

Measuring skills from education to the labour market

Author: Milward, Kirsty: milward.bose@gmail.com

Overview

There are a number of challenges and methodological issues for measuring skills acquisition. Socio-emotional and other non-cognitive skills are particularly challenging for measurement because they are intangible and difficult to isolate from a web of interactions and contextual factors. Nevertheless, a number of instruments have been developed, many of them with high income country contexts in mind.

Non-cognitive skills are often measured using self-reporting models employing tools such as multiple-question surveys before and after the intervention. The clear challenge for these approaches is that they are subject to respondent bias, as they rely on respondents' opinions on the level and types of skill present. To address respondent bias, some task-based and scaling tests – which can be observed/assessed by a third party – have been developed for non-cognitive skills, such as the Spanish Social and Personal Competencies Scale (CPS), among others.

Some of these tests using a question-answer format are be similar to surveys in seeking participant's responses to questions related to different skills. These tests, however, attempt to aggregate proxy behaviours, attitudes and other characteristics to observe a change in skill level.

One risk with task based tests is that they can be subject to circularity if effort and motivation are among the non-cognitive skills being tested, because completing the test requires effort. It is therefore important to standardise for the other skills that contribute to performance.

Random control trial (RCT) designs for skills interventions have the advantage of circumventing bias issues if properly administered, and of potentially establishing clear causation between an intervention and an outcome. But these are also subject to some challenges, such as that they are commonly at risk of selection contamination because – for example – individuals assigned to a control group may be offered a place if one becomes available. In addition, RCTs cannot by themselves establish why outcomes (such as raised levels of employment in better conditions) are caused, or why there is variation in across similar interventions and contexts – such as that in some contexts levels of employment may rise more among young men than young women. Qualitative research and monitoring over time can help to improve understanding of causation and variation.

Implications for MUVA

An RCT intervention design that clearly distinguishes the different components of the programme and allows comparison of results across different combinations of components will contribute to the knowledge base on what works for FEE with young women. However, care will be needed that selection processes for the intervention do not allow for contamination. An RCT design which can reveal variation in the different Mozambique contexts of implementation would enrich this knowledge. However, qualitative research will also be necessary to understand the reasons for context-based variation, and to enrich the picture of causation from input to outcome. Analysis of qualitative data will strengthen understanding of how and in what circumstances young women can establish pathways to FEE.

1 Approaches to measuring skills and skills programmes

As in the definitions and groupings of skills types discussed in other briefs of this series, there is also considerable overlap in the approaches taken to skills measurement for each skill type. While cognitive and non-cognitive skills may seem to lend themselves to different approaches, there is in fact some cross-fertilization in techniques.

However, type of skills throw up different challenges for measurement, with socio-emotional and other non-cognitive skills proving particularly challenging. This is because “such skills are largely intangible and difficult to isolate from the complex web of interactions and contextual factors that can contribute to their development, usage and impact” (UNICEF 2012).

Socio-emotional and other non-cognitive skills are particularly challenging for measurement because they are intangible and difficult to isolate from a web of interactions and contextual factors.

The first challenge is of course one of definition; it is essential to be clear exactly what skills are included in the group to be targeted for promoting and therefore tracked to measure acquisition. In addition, attitudes and behaviours, as part of the “soft skills” package, are notoriously difficult to measure.

There are at least 3 levels at which measurement is likely to be useful: at the level of programme methodology; at the level of skills acquisition; and at the level of overall programme impact or long-term outcomes.

- At the level of programme methodology, commentators suggest that for life/soft skills, it is important to develop process indicators to capture the teaching-learning process (such as monitoring indicators along the lines of “frequency of use of interactive methodology by trainers/teachers” – see GPYE 2014).
- For tracking skills acquisition, indicators are required that capture the skills attitudes and behaviours sought. These can be tracked either through a continuous assessment model, or through a pre and post-test assessment model.
- At the outcome/impact level, there are various options. Mixed method programme evaluation can synthesise survey and qualitative data to suggest patterns of impact as well as provide a picture of context and causation. The preferred method for establishing the fact of causation, however, is to place the programme within the design of a quasi-experimental model from the beginning, including randomised participation to establish statistical reliability at the outset.

1.1 Measuring skills acquisition

Despite challenges, a number of tools have been developed for measuring socio-emotional skills and higher-order thinking skills including – for example – communication, relationships and collaboration, critical thinking and decision making, and initiative and self-direction (see Wilson-Ahlstrom *et al* 2011 for detail on several instruments). Measuring skills acquisition is variously carried out through self-reporting models or through observable tests – either knowledge or task based.

Self-reporting

Self-reporting models have commonly been applied to non-cognitive skills including socio-emotional skills and personality traits such as the Big Five. Self-reporting surveys are a common tool used by personality psychologists (Heckman and Krautz 2012). A study on the effects on cognitive and non-cognitive skills on employment outcomes in central Asia, for example, used a self-reporting survey with modules “typical of most labor force surveys” and also modules that assessed the respondent’s cognitive (memory, literacy and numeracy) and non-cognitive skills (openness, workplace attitude, decision making, achievement striving and mind set factor). A similar road is taken by the World Bank STEP programme. Differences in pre and post programme responses are used to assess programme performance and understand effects.

Non-cognitive skills are often measured using self-reporting models employing tools such as multiple-question surveys before and after the intervention. The clear challenge for these approaches is that they are subject to respondent bias.

Self-reporting exercises can also make use of composite indicators to assess specific skills. In a programme aiming to enhance the economic empowerment of adolescent girls (EPAG) in Liberia (Adoho *et al.* 2014) two types of non-cognitive skills were measured: self-regulation and entrepreneurial skills. To measure self-regulation, researchers gave participants 11 statements about themselves related to goal setting, sticking to a plan and managing intense emotions; each participant indicated if she agreed with each statement. To measure entrepreneurial skills, the survey asked how well participants felt they could perform six tasks related to starting and running a business. In a similar survey design, a girls' life skills programme in India used self-reporting in a quantitative survey for an RCT study design. Questions aimed to capture information on proxy indicators for the soft skills of aspirations, self-esteem, empowerment and attitudes to the role of women (Delavallade 2015). In both these examples, outcomes were measured through self-reporting using multiple questions, without attempting to directly or independently measure acquisition of the skills.

Clearly, the major challenge for self-reporting models is that they are subject to respondent bias which can undermine reliability (MCF 2014).

Task based tests

Because of the potential unreliability of self-reporting models, and despite the fact that tests are more commonly associated with cognitive type skills, some task-based and scaling tests have also been developed for non-cognitive skills. One such test is the Social and Personal Competencies Scale, CPS for its Spanish acronym (Escala de Competencias Personales y Sociales). The CPS is a non-cognitive test that measures the effectiveness of the life skills module of the program in developing positive attitudes and values. It measures six basic competencies: leadership, behaviour in situations of conflict, self-esteem, abilities to relate with others, order, and empathy and communication skills (Ibarraran *et al.* 2012).

To address potential respondent bias, some task-based and scaling tests have been developed for non-cognitive skills, such as the Spanish Social and Personal Competencies Scale (CPS).

Tests can take a number of forms. Methods for assessing learning outcomes might include, for example (following UNICEF 2012)

- Measuring knowledge through multiple-choice questions or tasks involving poems, essays, posters;
- Measuring attitudes: through scalar attitude measurement tools, open ended questions and closed questions;
- Measuring skills: through close-ended questions, role plays and simulations, case study analysis, check lists;
- Measuring behavioural intent: through close-ended questions, case studies, simulations, checklists etc.

All of these are task-based tests of different forms. Question-answer format tests include studies that attempt to measure skills directly using a type of index measure or model to aggregate proxy behaviours, attitudes and other characteristics to observe a change in skill level. Brown *et al.* (2015) discuss the method used by Chhabra *et al.* (2008) to assess the refusal skills of youth (focusing on sexual encounters) by asking a series of questions around a respondent's confidence in 'being able to refuse their friends in order to avoid an uncomfortable or risky situation.' They also refer to Hazavehei *et al.* (2008) use of the "BASNEF" model (assessing beliefs, attitudes, subjective norms and enabling factors) to measure the ability to assert oneself – they do this by providing participants a series of questions related to each BASNEF characteristic.

Tests using a question-answer format can be similar to surveys. They attempt to aggregate proxy behaviours, attitudes and other characteristics to observe a change in skill level.

Challenges for task-based testing models

Krautz *et al* (2014), however, warn that the performance of task based tests depend essentially on the level of effort and motivation available to a participant – which may risk circularity if effort and motivation are some of the personality traits under scrutiny. Effort, they point out, depends on the incentives offered to exert the effort to perform the task, noting that test scores for young children can be improved by one standard deviation by offering candy for correct answers. Different incentives elicit different amounts of effort on the tasks used to measure skills (Krautz *et al* 2014; Brunello and Schlotter (2011)). They conclude that for measures of cognitive and non-cognitive skill that depend on performance on some task, it is necessary to standardise for incentives. In addition, since performance on most tasks depends on multiple skills it is important to standardise for the other skills – such as effort and others, such as literacy – that contribute to performance. Failing to do these standardizations can produce misleading estimates of the particular skill being measured (Krautz *et al* 2014).

Task based tests risk circularity if effort and motivation are among the non-cognitive skills being tested, because completing the test requires effort. It is therefore important to standardise for the other skills that contribute to performance.

UNICEF also observe that quantitative studies using task based tests also need to control for/account for a number of disparities among the participant population – disparities in social background, gender, labour market, and context-based/cultural variations, for example (UNICEF 2012).

1.2 Measuring intervention impact

Mixed methods programme evaluation to gauge programme impact might include a number of tools to assess different areas as well as provide robust material for triangulation. In addition to pre and post programme testing described above, impact assessments might include participant and employer satisfaction surveys; FGDs with participants and other stakeholders and other forms of post programme data collection among participants and employers on work, conditions, pay and behaviours to capture longer term outcomes as well as detail on the how and why of causation from programme input to effect (GPYE 2014).

However, some commentators (e.g. González-Velosa *et al* 2012) point out that although there is a large literature on, for example, fostering and developing soft skills among disadvantaged young people, the most effective way in which such skills can be developed has not yet been established – the available evidence is discussed further in other briefs of this series. To generate further evidence on the causation between different intervention components and outcomes, Random Control Trial formats for intervention design enable impact evaluations to establish:

- Whether more intensive interventions yield greater short-term results;
- The differential effects of individual interventions and combinations of interventions;
- The long-term effects of interventions;
- Whether interventions have heterogeneous impacts across different occupations or industries or other variables; and potentially
- The cost-effectiveness of interventions, which will require estimates of medium and long-term impacts.

For example, the Galpão Aplauso programme targeting at-risk youth in the favelas surrounding Rio de Janeiro, Brazil with the aims of increasing employment and earnings and reducing risky behaviours (Calero *et al.* (2014) was set up as an RCT. The programme had a number of input components: academic skills; language skills and life skills. At-risk youth who met certain eligibility criteria were at the outset randomly assigned to a programme or control group. Random assignment helped ensure a representative sample of the at-risk youth population and that the programme group and control group were comparable. Results of skills measurement were able to establish that the intervention had positive impacts on employment outcomes starting from four to five months after the end of the intervention.

Challenges for RCTs

A significant challenge for RCTs is that they can rarely capture information about context which can help explain why programmes have certain observed outcomes. As discussed further in other briefs from this series, RCTs have produced contradictory evidence from different countries about the impact of various youth employment schemes. Unfortunately, by themselves they do not have the explanatory power to throw light on the sources/causes of these variations.

RCTs can also be challenging to administer. As González-Velosa *et al* (2012) point out, a common problem is that programme staff may not be willing to deny the provision of training to eligible individuals. Contamination is, therefore, a common risk, since operators may provide services to individuals who were originally assigned to the control group.¹

Brown *et al* (2015) note that in order for RCTs to contribute to future knowledge in the sector, it is essential to include in communications materials adequate initial description of the study sample and a detailed description of the intervention so that comparison can be made across other knowledge sources.

RCTs for skills interventions are commonly at risk of selection contamination because – for example – individuals assigned to a control group may be offered a place if one becomes available.

In addition, RCTs cannot by themselves establish why outcomes are caused, or why there is variation in across similar interventions and contexts.

¹ This study mentions another limitation that I do not understand but may be useful: “Another limitation of the experimental design takes place when the program to be evaluated has, as a fundamental ingredient, a selection process based on characteristics of the participants that are correlated to labor market outcomes. By substituting this selection process for a random assignment, the experimental evaluation alters an essential aspect of the program. In this sense, the evaluated program differs in a fundamental way from the ongoing program that is the object of the evaluation.” (González-Velosa *et al* 2012).

References

- Adoho, F, Chakravarty, S, Korkoyah Jr., DT, Lundberg, M and Tasneem, A, 2014 "The Impact of an Adolescent Girls Employment Program: the EPAG Project in Liberia", Working Paper Series: 6832, The World Bank: Washington, DC.
- Brown, Annette N., Kristen Rankin, Mario Picon and Drew B. Cameron (2015) "The state of evidence on the impact of transferable skills programming on youth in low- and middle-income countries" 3ie Scoping Paper 4, International Initiative for Impact Evaluation (3ie).
- Brunello, Giorgio and Martin Schlotter (2011) "Non Cognitive Skills and Personality Traits: Labour Market Relevance and their Development in Education & Training Systems", IZA discussion Paper No 5743, May.
- Calero, Carla, Carlos Henrique, Veronica Gonzales, Jochen Kluge, and Yuri Soares (2014) 'Can Arts-Based Interventions Enhance Labor Market Outcomes among Youth? Evidence from a Randomized Trial in Rio de Janeiro', IZA Discussion Paper No. 8210 May
- Chhabra, R, Springer, C, Rapkin, B and Merchant, Y, (2008) "Differences among male/female adolescents participating in a school-based teenage education program (STEP) focusing on HIV prevention in India", *Ethnicity and Disease*, 18, pp. 123–127.
- Delavallade, Clara; Alan Griffith, and Rebecca Thornton (2015) – "Girls' Education, Aspirations and Social Networks in India: Evidence from a Randomized Trial in Rural Rajasthan", Available at <https://www.aae.wisc.edu/mwiedc/papers/2015/Thornton-Aspirations%20and%20Social%20networks.pdf>
- Global Partnership for Youth Employment (GPYE) (2014), "Strengthening Life Skills for Youth: A Practical Guide to Quality Programming", International Youth Foundation (IYF) and World Bank.
- González-Velosa, Carolina, Laura Ripani and David Rosas-Shady (2012) "How Can Job Opportunities for Young People in Latin America be Improved?" Inter-American Development Bank TECHNICAL NOTES No. IDB-TN-345, May.
- Hazavehei, SMM, Sharifirad, GR and Kargar, M (2008), "The Comparison of Educational Intervention Effect Using BASNEF and Classic Models on Improving Assertion Skill Level", *Journal of Research in Health Services*, 8(1), pp. 1–11.
- Heckman, James J. and Tim D. Kautz (2012) "Hard Evidence on Soft Skills" National Bureau of Economic Research, Working Paper 18121, June.
- Ibarraran, Pablo, Laura Ripani, Bibiana Taboada, Juan Miguel and Brigida Garcia (2012) "Life Skills, Employability and Training for Disadvantaged Youth: Evidence from a Randomized Evaluation Design", IZA Discussion Paper No. 6617, Institute for the Study of Labor, June.
- Kautz, Tim, James J. Heckman, Ron Diris, Bas ter Weel, Lex Borghans (2014) "Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success", IZA DP No. 8696, Institute for the Study of Labor, December.
- MCF (Master Card Foundation) (2014) "Training Youth in Transferable Skills: State of the Evidence" Part of the Akazi Kanoze 2 Initiative, Education Development Centre. Available at <http://www.mastercardfdn.org/training-youth-in-transferable-skills/>
- Paris, Anne Hilger, Christophe J. Nordman, Leopold R. Sarr (2015) "Cognitive and Non-cognitive skills, Social networks, and Wages in Bangladesh" DRAFT, June 10. Available at http://www.iza.org/conference_files/worldb2015/hilger_a22130.pdf
- UNICEF (2010), "Life skills learning and teaching: principles, concepts and standards" New York.
- UNICEF (2012) "Global Life Skills Education Evaluation: UNICEF Draft Final Report" Education for Change Ltd., February.
- World Bank Group (2014) "STEP skills measurement program – Snapshot", Available at [http://www.worldbank.org/content/dam/Worldbank/Feature%20Story/Education/STEP%20Snapshot%202014_Revised_June%2020%202014%20\(final\).pdf](http://www.worldbank.org/content/dam/Worldbank/Feature%20Story/Education/STEP%20Snapshot%202014_Revised_June%2020%202014%20(final).pdf)
- Wilson-Ahlstrom, Alicia; Nicole Yohalem, David DuBois, and Peter Ji, Adler (2011) "From Soft Skills to Hard Data: Measuring Youth Program Outcomes", The Forum for Youth Investment, September