reliability for the baseline and the follow-up assessments. The two employability sub-domains with the lowest alpha coefficients (in the mid-0.70s) during the follow-up, though still an acceptable reliability, were "Self-Control" and "Positive Self-Confidence." Overall, alpha reliabilities increased slightly during the follow-up assessment. In summary, the reliabilities of the sub-domains are good to excellent and thus can be used individually to assess change and association with outcomes in addition to the overall domain, employability.

Table 2: Mexico Cohort 1 EA Tool Domain & Sub-Domain Reliabilities (N=263)

EA Tool		**Girls** (n=173)		**Boys** (n=190)		**Total** (N=263)	
		Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up
Domain	**Employability**	**0.96**	**0.97**	**0.96**	**0.95**	**0.96**	**0.96**
Sub-domains	Positive Self-Concept	0.85	0.90	0.79	0.84	0.83	0.88
	Self-Control	0.77	0.84	0.81	0.79	0.78	0.82
	Social Skills	0.89	0.91	0.87	0.87	0.88	0.90
	Communication Skills	0.85	0.90	0.87	0.84	0.86	0.88
	Higher Order Thinking Skills	0.84	0.88	0.86	0.86	0.84	0.87
	Job Search Skills	0.89	0.88	0.87	0.92	0.88	0.89

Cohort 2 – This cohort also used a retrospective baseline approach. And, similar to cohort 1 from Mexico, at the baseline and follow-up, the overall domain, "employability," had an alpha coefficient was greater than 0.90, which indicates excellent reliability. This finding reinforces that the adaptation process of converting the English version of the EA Tool into Spanish language was effective in producing a reliable measurement for employability overall for both girls and boys.

Table 3: Mexico Cohort 2 EA Tool Domain & Sub-Domain Reliabilities (N=296)

EA Tool		**Girls** (n=144)		**Boys** (n=152)		**Total** (N=296)	
		Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up
Domain	**Employability**	**0.96**	**0.97**	**0.93**	**0.97**	**0.95**	**0.97**
Sub-domains	Positive Self-Concept	0.84	0.89	0.73	0.90	0.79	0.90
	Self-Control	0.82	0.85	0.73	0.85	0.78	0.85
	Social Skills	0.87	0.89	0.88	0.93	0.88	0.92
	Communication Skills	0.86	0.90	0.76	0.91	0.81	0.90
	Higher Order Thinking Skills	0.83	0.85	0.77	0.89	0.81	0.88
	Job Search Skills	0.91	0.88	0.88	0.89	0.90	0.90

[2] A retrospective baseline is conducted to reduce response shift-bias in which youth "over rate" themselves at the baseline and then once completing the training/program, and learning how much they did not know before entering the training/program, rate themselves the same or lower at the follow-up. Thus, a retrospective baseline is used to reduce "no change" or "negative change" in scores. However, these results seem to indicate that a retrospective baseline may also improve internal reliability of the overall EA Tool as well as the six sub-domains since after completing the training youth are more familiar with employability concepts, terms, and phrases.

The six employability sub-domains had alpha coefficients ranging from good to excellent reliability for the baseline and the follow-up assessments. The employability sub-domain with the lowest alpha coefficients (mid 0.80s) though still an acceptable reliability at the follow-up was "Self-Control." Overall, alpha reliabilities increased slightly during the follow-up assessment. Like cohort 1, the reliabilities of the sub-domains are good to excellent and thus can be used individually to assess change and association with outcomes in addition to the overall domain, employability.

Indonesia (Cohort 2)

Cohort 2 **–** The time interval between the baseline and the follow-up assessment varied, due to different implementing partners, but in general the follow-up assessment was administered one year after the baseline assessment. The Cronbach alpha coefficient at the baseline and follow-up for the overall domain, "employability," was good for both girls and boys (0.84 – 0.88). However, of the six sub-domains only "Job Search Skills" obtain an adequate level of reliability for girls and boys at the baseline and follow-up with the remaining five sub-domains having poor reliability (alpha reliability 0.69 or below).

Table 4: Indonesia Cohort 2 EA Tool Domain & Sub-Domain Reliabilities (N=215)

EA Tool		**Girls** (n=100)		**Boys** (n=115)		**Total** (N=215)	
		Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up
Domain	**Employability**	**0.85**	**0.84**	**0.88**	**0.88**	**0.87**	**0.86**
Sub-domains	Positive Self-Concept	0.55	0.51	0.51	0.58	0.53	0.56
	Self-Control	0.54	0.38	0.59	0.67	0.57	0.57
	Social Skills	0.65	0.67	0.68	0.62	0.66	0.65
	Communication Skills	0.73	0.49	0.59	0.67	0.66	0.59
	Higher Order Thinking Skills	0.51	0.60	0.69	0.64	0.62	0.62
	Job Search Skills	0.73	0.82	0.80	0.87	0.77	0.85

The two sub-domains with the poorest reliabilities were "Self-Control" and "Communication Skills," especially for girls during the follow-up (0.38 and 0.40 respectively). These results indicate much more refining of the local language adaptation needs to occur for these five sub-domains, especially among girls. Based on these alpha coefficients, only the overall "Employability" and the sub-domain, "Job Search Skills," scores should be analyzed to assess the performance of the S2S programs. Not until the alpha reliabilities are 0.70 or greater can the other five sub-domains be used individually in analysis of change from baseline and follow-up.

Bangladesh (Cohort 4)

Cohort 4 – The time interval between the baseline and the follow-up assessment was one year. The Cronbach alpha coefficient at the baseline and follow-up for the overall domain, "employability," was good for both girls and boys (0.90 or larger). Of the six sub-domains, Self-Control had very poor reliability (0.28 - 0.45) for girls and boys during the baseline and follow-up assessment. Similarly, Positive Self-Concept never achieved an adequate level of reliability (0.70 or more) for girls and boys as well as at the baseline and follow-up assessments. The three sub-domains which had adequate to good reliabilities for girls and boys at both the baseline and follow-up are Social Skills, Higher Order Skills, and Job Search Skills.

Table 5: Bangladesh Cohort 4 EA Tool Domain & Sub-Domain Reliabilities (N=149)

EA Tool		Girls (n=74)		Boys (n=75)		Total (N=149)	
		Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up
Domain	**Employability**	**0.91**	**0.91**	**0.90**	**0.91**	**0.90**	**0.91**
Sub-domains	Positive Self-Concept	0.56	0.51	0.67	0.64	0.62	0.59
	Self-Control	0.52	0.39	0.28	0.51	0.43	0.45
	Social Skills	0.75	0.74	0.75	0.74	0.75	0.74
	Communication Skills	0.72	0.68	0.76	0.67	0.74	0.67
	Higher Order Thinking Skills	0.74	0.70	0.75	0.72	0.75	0.71
	Job Search Skills	0.88	0.82	0.83	0.82	0.86	0.82

For the Bangladesh cohort 4 youth, only the overall domain of "Employability" and the three sub-domains of Social Skills, Higher Order Skills, and Job Search Skills could be used to for assessing program outcomes.

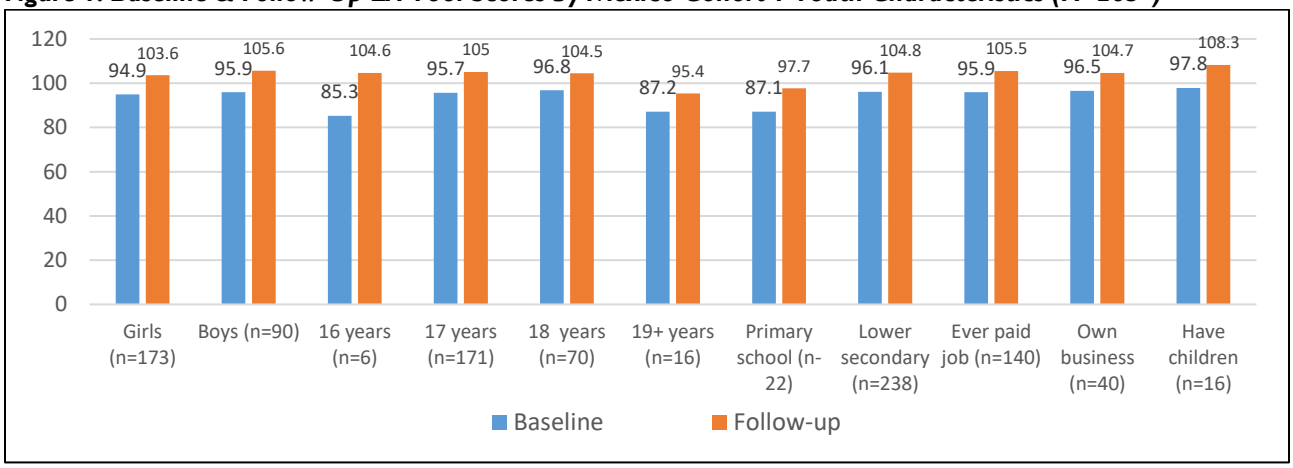
7. YOUTH CHARACTERISTICS & EA TOOL SCORES

As mentioned earlier, various youth characteristics were recorded by the S2S programs in these three countries, which included most commonly the characteristics of gender, age, level of education, having a child/ren, if the youth every had paid job, and if the youth ever started her/his own business. Thus, the association between these characteristics and the EA Tool scores are examined below.

Mexico (Cohort 1 & Cohort 2)

Cohort 1 – The following characteristics for the Mexico cohort 1 youth were examined in relationship to baseline and follow-up EA Tool scores: gender, age, level of education, having children, if the youth ever had paid job, and if the youth ever started her/his own business. As expected, on average, baseline scores were lower than follow-up scores.

Figure 1: Baseline & Follow-Up EA Tool Scores by Mexico Cohort 1 Youth Characteristics (N=263*)



* 2 missing cases for "ever paid job" and 5 missing cases for "own business".

The average baseline score was 95, which is 79% of the total EA Tool score possible (120) and slightly increased to an average score of 104 at the follow-up study or 87% of the total EA Tool score possible. Overall, the average increase in the EA Tool was about 9% - 11%, except for the 6 youth who were 16 years of age who experienced a 23% increase (85.3 at baseline and 104.6) at follow-up). The overall change from the baseline score (95) and follow-up score (104) was statistically significant (t-value=11.24, p<0.00). However, based on test/retest analysis (Dershem, 2016) during the development of the EA Tool, the EA

Tool has a standard error of measurement ranging from 7% - 8%.[3] Thus, EA Tool scores must increase greater than approximately 9% to begin measuring "real change." For all youth, except those 6 youth who were 16 years of age, the percentage increase in EA Tool score at the follow-up over the baseline scores (9% to 11%), which means that little meaningful change occurred in the follow-up scores. In short, even though the amount of increase in the EA Tool scores was statistically significant, nonetheless, the proportion of change was close to the standard error of measurement for the EA Tool, indicating little change.

Examining baseline and follow-up scores more closely, not all youth had a follow-up score higher than their respective baseline score. That is, 16.3% (or 43) youth had a follow-up score lower than their baseline score and 10.3% (or 27) youth had similar baseline and follow-up scores. Girls, more than boys, were more likely to have no change in baseline and follow-up scores. Of the six sub-domains, on average for both girls and boys, baseline and follow-up scores were the same (no change) for Self-Control and only a very small increase (0.1) for Social Skills, which as not statistically significant. The largest increase from baseline to follow-up scores was for Job Search Skills; that is, a baseline score of 14.0 and a follow-up score of 16.9, or a 20.7% increase, which is substantially larger than the EA Tool's standard error of measurement as well as statistically significant (t-value=13.13, $p<0.00$).

In the table below, OLS regression was used to regress change in EA Tool scores on six youth characteristics (gender, age, level of education, ever had job, every had business, and baseline score) to examine which, if any, were significantly related to change in the EA Tool score. The only youth characteristic statistically significantly related to the change in baseline and follow-up EA Tool scores was the baseline EA Tool score. That is, youth who scored the lowest at the baseline experienced the greatest increases, when statistically controlling for all other youth characteristics. In other words, youth with the highest EA Tool scores at the baseline experienced negative or no change at the follow-up. This result occurred for the other youth cohorts discussed below.

With the baseline score as the only significant predictor of change in EA Tool score, this result could be due to several factors: 1) a substantial proportion of youth who entered the S2S program had high levels of employability skills and, thus, only those youth with the lowest levels of employability skills at the baseline increased at the follow-up, which means there were very different levels of employability skills among youth in this cohort, which is why baseline EA Tool score results should be examined prior to programming so that appropriate and relevant training is provided to each group or 2) even with the use of a retrospective approach, this youth cohort, nonetheless, a degree of overestimation of skills level at the baseline occurred resulting in youth assessing their skills level had not changed or decreased (negative change) at the follow-up.[4]

[3] The only test / retest reliability, which establishes the standard error of measurement, of the EA Tool was conducted in the Philippines in 2006. Each S2S country office would need to conduct a test / retest in order to establish the standard error of measurement for their version of the EA Tool; however, until such test / retest reliabilities are conducted in each country the original test / retest reliability (7%-8%) is assumed to apply to all versions of the EA Tool.

[4] Many researchers and evaluators attribute this type of negative change mostly, but not solely, to the "overconfidence effect." The Overconfidence Effect refers to the subjective confidence a person has in their own ability is greater than their objective (actual) ability. In terms of S2S program, this means that youth at the baseline assessment, due to some degree of overconfidence in their abilities, rate themselves relatively high on the six sub-domains, which results in a relatively high EA Tool score. Then, after attending and learning from the S2S trainings and activities, the youth, reflecting on what they have learned, realizes s/he rated her/himself to high and at the follow-up rates her/himself lower or at the same level.

Table 6: Linear (OLS) Regression of Change in Baseline/Follow-up EA Tool Score on Mexico Cohort 1 Youth Characteristics (N=263)

Youth Characteristic	Statistical significance
Age	Not significant
Gender	Not significant
Level of education	Not significant
Every had job	Not significant
Ever own business	Not significant
EA Tool Baseline Score	**Yes – negative relationship (p< 0.05)**

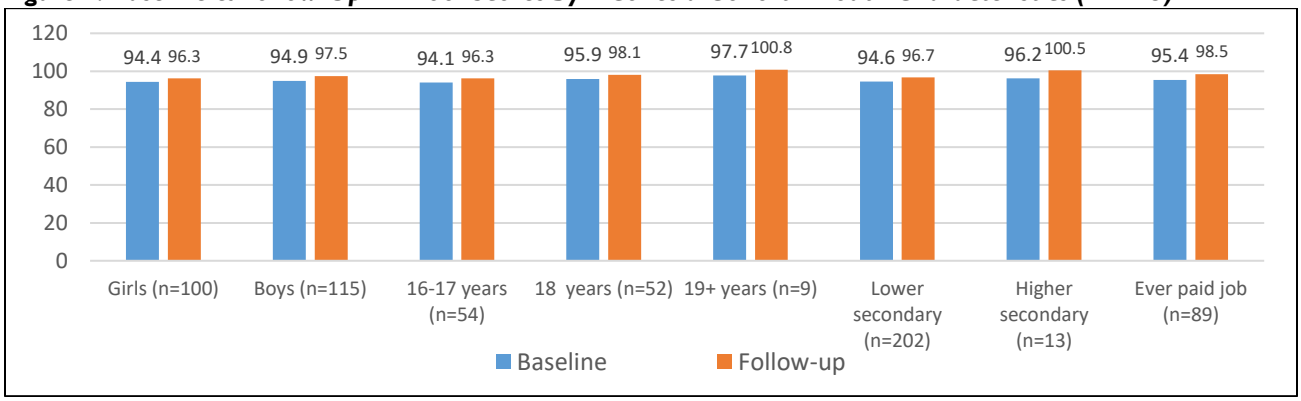
Cohort 2 – Due to several reasons, the country office was unable to match baseline and follow-up data for each youth in this cohort. Without matching the baseline and follow-up scores it is not possible to examine change in EA Tools scores. Therefore, it is not possible to examine change in EA Tool scores by youth characteristics.

Indonesia (Cohort 2)

Cohort 2 – Among the Indonesia cohort 2 youth, the average score for Employability at the baseline was 94.7 and at the follow-up was 96.9, which represents a 2.3% increase. This increase is statistically significant (t-value=3.95, p<0.00), however, with the standard error of measurement of the EA Tool ranging from 7% - 8% this indicates that basically, on average, no substantial change occurred among these youth in employability skills. None of the follow-up scores increased greater than 8% over baseline scores regardless of youth characteristics.

One contributor to the overall EA Tool score not significantly increasing from baseline to follow-up was that a relatively larger percentage of youth, 36.4% (56), had a follow-up score that was lower than their baseline score. In addition, 9.1% (14) of youth had the same EA Tool score at the follow-up as they had at baseline. Thus, the EA Tool score remained the same or declined for slightly less than one-half (45.5% or 70) of all youth in this cohort.

Figure 2: Baseline & Follow-Up EA Tool Scores by Indonesia Cohort 2 Youth Characteristics (N=215)



Since the sub-domain, Job Search Skills, was the only sub-domain with an adequate level of reliability it is possible to examine whether was significant change in this sub-domain. Overall, there was a 10% increase in follow-up over the baseline score (13.1 and 11.9 respectively), which is a statistically significantly difference (t-value=6.41, p<0.00) and slightly more than the standard error of measurement for the EA Tool. Among the youth characteristics, only youth 16-17 years of age had an increase in Job Search Skills (11% increase) greater than the standard error of measurement. In summary, Job Search Skills increased at the follow-up but only slightly once the measurement error was accounted for.

Using OLS regression, the change in the EA Tool scores, even though very small, was regressed on various youth characteristics to see if any were significantly associated with this slight change. The table below shows which youth characteristics were significantly related to changes in the EA Tool scores.

Table 7: Linear (OLS) Regression of Change in Baseline/Follow-up EA Tool Score on Indonesia Cohort 2 Youth Characteristics (N=215)

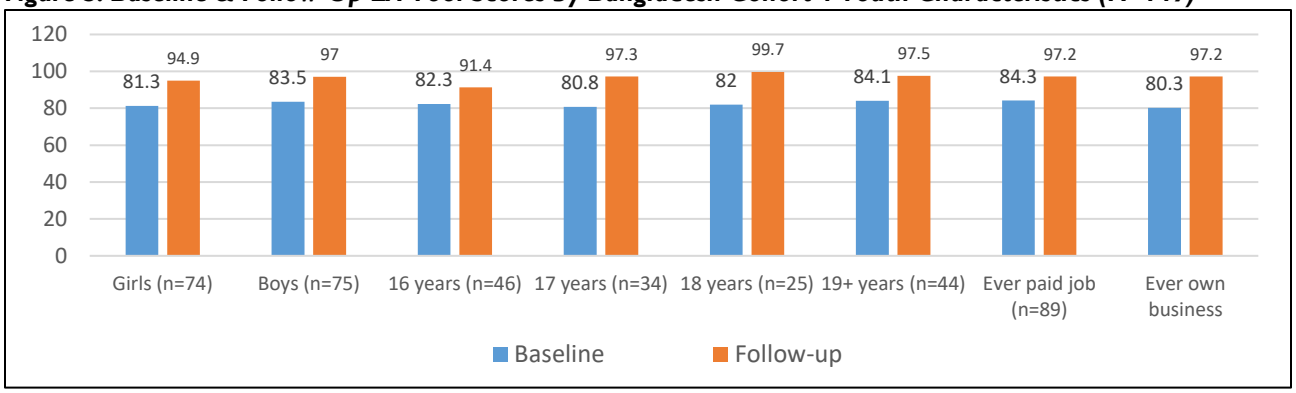
Youth Characteristic	Statistical significance
Age	Not significant
Gender	Not significant
Level of education	Not significant
Ever had a paid job	**Yes - positive (p< 0.05)**
EA Tool Baseline Score	**Yes - negative (p<0.00)**

Thus, similar to Mexico, those youth with some change in their EA Tool scores, , baseline EA Tool score was inversely correlated with change in EA Tool score - meaning lower baseline scores were associated with larger changes in scores, and higher baseline scores were associated with smaller changes in scores. Additionally, having a previous paid job was positively associated with increases in EA Tool score from baseline to follow-up.

Bangladesh (Cohort 4)

Cohort 4 – Among the Bangladesh cohort 4 youth, the average score for Employability at the baseline was 82.4 and at the follow-up was 95.9, which represents a 16.4% increase. This increase is statistically significant (t-value=10.34, p<0.00) and is substantially larger than the standard error of measurement of the EA Tool (7% - 8%). The increase in the baseline and follow-up scores were, on average, the same for girls (16.7%) and boys (16.2%). Among the three other characteristics, the largest increase in the EA Tool score (21.6%) was reported by youth 18 years of age followed by youth who had ever created a business (21.0%).

Figure 3: Baseline & Follow-Up EA Tool Scores by Bangladesh Cohort 4 Youth Characteristics (N=149)



Despite these overall increases on average, 17.4% (26) of these youth had follow-up scores lower than their baseline scores representing a decline in EA Tool score. Only 1.3% (2) experienced no change in that their baseline and follow-up scores were identical.

Three sub-domains were sufficiently reliable at the baseline and follow-up (Social Skills-0.74, Higher Order Thinking Skills - 0.71, and Job Search Skills - 0.82) to asses change. In rank order, the largest increases on average was for Job Search Skills (30.2%), which was statistically significant (t-value=10.62, p<0.00), followed by Higher Order Thinking Skills (15.2%) which was statistically significant (t-value=7.09, p<0.00), and Social Skills (12.8%) which was statistically significant (t-value=7.77, p<0.00). For Job Search Skills, girls experienced a slightly higher increase (33.6%) over their baseline score than boys (28.2%), but this difference was not statistically significant.

Using OLS regression, the change in the EA Tool scores, was regressed on various youth characteristics to see if any were significantly associated with this slight change. The table below shows which youth characteristics were significantly related to change in the EA Tool scores. Regardless of gender, age, level of education or ever having a paid job, the most significant factor related to an increase in the EA Tool score at the follow-up was the baseline score. That is, youth with the highest baseline scores had the smallest increase in follow-up scores, and youth with the lowest baseline scores had the largest increase in the follow-up scores.

Table 8: Linear (OLS) Regression of Change in Baseline/Follow-up EA Tool Score on Bangladesh Cohort 4 Youth Characteristics (N=149)

Youth Characteristic	Statistical significance
Age	Not significant
Gender	Not significant
Level of education	Not significant
Ever had a paid job	Not significant
EA Tool Baseline Score	**Yes – negative (p<0.00)**

8. Employability Skills and Employment Outcomes

The follow-up survey, which was conducted approximately 3-6 months after youth completed the program, asked youth if they had a paid job for longer than 1 month since completion of the program and if they were currently working in that job. Probit regression was used to examine the relationship between this employment outcome and the EA Tool follow-up score as well as demographic and employment experience characteristics. The tables below present the results from probit regressions.

Mexico (Cohort 1 & Cohort 2)

Cohort 1 – Of the 263 Mexico cohort 1 youth, 150 (57.0%) reported gaining paid employment after 3-6 months after the S2S program, with 109 (41.3%) reporting remaining unemployed,[5] and 4 (1.6%) not responding to this question. Moreover, 13 youth did not respond to one or more of the characteristics in the regression and thus were not included in the analysis. If youth answered yes to whether they had a paid job for longer than 1 month since completion of the program, they were also asked whether they were still currently working in that job.

Since the increase in EA Tool scores were slightly more than the standard error of measurement for the EA Tool (i.e., low variance) it is not surprising that many youth characteristics were not statistically significantly related to gaining employment. Of the six youth characteristics analyzed, the only youth characteristic statistically significantly related to these youth gaining employment was if a youth had ever had a paid job prior to joining the S2S program. As shown in the table below, youth who entered the S2S program having previously worked had a 138%-point likelihood of gaining employment after graduating from S2S program comparted to youth without any previous job experience.

Table 9: Probit Regression Results Predicting Whether Youth in Mexico Cohort 1 Gained Employment at Follow-up (N=254*)

Youth Characteristic	Percentage-Point Likelihood in having a paid job since graduation from S2S Program	Statistical significance
Age	18.2	Not significant
Level of Education	32.9	Not significant
Gender	34.1	Not significant
Paid Job Ever	**138.0**	**Yes (p<0.000)**

[5] The 41 youth who pursued further education at the end of the S2S program in this cohort were included in the unemployed category for the probit analysis.

Attempted Own Business	44.1	Not significant
EA Tool Follow-up Score	0.7	Not significant

* 13 missing cases due to no data for one or more characteristics.

Cohort 2 – As mentioned earlier, due to several reasons, the country office was unable to match baseline and follow-up data for each youth in this cohort. Without matching the baseline and follow-up scores it is not possible to examine change in EA Tools scores. Therefore, it is not possible to examine the relationship between EA Tool scores and gaining employment.

Indonesia (Cohort 2)

Cohort 2 – Of the 215 Indonesia cohort 2 youth, 142 (66.0%) reported gaining paid employment 3-6 months after the S2S program with 73 (34.0%) reporting remaining unemployed.

Table 10: Probit Regression Results Predicting Whether Youth in Indonesia Cohort 2 Gained Employment at Follow-up (N=215)

Youth Characteristic	Percentage-Point Likelihood in having a paid job since graduation from S2S Program	Statistical significance
Age	8.0	Not significant
Level of Education	- 20.0	Not significant
Gender	31.0	Not significant
EA Tool Follow-up Score	- 1.5	Not significant
Paid Job Ever	**90.0**	**Yes (p<0.00)**

Since the increase in EA Tool scores were slightly more than the standard error of measurement for the EA Tool (i.e., low variance) it is not surprising that many youth characteristics were not statistically significantly related to gaining employment. The table above shows that of the five youth characteristics analyzed, the only youth characteristic statistically significantly related to these youth gaining employment was if a youth had ever had a paid job prior to joining the S2S program. As shown in the table below, youth who entered the S2S program having ever had a paid job had a 90%-point likelihood of gaining employment after graduating from S2S program compared to youth without any previous job experience.

Bangladesh (Cohort 4)

Cohort 4 – Of the 149 Bangladesh cohort 4 youth, 69 (46.3%) reported gaining paid employment 3-6 months after the S2S program with 65 (43.6%) reporting remaining unemployed, and 15 (10.1%) not responding to this question. The table below presents the probit regression results in which being employed after the S2S program was regressed on five youth characteristics for the 134 youth who responded to the employment question.

Table 11: Probit Regression Results Predicting Whether Youth in Bangladesh Cohort 4 Gained Employment at Follow-up (N=134*)

Youth Characteristic	Percentage-Point Likelihood in having a paid job since graduation from S2S Program	Statistical significance
Age	18.4	Not significant
Gender	- 23,6	Not significant
EA Tool Follow-up Score	**4.9**	**Yes (p<0.01)**
Paid Job Ever	50.1	Not significant
Ever created own business	32.6	Not significant

*15 missing cases on question about current employment.

Of the five youth characteristics, the only statistically significant predictor for gaining employment was the EA Tool follow-up score. That is, youth with the highest follow-up scores were more likely to be employed

after S2S program than youth with the lowest EA Tool scores. As indicated in section 8 above, of the six sub-domains, the largest increase in follow-up scores for Bangladesh cohort 3 was for Job Search Skills.

9. Summary & Recommendations

Summary

Internal Reliability – These results show that the EA Tool has good to very good internal reliability as-a-whole, but less reliability by sub-domain. In Mexico both the overall EA Tool as well as sub-domains had good to excellent reliabilities and, thus, the S2S program can assess outcomes from the overall EA Tool scores and by each sub-domain, which is most likely due to a good local language adaptation process and somewhat by using a retrospective baseline approach. However, in Bangladesh and Indonesia, only the overall EA Tool score has adequate reliability (0.70 or more) to assess change but not by sub-domains. In Bangladesh the EA Tool requires more work at local language adaptation to increase internal reliability for three of the six sub-domains (Positive Self-Concept, Self-Control, and Communication Skills). In Indonesia, much more work is needed at local language adaptation since all sub-domains, except Job Search Skills, have poor reliabilities at both the baseline and follow-up.

Change in EA Tool Scores – Of the three countries and four youth cohorts examined, the percentage increase in the follow-up over the baseline score was only significant among the Bangladesh cohort 4 youth. All other youth cohorts had follow-up scores that were within the EA Tools standard error of measurement (7% - 8%). When examining the six sub-domains, the only one that showed substantial increase among most youth cohorts was "Job Search Skills." These findings suggest that S2S employability training was not effective in increasing the vast majority of the five sub-domains of Self-Concept, Self-Control, Social Skills, Communication Skills, and Higher Order Thinking Skills. These results appear to indicate that most S2S employability training focuses on job search skills since this is the sub-domain that showed the largest increase, though not enough to increase the overall EA Tool score beyond the EA Tools standard error of measurement. Other possible reasons for little change in baseline and follow-up scores are that training on the intangible soft skills are not implemented by S2S staff as well as tangible job search skills or that youth have a greater challenge in assessing change of intangible soft skills.

Predicting Change in EA Tool Score – When regressing the change in EA Tool scores, from baseline to follow-up, on youth characteristics, the only significant predictor of the follow-up EA Tool score for all youth cohorts was the baseline EA Tool score, which showed a significant but negative relationship. That is, youth who had a low EA Tool score at the baseline had a higher EA Tool score at the follow-up and visa versa that youth with high EA Tool scores at the baseline had lower EA Tool scores at the follow-up. (One explanation for this finding will be discussed in the recommendations section below.) The only exception was for youth cohort 2 in Indonesia, where a youth "ever having a paid job" was statistically significantly related to positive change in the EA Tool score, which suggest that employment experiences can enhance employability learning.

EA Tool Scores and Gaining Employment – With little change in the baseline and follow-up EA Tool scores, significant relationships between increased EA Tool scores and youth characteristics were difficult to find. Of the various youth characteristics examined (gender, age, and education, at EA Tool scores) in relationship to gaining employment after completing the S2S program, for youth in Mexico and Indonesia the best predictor of employment status after completing the S2S program was whether youth had been previously employed. Among Bangladesh youth, the only statistically significant predictor of gaining employment was the EA Tool follow-up score, which was positively related. That is, youth with higher EA Tool scores at the follow-up assessment were more likely to be employed. As presented earlier, youth cohort 4 in Bangladesh reported significantly higher EA Tool at the follow-up compared to the baseline assessment, of which the largest increase occurred for Job Search Skills followed by Higher Order Thinking

Skills and then Social Skills. This finding supports S2S's Theory of Change that youth involved in S2S training gain critical employability soft skills that lead to actual employment outcomes.

Recommendations

1. **Addition Local Language Adaptation is Needed to Refine the Wording of EA Tool Questions in Bangladesh and Indonesia to Increase Reliability** – To accurately measure change in EA Tool scores, internal reliability must achieve an adequate to excellent rating. Clearly, the Cronbach alpha coefficients of internal reliability for the overall EA Tool in Bangladesh (0.90) and Indonesia (0.87) are quite good; however almost one-half of the sub-domains have poor to very poor reliabilities. If the S2S program in these two countries would like to be able to assess change by sub-domain then additional local language adaptation and pilot-testing needs to occur. Based on these results, only EA Tool in the S2S Mexico program can assess both overall changes in employability skills and changes at the six sub-domains levels. However, in Bangladesh and Indonesia ONLY the overall EA Tool score can be used to assess change.

2. **Take the EA Tool's Standard Measurement Error into Account when Assessing Statistically Significant Change in Scores** – In all three countries, the change from baseline to follow-up in the EA Tool scores represented a 7% to 11% increase. As reported in Dershem (2016) based on several test/retest studies, the EA Tool has a standard measurement error ranging from 7% to 9%. Thus, even though in some cases a 7% increase may be statistically significant, based on the EA Tool's standard error of measurement, this increase is most likely not meaningful (actual) change since it could be due to measurement error. That said, there were cases in which the percentage change from the baseline and follow-up EA Tool scores was greater than the standard error of measurement, such as 6 youth who were 16 years of age in Mexico's S2S program who experienced a 23% increase in their EA Tool score. In such cases, the S2S program can cite as "meaningful" change.

3. **Addressing Lower Follow-up Scores than Baseline Scores (negative change)** – In all of the countries a percentage of youth had an EA Tool follow-up score that was lower than her/his baseline EA Tool score, which represents a decline in EA score after attending the S2S program. Negative change in these countries ranged from a low of 16% in Mexico's cohort 1 to 35% in Indonesia's cohort 2.

   This finding is common among training program evaluation (Klatt & Powell, 2005, Moore & Tananis, 2009). In addition, Dershem (2016) reported this finding in the development of the EA Tool due to the "ceiling-effect" in which youth rated themselves high to very high on the six sub-domains at the baseline, prior to participating in the S2S program, which would limit potential increase in the EA Tool during the follow-up.

   Many researchers and evaluators attribute this type of negative change mostly, but not solely, to the "*overconfidence effect*." The *Overconfidence Effect* refers to the subjective confidence a person has in their own ability is greater than their objective (actual) ability. In terms of S2S program, this means that youth at the baseline assessment, due to some degree of overconfidence in their abilities, rate themselves relatively high on the six sub-domains, which results in a relatively high EA Tool score. Then, after attending and learning from the S2S trainings and activities, the youth, reflecting on what they have learned, realizes s/he rated her/himself to high and at the follow-up rates her/himself lower or at the same level.

   Negative change does not necessarily mean the program was not successful in building employability skills. Rather, often negative change represent a "*response-shift*." *Response Shift* occurs when youth's self-rating changes between the baseline and follow-up assessments due to the influence of the S2S program training.

One method to control for *Response Shift* bias is to use a retrospective pre-test, which involves administering the baseline and follow-up assessments at the same time, such as after the completing of the training. Klatt and Powell (2005) provide a brief overview of a retrospective pre-test (baseline), describing both strengths and weaknesses, as well as designing "post-test then pre-test" questions. No doubt, using a retrospective pre-test will not eliminate negative change, but it has the potential, when done correctly, to resolve much of it.

The S2S program may want to consider selecting either Bangladesh or Indonesia to use a retrospective baseline for a cohort and see if there is a reduction in the percentage of youth with "no change" or "negative change" at the end for the training as well as improving internal reliability. The Mexico country office could support in this effort either via Skye or a TDY.

4. **Reviewing S2S' Program Content to Address Little Change in EA Tool's Sub-Domains** – The six domains of the EA Tool are meant to measure changes due to training and activities S2S youth experience when participating in a S2S program. In all three countries, the findings in this report show little to no increase occurred in the six sub-domains. The one sub-domain which showed some increase was Job Search Skills. These findings suggest that the S2S program training and activities, in these countries, are focused primarily on introducing, refreshing and/or reinforcing job search skills more so than the other five skills of positive self-concept, self-control, social skills, communication skills, and higher order thinking skills or that the training is not successfully improving these skills in youth. If this is the case, then in order for S2S to show increases in these sub-domains the design, curriculum, and quality of implementation of S2S program will need to be reconsidered.

5. **Decreasing the Amount of Time Between Baseline and Follow-up Assessment –** One of the main objectives of the EA Tool in S2S programs is to assess the quality and effectiveness of the training youth receive. The general rule for assessing a training is that the follow-up assessment of a training should occur not long after the completion of the training because 1) a participant has better recall of what they have learned and 2) if an increase in learning occurred then this increase can reasonably attributed to the training.

   All the follow-up assessments in this report were conducted almost one year after the baseline was conducted, which raises several issues. First, is how accurately S2S participants can recall how much they learned from a training that may have occurred more than 6 to 8 months earlier? Recall error would most likely decrease the follow-up scores. Second, over a one year period of time, most likely non-S2S factors (maturation, social media, etc.) influenced youth's employability skills and, thus, any increase in the EA Tool at the follow-up may or may not be attributable to the S2S training.

   Recall and contribution of the S2S training are problematic issues when the follow-up assessment is conducted long after the training is completely regardless of whether a prospective or retrospective baseline approach is used. Therefore, the S2S program should determine an optimal time to conduct the follow-up assessment in order to ensure youth can reasonably recall what they learned and that S2S can reasonably claim that the S2S training was one of the main contributors to any increase in the EA Tool follow-up score.

References

Dershem, Larry (2016), "*Skills to Succeed Employability Assessment Tool*," report to the SCUS' Hunger and Livelihoods Department, Washington, DC.

International Labor Organization (2013). "*Enhancing youth employability: What? Why? And How? Guide to core work skills*". Retrieved from https://www.ilo.org/wcmsp5/groups/public/---ed_emp/---ifp_skills/documents/publication/wcms_213452.pdf

Kautz, T., et al. (2014). "*Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success*", OECD Education Working Papers, No. 110, OECD Publishing, Paris, Retrieved from: https://www.oecd.org/education/ceri/Fostering-and-Measuring-Skills-Improving-Cognitive-and-Non-Cognitive-Skills-to-Promote-Lifetime-Success.pdf

Klatt, John and Ellen Taylor-Powell. (2005). *Synthesis of Literature Relative to the Retrospective Pretest Design*, Panel presentation for 2005 Joint CES/AEA Conference, Toronto, October 29.

Klatt, John & Ellen Taylor-Powell. (2005) *Program Development and Evaluation. Using the Retrospective Post then-Pre Design, Quick Tips #27*. University of Wisconsin-Extension, Madison, WI.

Klatt, John & Taylor-Powell. (2005) *Program Development and Evaluation. Quick Tips #28: Designing a Retrospective Post-then-Pre Question*. University of Wisconsin – Extension, Madison, WI.

Moore, Debra and C. Tananis. (2009). *Measuring Change in a Short-Term Educational Program Using a Retrospective Pretest Design*, 2009American Journal of Evaluation 30(2):189-202.

Ying, Christopher. 2018. *Results From a Cross-Country Analysis of Skills Acquisition and Employment Outcomes*. A report to Skills to Succeed Program, Hunger & Livelihoods Department, Save the Children-US, Washington, DC.