

# Do Men Really Have Greater Socio-Emotional Skills Than Women?

Evidence from Tanzanian Youth

*Rachel Cassidy*

*Smita Das*

*Clara Delavallade*

*Elijah Kipchumba*

*Julietha Komba*



**WORLD BANK GROUP**

Africa Region

Gender Innovation Lab

August 2024

## Abstract

Individuals' socio-emotional skills (SES), and their perceptions of their skill levels, matter for labor market outcomes and other welfare outcomes. Men appear to have higher levels of SES than women, but this gender gap is typically documented in self-reported measures. Few studies use measures beyond self-reports—or seek to measure SES granularly and rigorously in large samples, especially in low- and middle-income countries. This paper deploys novel sets of self-reported and behavioral measures of 14 SES in a sample of more than 4,000 male and female youth not in full-time education, employment or training, in urban and peri-urban Tanzania. The findings show that

men score higher than women on all 12 positively-worded self-reported measures. In contrast, gender gaps in behavioral measures are only observed for a few skills, and are far smaller in magnitude. The paper provides suggestive evidence that this pattern reflects men's overestimation of their own skills, rather than women's underestimation. In particular, there is a larger gap between self-reported and behavioral measures among men. Men's self-reports, and the gap between their self-reported and behavioral measures, are strongly correlated with measures of their social desirability and gendered beliefs about abilities—but this does not hold for women.

---

This paper is a product of the Office of the Gender Innovation Lab, Africa Region. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at [rcassidy@worldbank.org](mailto:rcassidy@worldbank.org) and [cdellavallade@worldbank.org](mailto:cdellavallade@worldbank.org).

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.*

# Do men really have greater socio-emotional skills than women? Evidence from Tanzanian youth \*

Rachel Cassidy    Smita Das    Clara Delavallade    Elijah Kipchumba  
Julietha Komba

**Keywords:** youth employment, socioemotional skills, measurement, gender, social desirability

**JEL Classification:** J16, J24, 015

---

\*Cassidy: World Bank Africa Gender Innovation Lab (GIL), rcassidy@worldbank.org; Das: Innovations for Poverty Action and World Bank GIL, sdas@poverty-action.org; Delavallade: World Bank GIL, cdelavallade@worldbank.org; Kipchumba: Trinity College Dublin, kipchume@tcd.ie; Komba: BRAC Tanzania, kombajulietha88@gmail.com. This paper is a product of the World Bank GIL in the Africa Region Chief Economist office. We thank Innovations for Poverty Action and BRAC Tanzania, especially Munshi Sulaiman for their collaboration with project design and implementation, as well as Ariel Gruver, Alev Gurbuz Cuneo and Josephine Tassy for excellent research assistance. We also thank Aidan Clerkin, Julian Jamison, Estelle Koussoubé and Léa Rouanet for useful comments on earlier drafts of this paper. This work was funded by the International Development Research Centre (IDRC), Wellspring Philanthropic Fund and the World Bank Umbrella Fund for Gender Equality (UFGE). The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

# 1 Introduction

Global labor trends have increased the demand for jobs that are more intensive in skills that cannot be automated, such as socio-emotional skills (SES), including in low- and middle-income countries (Deming, 2017). Recent evidence shows that SES — the ability to understand emotions and navigate personal and social situations effectively — are beneficial not only for educational outcomes (Chioda et al., 2021) and psycho-social outcomes (Ganimian et al. 2020; Bossuroy et al. 2022), but also for labor market outcomes (Prada et al. 2019; Allemand et al. 2023). Policymakers now thus recognize that SES are foundational alongside literacy and numeracy, and that investing in SES has the potential to yield high economic returns (Cunningham et al., 2022). Not only do actual skill levels matter, but individuals’ perceptions of their skills affect their economic outcomes (Reuben et al. 2017; De Martino et al. 2022). Some literature has found a female disadvantage in SES (Ajayi et al., 2022), and interventions have focused on equipping women with SES to improve their economic empowerment and gender equality (Ashraf et al., 2020; Bandiera et al., 2020; Edmonds et al., 2021). However, this gender gap is typically documented in self-reported measures, which may capture individuals’ perceptions of their skills, social desirability bias and other forms of response bias in addition to actual skill levels. In contrast, behavioral measures may be less subject to reporting biases, and more objectively capture actual skill levels (Duckworth and Yeager, 2015). Understanding if an observed male advantage in SES reflects a gap in actual skill levels — or rather response bias or a gap in perceptions, due to men overestimating or women underestimating their skills — has important implications for the diagnosis of skills gaps, and design and targeting of interventions to support labor market outcomes and women’s empowerment.

This paper deploys a novel, validated set of self-reported and behavioral measures of 14 SES (Das et al., 2024), among over 4,000 male and female youth who are not in education, employment or training (NEET). NEET youth are a population of particular interest for active labor-market policies, especially in Sub-Saharan Africa where 62 percent of the pop-

ulation are aged under 25 (Karkee and Niall, 2023) and the youth NEET rate is 26 percent. Our sample is drawn from NEET youths in urban and peri-urban Tanzania — where the youth NEET rate is 14 percent. We investigate the extent and correlates of gender differences in SES, as measured by self-reported scales and behavioral measures — which consist of situational judgement tests (SJTs) and tasks.

Our main results are threefold. First, we find a significant male advantage in all 12 self-reported SES measures that are positively framed – i.e., where answering more affirmatively (e.g., agree, strongly agree) leads to a higher score. The gender gap is 0.20 standard deviations in the aggregate self-reported SES measure. This male advantage is associated with higher social desirability bias, gender-biased beliefs on abilities, and lower cognitive ability. Second, gender differences disappear for the majority of the skills when using behavioral measures. Of the 14 skills, a small gender gap in behavioral measures is found for three skills and two sub-skills with a magnitude of 0.04 to 0.08 standard deviation units. Third, the gap in self-reported and behavioral measures is also higher at higher levels of social desirability, lower levels of cognitive ability and among individuals holding more regressive gender beliefs. The gap between self-reports and behavioral measures is also higher among men who view men as generally better equipped with problem-solving and decision-making skills than women, while the gap is not correlated with such beliefs among women. We take this as suggestive evidence that the gap between self-reported skills and behaviorally-measured skills among men partly reflects men’s overestimation of their own skills due to biased beliefs around men being more skilled in certain domains than women, rather than women’s underestimation of their skills.

We expand existing literature in three ways. First, this paper contributes to a wider discussion on gender gaps in SES. There is mixed evidence on gender differences in self-reported SES levels. A recent cross-country study using data from more than 40,000 individuals across 17 African countries, including the self-reported data in this study, documents a significant male advantage in all self-reported skills except self-control (Ajayi et al., 2022). However,

these gaps are often small to moderate in size. A review of 46 meta-analyses supports the hypothesis that men and women do not have vastly different soft skills (Hyde 2005; Costa Jr et al. 2001), although evidence is scarce outside high-income countries. One example of a study from a middle-income country is Napolitano et al. (2021), who find no gender gap in measures of growth mindset, mastery orientation and grit among Indonesian adolescents. By examining 14 socio-emotional skills, this paper expands the evidence on gender gaps self-reported SES; and by including behavioral SES measures, the paper provides evidence on the robustness of apparent gender gaps to different measurement methodologies. The list of skills was designed to be as exhaustive and mutually exclusive as possible (Delavallade et al., 2020), and examine skills tied to labor outcomes, with enough granularity to examine gender differences. The framework includes four sub-categories: self-awareness skills (emotional awareness, self-awareness); social awareness skills (listening, empathy); self management skills (emotional regulation, self-control, perseverance, personal initiative, problem-solving and decision-making); and relationship management skills (expressiveness, interpersonal relatedness, influence, negotiation, and collaboration). Our evidence suggests that gender gaps in self-reported SES may be associated with tendencies to misreport or misperceive one’s own skills, which differ systematically by gender, and may not persist when moving away from self-reported measures. Such evidence matters for designing interventions and policies aiming to improve both female and male labor market outcome — for example, informing whether SES training programs should particularly target women, and whether curricula should focus on particular skills based on gender.

Second, our paper adds to the rapidly expanding literature on SES measurement, and in particular the nascent body of literature from low- and middle-income countries (LMICs) (Valerio et al. 2016; Laajaj and Macours 2021; Danon et al. 2023). Accurately measuring SES remains a challenge, especially in LMICs, due to linguistic and cultural translation issues, low levels of literacy, response scales that can be difficult to remember or interpret, offline data collection, time requirements and western bias in scoring methods (Laajaj et al., 2019; Soto

et al., 2021; Dinarte Diaz et al., 2022). Moreover, self-reported skills measures face challenges including acquiescence bias, i.e., the tendency to agree with yes/no questions regardless of their content; reference bias, i.e., respondents’ use of different comparison groups or frames of reference in rating their abilities; and social desirability bias, i.e., the tendency of subjects to respond to test items in such a way as to present themselves in socially acceptable terms. All of these biases may differ for men and women, especially if men and women compare themselves to different reference groups or have different ideas of which skills are socially desirable for their gender (Feingold, 1994; Eagly and Wood, 2012). Limiting such biases has motivated innovations in measurement using more concrete, scenario-based, observation-based or task-based measures — which rely less on subjective self-judgments, suffer less from reporting bias, and have the potential to provide more accurate measures of skills (Duckworth and Yeager, 2015) and related norms (Brar et al., 2023).<sup>1</sup> We expand this line of work first by examining a larger set of 14 SES, with measures extensively adapted, piloted and validated in low and middle-income contexts (Das et al., 2024; Clerkin et al., 2024), and second by systematically comparing self-reports and behavioral measures for each skill.

Third, this paper contributes to the literature on men’s overestimation of their skills and abilities, and related challenges when studying gender gaps using self-reports. There is growing evidence on a gender gap in overconfidence (Beyer, 1990; Lundeberg et al., 1994; Bengtsson et al., 2005; Niederle and Vesterlund, 2007; Reuben et al., 2012, 2014, 2017; Bordalo et al., 2019; Exley and Kessler, 2022). Overconfidence in one’s own abilities helps explain why men predominate as leaders (as demonstrated by Reuben et al. (2012) in a lab setting in the US); and may contribute to gender differences in college major choices and labor market expectations (Reuben et al., 2017), sub-optimal educational/career choices (Wang and Degol, 2017), and sub-optimal hiring choices when employers use self-reports on performance (Reuben et al., 2014). Much of the evidence points to men overestimating their abilities rather than women underestimating their own. For example in separate lab experi-

---

<sup>1</sup>In contrast, Boon-Falleur et al. (2022) show how self-reported measures outperform behavioral tasks measuring SES in one high-income country school setting.

ments in the US [Bench et al. \(2015\)](#); [Bordalo et al. \(2019\)](#); [Exley and Kessler \(2022\)](#) found overconfidence among men in domains (mathematics and science) traditionally regarded as “male” from as early as middle school in conformity with gender stereotypes. That said, a few studies have shown that women exhibit a gap between self-assessments of ability and actual ability — which influences education and employment aspirations ([De Martino et al., 2022](#)) — and may be driven by everyday sexism and the internalized belief that women are less capable ([Correll, 2004](#); [Seron et al., 2016](#)).<sup>2</sup> Women have been shown to rank lower than men on self-assessments of intelligence across 12 countries ([Von Stumm et al., 2009](#)) while no gender difference was observed in objective assessment ([Jensen, 1999](#)). Men’s overestimation and women’s underestimation of their own ability has been shown to be due in part to implicit gender stereotypes ([Reuben et al., 2014](#); [Exley and Kessler, 2022](#)). Our paper adds to this evidence by showing that the gap between men’s (but not women’s) self-reported skills and behaviorally-measured skills is correlated with their reports of beliefs about men being more skilled than women in certain domains, suggesting that men may overestimate their SES more than women underestimate their own. If apparent gender gaps in SES are spuriously driven by men’s overestimation of their own individual skills, and/or their internalization of societal beliefs that men in general are stronger at particular skills, then policy focus on training seemingly less skilled women may be misplaced, and instead interventions may seek to update individuals’ beliefs about their abilities and about men and women’s relative abilities. Interventions to update beliefs may be particularly relevant when i) better informing men’s educational and occupational aspirations, and (ii) processes where selections happen with heavy reliance on written self-reported assessments, or verbal self-reports such as in interviews – for example hiring decisions, loan applications, or broader selection into programs.

---

<sup>2</sup>In contrast, and in line with the same hypothesis, in Uganda [Campos et al. \(2015\)](#) found that women who worked in male-dominated occupations demonstrated greater self-efficacy and were less concerned with social judgments.



## 2 Context, data and methodology

### 2.1 Data source

Our analysis draws from data collected between April and June 2021, from 4,459 individuals spread over 40 communities in three Tanzanian cities: Dodoma, Dar es Salaam, and Iringa. The dataset constitutes the baseline of a randomized control trial of an SES training intervention, conducted in collaboration with the international development organization BRAC.<sup>3</sup> The analysis in this paper focuses specifically on the baseline information collected on individual respondents’ demographics, own and parental education history, SES, cognitive ability, social desirability, and gender attitudes.

### 2.2 Measuring socio-emotional skills

A measurement team including some of the authors of this paper developed a framework of 14 SES (Appendix Figure A1) — informed by existing frameworks, consultations with psychologists, focus group discussions, literature on which SES matter for labor outcomes, and theory on gender gaps in levels and returns to SES (Delavallade et al., 2020). The list of skills includes four “awareness” SES (emotional awareness, self-awareness, listening, and empathy) and 10 “management” SES (emotional regulation, self-control, perseverance, personal initiative, problem-solving and decision-making, expressiveness, interpersonal relatedness, influence, negotiation, and collaboration). An alternative way to categorize the skills is that “intra-personal” skills are composed of self-awareness and self-management while “inter-personal” skills are composed of social awareness and relationship management. Throughout our analysis, we present results for these two alternative classifications of SES. Examples of the measures used for each skill can be found in Appendix A6

---

<sup>3</sup>The RCT study, forthcoming, tests the impact of trainings focused on different sets of SES. It is designed to examine which SES are most teachable to vulnerable male and female youth, and which matter most for economic empowerment. The results are expected to inform updates to one of BRAC’s flagship programs, the Empowerment and Livelihood program (ELA), which provides training and mentorship via safe spaces.

For each of the 14 skills, our baseline survey included one self-report scale and one behavioral measure: either a situational judgment test (SJT, for nine skills) or a task-based measure (for five skills). The SES measures were rigorously developed over several years and validated in three Sub-Saharan African countries (Das et al., 2024; Clerkin et al., 2024). Among the self-report scales, five skills are measured using original items which were developed by the measurement team based on theory, and nine are adapted from existing scales with the addition of selected original items.<sup>4</sup> For each skill, the self-report scale includes six to 12 items and utilizes a five-point Likert response scale.

Turning to the behavioral measures, the measures of negotiation and perseverance are based on existing measures, and the remaining measures are original, designed by the measurement team based on theory. The SJTs, used for nine skills, each involve two to three scenarios. Each scenario is followed by a list of several “good” actions, associated with the successful use of a skill, and a few “poor” actions, associated with the poor use of a skill. Individuals are then asked their likelihood of taking each action. The scenarios were designed to examine SES in the context of economic empowerment, and to constrain the context in which individuals reported their likely behaviors in order to improve the comparability of responses from individuals with differing socioeconomic and demographic backgrounds.<sup>5</sup> For example, one SJT examining *problem solving and decision-making* presented a scenario in which the individual owns a convenient store and profits have gone down because the price of lotion has increased. The individual is given some possible solutions and then asked how likely they are to “Calculate the effect of each solution on your profits” or “Decide which of these solutions to do, based on your gut feeling”. The former is considered a “good” action associated with the use of *problem solving and decision-making* skills, while the latter

---

<sup>4</sup>The sources for these items can be found in Das et al. (2024), but include influential papers such as Schutte et al. (1998), Schwarzer and Jerusalem (1995), Duckworth and Quinn (2009), and Frese et al. (1997).

<sup>5</sup>Scenarios and responses were developed using critical incident sourcing and surveys of possible responses as recommended by Cabrera and Nguyen (2001). The list of actions following each scenario mirrors the content of the self-reported scales, such that the included definitions are as similar as possible. Measures utilize an adapted format that allows them to be administered verbally while minimizing cognitive load for respondents.

is considered a “poor” action.

The tasks used to measure the five remaining skills include a simulated SMS conversation with responses encoded by enumerators to measure collaboration, an enumerator post-survey assessment to measure self-control, responses to scenarios involving conflict,<sup>6</sup> a frustrating puzzle to measure perseverance,<sup>7</sup> and a listening prompt followed by enumerator assessments of active listening and questions to assess comprehension. Measurement examples are provided in Appendix Table A5.

The score for each skill using the self-report measures was calculated as a simple average of item responses.<sup>8</sup> Similarly, the scores for task-based measures were based on simple averages.<sup>9</sup> For the SJTs, item lists for “good” actions – associated with utilizing a given skill – were separated into scenario-based dimensions, and a geometric mean was used to combine them. The average score for the “poor” actions — associated with poor use of a skill — were then subtracted from the score for “good” actions.

Final scores for individuals’ skills were standardized by subtracting the mean and dividing by the standard deviation of the scores from the male sample. Scores for each skill category were then aggregated by calculating the geometric mean of the included individual skills. A modified geometric mean was utilized to prevent scores from zeroing out in the event that the score from one skill was zero. A summary of the psychometric properties for both measurement types can be found in Appendix table A5. In order to develop these measures, the team underwent an extensive process involving qualitative interviews, cognitive interviews, reviews by psychologists and SES trainers, and psychometric testing in

---

<sup>6</sup>This task was adapted from one originally designed to measure negotiation. While the task was tested and could not capture negotiation, the adapted version captures aspects of empathy (Selman et al., 1986).

<sup>7</sup>Over the course of four rounds, individuals were asked to select an easy or difficult puzzle where they would need to count the number of triangles in a given figure. The task was adapted from one used in Alan et al. (2019).

<sup>8</sup>The team also ran the analysis using factor scores, available on request; but results were highly similar and the simple average scores were considered simpler to utilize and interpret.

<sup>9</sup>For empathy, an average was taken of items focused on the pleasure or happiness of the respondent after each scenario. For perseverance, an average was taken across 4 rounds of the game; for each round, the score was based on the difficulty of the puzzle and whether the individual answered or did not answer. Whether the response was correct or incorrect did not affect the score, and those who selected a difficult puzzle obtained more points. Those who quit after a given round received zero points.

several countries. Testing included an examination of inter-item reliability, exploratory and confirmatory factor analysis, the relationship between skills, and concurrent and predictive validity with employment variables. Results are reported in forthcoming papers (Das et al., 2024; Clerkin et al., 2024).

All measures perform well based on common standards, particularly because behavioral measures often face greater psychometric challenges. Cronbach’s alpha for both self-reported and behavioral measures is largely above the common threshold of 0.7 for the 14 skills with the exception of 0.56 for the behavioral measure of empathy, 0.68 for the behavioral measure of maintaining relationships, and 0.61 for the self-reported measure of respectful listening. The goodness of fit statistics from the confirmatory factor analysis, also fall within accepted thresholds: the comparative fit index (CFI) and Tucker-Lewis index (TLI) all fall above 0.90; the standardized root mean square residual (SRMR) all fall below 0.06; and the root mean square error of approximation (RMSEA) largely falls below 0.08 with the exception of the behavioral measures for empathy and expressiveness. While these statistics are often found with commonly used measures, we would caution against over-interpretation of the behavioral results for empathy and the self-reported results for respectful listening. Three behavioral measures are tasks that do not have enough similar items to examine these psychometric qualities: self-control, perseverance, and collaboration.

### **2.3 Measuring other variables of interest**

The survey included three additional measures expected to influence SES measurement: cognitive ability, social desirability, and beliefs as to whether men or women were considered better at SES. The latter two are closely tied to Eagly and Wood (2012)’s social role theory and the social and individual regulation which influence which skills are developed by men and women. While it would have been preferable to also examine acquiescence bias, no measure was available despite the team’s efforts: several attempts to include both positively and negatively framed items in the same scale were unsuccessful in producing unidimensional

measures, or highly correlated multidimensional measures. Cognitive ability is measured using a test inspired by Raven’s matrices (Raven, 1936). For each of six pictures, individuals are given a score of 1 for a correct response and 0 for an incorrect response, then divided by a total of six questions such that the final average score falls between 0 and 1.

The social desirability index (SDI) utilized is based on the 8-item impression management dimension of a short form of the Balanced Inventory of Desirable Responding (BIDR; Hart et al. (2015)) with a 5-point Likert response scale. The items used to assess SDI generally include a series of actions that are deemed socially desirable but uncommon, or socially undesirable but common. Here, the team excluded the dimension of the longer-form BIDR that is focused on self-deceptive enhancement, primarily due to time limitations. An average of the responses to the eight included items is taken, resulting in a score falling between 1 and 5.

Finally, a variable on equitable beliefs is constructed based on the reversed Likert response to the statement *“By nature, men are better at problem solving and decision making than women.”*. The question focused on this particular skill since gender-biased beliefs about this skill were a common finding in qualitative data collection during the SES measure development. The score for each individual is an integer between 1 and 5 — and by its reverse nature, a higher score represents a more gender-equitable belief.

## **2.4 Sample description**

The sample was selected using a community sampling frame of streets (a local administrative unit) which met the following criteria: they had at least 120 eligible young men and 120 eligible young women; community leaders confirmed interest in the program; and there was an existing venue suitable for hosting a training. 40 communities were randomly selected by Stata from this list, stratified by city. Within selected communities, we conducted a listing to create a sampling frame of all eligible individuals — defined as those age 16 to 27, not in full-time education or attending boarding school, and not in full time formal salaried

employment. From this listing, 60 women and 60 men were randomly selected in Stata per community. The final sample is restricted to respondents who answered all questions on SES and comprises an average of 111 individuals per community, of which 50 percent are women.

Table 1 displays summary statistics for the study sample. 73 percent of the sample was located in Dar es Salaam, 20 percent in Dodoma, and eight percent in Iringa. Individuals have a mean age of 21 and mean of nine years of education, slightly higher than that of their fathers (8.1 years) and mothers (7.6 years). The parental education level for male participants is slightly higher than that of female participants — perhaps reflecting differential selection into NEET status by gender. Male participants score significantly higher on test of cognitive ability, and report lower social desirability and less gender-equitable beliefs regarding problem-solving and decision-making skills. This is inline with literature on men’s advantage in cognitive ability - an artefact of socio-cultural context (Hyde and Mertz, 2009; Bordalo et al., 2019), higher social desirability among women (Agut et al., 2022) and lower gender-equitable beliefs among men (Borgonovi et al., 2023). Table 2 includes summary statistics for each of the SES aggregates included in the analysis, disaggregated by gender — while Appendix Tables A1 and A2 show descriptive statistics for disaggregated self-reported and behavioral measures respectively. All scores are standardized based on the male sample.

## 3 Results

### 3.1 Empirical strategy

We first use an analysis of covariance (ANCOVA) estimator to assess whether there are gender gaps in self-reported and behavioral measures of SES. Here we examine each skill, at various levels of aggregation, in a separate regression. We first estimate the gender gaps in SES controlling for basic socio-demographic characteristics — age, father’s education and mother’s education, and city fixed effects. We include enumerator fixed effects for precision and consistency, since enumerators may consciously or unconsciously influence

both measurement noise and the levels of reported skills — for example via the level of explanation they provide or other unintended nudges, or if respondents report differently to enumerators of different genders (see Di Maio and Fiala, 2020; Laajaj and Macours, 2021; Rodriguez-Segura and Schueler, 2023). Importantly, these effects may also interact with the respondent’s gender.<sup>10</sup> Next, we expand the controls to include key characteristics of interest that might be associated with individuals’ self-reported SES — years of education, cognitive ability, social desirability, and beliefs on gendered abilities. Finally, to test whether these additional correlates play a differential role by gender and help in explaining purported gender gaps, we interact these variables with the gender dummy.

---

<sup>10</sup>Our baseline survey was administered by 52 enumerators, equally split by sex, organized into six teams of approximately nine members each. Each team was responsible for surveying respondents within their assigned city: four teams in Dar es Salaam, and one team each for Dodoma and Iringa. The geographically limited teams were formed, balancing administrative and logistical burdens as well as minimizing survey biases. First, we purposely chose six team leaders, one for each team, for administrative purposes. Second, the remaining 46 enumerators chose a city with which they were familiar for logistical considerations. Enumerators who chose Dodoma and Iringa were automatically members of the single team assigned to these cities. Enumerators who chose Dar es Salaam were randomly assigned to one of its four teams. These teams were then randomly assigned to communities within their assigned cities. Finally, within an assigned community, enumerators (members of a team) were randomly assigned to specific respondents without blocking by enumerator or respondent characteristics.

Thus our saturated estimating equation is as follows:

$$\begin{aligned}
SES_{iec} = & \beta_0 + \beta_1 F_{iec} + \beta_2 Edu_{iec} + \beta_3 Edu_{iec} * F_{iec} + \beta_4 Cog_{iec} + \beta_5 Cog_{iec} * F_{iec} \\
& + \beta_6 SDI_{iec} + \beta_7 SDI_{iec} * F_{iec} + \beta_8 Beliefs_{iec} + \beta_9 Beliefs_{iec} * F_{iec} + \beta_{10} X'_{iec} + \lambda_{ec} + \varepsilon_{iec}
\end{aligned} \tag{1}$$

$SES_{iec}$  is a skill for individual  $iec$ , at various levels of aggregation of skills as described in Appendix Figure A1. We use two different types of SES measures: (i) a self-reported measure, and (ii) a behavioral measure, as described above.  $F_{iec}$  is a binary variable equal to one if respondent  $iec$  is a woman.  $Edu_{iec}$  is an interval variable ranging from 0 to 18 representing years of schooling.  $Cog_{iec}$  is individual  $iec$ 's cognitive score ranging from 0 to 1 as described in Section 2.3.  $SDI_{iec}$  is the SDI score ranging from 1 to 5 as described in Section 2.3.  $Beliefs$  is a score ranging from 1 to 5 as described in Section 2.3.  $X'_{iec}$  is a vector of socio-demographic controls, namely age and mother's and father's education,  $\lambda_{ec}$  are enumerator and city fixed effects, and  $\varepsilon_{iec}$  is the error term which is robust to individual heteroskedasticity. Our estimates do not imply a causal relationship. Nonetheless, we argue that we recover consistent estimates of the association between our variables of interest and measured skills, conditional on socio-demographic controls in addition to enumerator and city fixed effects — i.e., we argue that with these controls, there is little further concern about omitted variable bias.

The second part of our analysis focuses on the correlates of the gap between an individual's self-reported skill and their score for the same skill using a behavioral measure. Our other key estimating equation is as follows:

$$\begin{aligned}
SESGap_{iec} = & \beta_0 + \beta_1 F_{iec} + \beta_2 Edu_{iec} + \beta_3 Edu_{iec} * F_{iec} + \beta_4 Cog_{iec} + \beta_5 Cog * F_{iec} \\
& + \beta_6 SDI_{iec} + \beta_7 SDI_{iec} * F_{iec} + \beta_8 Beliefs_{iec} + \beta_9 Beliefs_{iec} * F_{iec} \\
& + \beta_{10} SESBehav_{iec} + \beta_{11} X'_{iec} + \lambda_{ec} + \varepsilon_{iec}
\end{aligned} \tag{2}$$

$SESGap_{iec}$  is a measure of the gap obtained by subtracting the behavioral measure from the self-reported measure, akin to Reuben et al. (2017), for individual  $iec$ , for a skill at various



levels of skill aggregation as described in Appendix.  $Edu_{iec}$ ,  $Cog_{iec}$ ,  $SDI_{iec}$ ,  $Beliefs_{iec}$  and  $X'_{iec}$ ,  $\lambda_{ec}$  and  $\varepsilon_{iec}$  are as described above.  $SESBhav_{iec}$  is the behavioral measure for the skill considered in the outcome. Holding the behaviorally elicited level of SES constant assures comparisons of respondents with similar levels SES and thus  $SESGap_{iec}$  is only a function of the other explanatory variables.

### 3.2 Gender gaps in SES

We begin by reporting ordinary least squares (OLS) estimates of gender differences in aggregate SES scores in Table 3, obtained from estimating various versions of Equation 1. We first report the estimated raw gender gap controlling for age, father’s education, mother’s education, enumerator and city. Next, we report estimates of a specification that also includes measures of cognition, education levels, social desirability and gender norms. Last, we add interactions with the gender dummy. For each (sub-)aggregate, we present gender gap estimates using both a self-reported scale and a behavioral scale in Tables 3 and A3 as well as Figure A2. We also report gender gaps in scores for disaggregated SES measures in Figures 1 and 2.

Table 3 shows that using self-reported measures we observe a gender gap — a significant male advantage — in overall SES scores. Specifically, young women report SES levels that are 0.20 standard deviation units lower than young men on average, after adjusting for differences in respondents’ age, parental education as well as enumerator and city fixed effects (Column 1). We find a similar conditional gender gap when we further partition self-reported SES into sub-aggregates. For example, we see a male advantage of 0.16 standard deviation units in awareness-related SES (Table 3, Column 7) and of 0.20 standard deviation units in management-related SES (Table 3, Column 13). A similar gap is observed in sub-aggregates comprising intrapersonal skills (Table A3, Column 1) and interpersonal skills (Table A3, Column 7) SES. Further examining disaggregated SES using the self-report measures, our results in Figures 1 and 2 indicate a conditional gender gap, ranging from 0.05 to 0.2 standard

deviations, for all self-reported SES measures except self-control and a domain of listening – respectful listening. The magnitude of the gender gaps in self-reported measures persists even after accounting for variables that are closely linked to individuals’ SES (education and cognitive levels, social desirability and beliefs on gendered abilities). For example, Column 2 of Table 3 shows that the male advantage marginally reduces to 0.17 standard deviations when these controls are introduced.

This observed gap could either be an actual gender gap in SES, a reflection of gender differences in reporting bias or in the self-perception of one’s own SES, or a combination of these factors. We first note that women seem to report levels statistically similar to men in only two skills: self-control (Figure 2) and respectful listening (Figure 1).<sup>11</sup> Unlike the other skills measures, these two skills rely on negatively-framed scales, e.g., “*I say inappropriate things.*” or “*Sometimes I can’t stop myself from doing something, even if I know it is wrong.*”<sup>12</sup> This finding may indicate a reporting phenomenon wherein men are more likely to agree to statements whether positively or negatively framed. We argue that there may be a stronger acquiescence bias among men, which we provide suggestive evidence for below.

To investigate whether men’s advantage in self-reported SES measures may be an artefact of self-reporting, we next turn to behavioral measures. Behavioral measures have the potential to be less biased than self-reports, since in SJTs the “desirable” answer may be less evident, and in task-based measures the objective is to measure the skill being performed directly. Columns 4, 10 and 16 of Table 3 estimate the conditional gender gap using the behavioral measures of SES. At the (sub-)aggregate level, the magnitude of the gender difference in SES is at least five times smaller than that observed for self-reported scores, and barely achieves significance at conventional levels. We further find that gender differences persist in only three skills and two sub-skills out of the 14 SES disaggregated skills when

---

<sup>11</sup>While estimates are statistically similar at conventional levels, coefficient signs suggest women report higher levels compared to men in these two skills.

<sup>12</sup>In addition, items of opposite valence never loaded on the same scale. Issues arising from combining positively and negatively framed items have been documented in previous literature, see for instance [Chyung et al. \(2018\)](#).

we consider behavioral SES measures (Figures 1 and 2). Gender differences persist in skills mostly related to self-management SES (Figure A2). Specifically, we see a significant persistent female disadvantage in a dimension of listening – active listening – in addition to emotional regulation and perseverance. There are marginally significant (at the 10% level) female disadvantages in PSDM and a dimension of interpersonal relatedness – networking. Overall, these results are more in line with Hyde (2005) and several other meta-analyses which have found that gender differences in SES are small or insignificant.

It is noteworthy that the gender gaps observed in SES measures elsewhere in the literature often do not conform to theory nor existing literature. While women are often expected to develop more *communal* skills while men are expected to develop *agentic* skills (Eagly and Wood, 2012), self-reports are generally not found to be higher for women for skills such as listening, empathy, interpersonal relatedness, and collaboration. Among behavioral measures, the slight male advantage we observe in active listening is also theoretically unexpected, but we observe no male advantage in *agentic* skills such as personal initiative and expressiveness. The slight male advantage in networking is more expected, and the male advantage in emotional regulation matches a common finding in the literature (Ajayi et al., 2022).

### 3.3 Heterogeneity of the gender gap in SES

We next assess whether measures of cognition, education levels, social desirability and beliefs on gendered abilities are associated with self-reported SES — and whether these associations differ by gender, potentially explaining the purported gender gap in self-reported SES. For example, if self-reported measures are more strongly correlated with social desirability bias relative to behavioral measures, and if men have higher levels of social desirability or the correlation is stronger among men, then this may explain the pattern of gender gaps outlined above. Therefore, in addition to the controls used in estimating the conditional gender gap in Section 3.2, our regressions in this section include measures of cognition, education levels, social desirability and gender norms as well as their interaction with the gender dummy; see

Equation 1.

As expected, we find a positive correlation between cognition and SES levels irrespective of the type of measures used. Coefficients corresponding to cognition in Table 3 are qualitatively similar between self-reported and behavioral measures (Columns 2 and 5, 8 and 11, 14 and 17). This correlation may reflect a true positive correlation between cognitive and non-cognitive (SES) skills, and/or that the skill measures are easier to understand and to answer positively for those with higher cognitive ability. We see a similar positive correlation between SES and education levels, though with a much weaker magnitude compared to cognitive skills. The correlation between education levels and SES does not differ by gender — the interaction between education and gender is almost identically zero and insignificant. The same holds for behavioral measures (columns 6, 12, 18). This finding could indicate that young men and young women realize similar returns to education in terms of SES irrespective of their gender, or may indicate that selection into education based on SES does not vary by gender in this context.

Meanwhile, the association between cognitive skills and the self-reported SES measures differs across women and men. Specifically, women have a stronger association between cognitive skills and self-reported SES compared to men. For example, compared to men of similar cognitive levels, women’s overall self-reported SES is 0.22 standard deviations higher (Column 3 of Table 3). As a corollary, the gender gap in self-reported SES is larger at lower levels of cognitive ability. We see similar gender-differentiated associations between cognitive and non-cognitive ability (SES) when self-reported SES are further partitioned into either awareness-related versus management-related SES (Table 3, Columns 9 and 15) or intrapersonal versus interpersonal SES (Table A3, Columns 3 and 9). Turning to behavioral measures, in contrast we do not see a gender-differentiated role of either cognition or education levels. This may indicate that the relationship between cognitive skills and self-reported SES is more reflective of reporting or understanding of self-report measures.

Our results also show a significant positive correlation between social desirability and

SES when using both self-reported and behavioral measures, but that the correlation corresponding to self-reported measures is almost twice that of behavioral measures (Table 3, Columns 2 and 5, 8 and 11, 14 and 17). In addition, we find that the correlation between social desirability and self-reported measures of SES is smaller for women than for men. By contrast, the gender gap in these skills is not significantly correlated with social desirability when using behavioral SES measures (Table 3, columns 6, 12 and 18). When behavioral measures are used, the point estimates very close to zero may indicate the absence of a gender-differentiated role for social desirability in SES when elicited via behavioral measures. Further exploration shows these results are robust to the inclusion of enumerator gender and its interaction with the gender of the respondent.<sup>13</sup>

Next, we examine whether gender-equitable beliefs are associated with self-reported SES. We focus on respondents' endorsement of a statement that men have greater *problem-solving and decision-making (PSDM)* skills than women. Our measure of gender beliefs is reversed such that a higher score represents a more gender-equitable belief (See Section 2.3). Men who endorse less the statement that men have greater PSDM than women also have lower self-reported SES, significant at 5% (Table 3, Columns 3, 9, and 15). On the contrary, young men who show less endorsement that men have greater PSDM than women exhibit higher SES levels when behavioral SES measures are used. Further, young women's view on gender equality in PSDM are not associated with their SES, irrespective of whether self-reported or behavioral measures of SES are used. Taken together, these results could point to gender norms playing a role in how young men report or perceive their skill levels, but not young women.

### 3.4 Do young men overestimate their SES?

Three striking findings have emerged so far. First is the absence of gender gap in a majority of SES when using behavioral measures. Second is a female disadvantage in self-reported

---

<sup>13</sup>Results are available upon request.

measures that widens among youths who have lower cognitive ability or demonstrate higher social desirability. Third is that young men who endorse gender-unequal skill-related beliefs have higher self-reported SES — a pattern not observed among young women, and largely missing in both genders when using behavioral SES measures. These findings suggest that men’s self-reports, and not women’s, may be artificially inflated. We cannot experimentally distinguish between women’s underestimation and men’s overestimation. However, to test further the hypothesis that men overestimate their skills more than women underestimate theirs, we use the gap between individuals’ self-report and their behavioral measure for the same skill as a proxy for skill overestimation. And we examine the relationship between this gap and the covariates used in Section 3.3, in addition to controlling for actual SES levels using the behavioral measures, by estimating OLS coefficients of Equation 2.

We expect a gender gap in favor of men in this skill assessment gap. In addition, if men overestimate their skills more than women underestimate theirs, we expect men’s skill assessment gap to be more correlated with gender beliefs and by social desirability than women’s. In contrast, if young women underestimate their skills more than men overestimate theirs — for example because women face higher pressure to conform to prevailing gender norms such as on women being modest — we should observe a stronger association between the skills assessment gap and biased beliefs on gendered abilities or social desirability among women.

Consistent with the former explanation, results in Table 4 show a higher degree of overestimation in SES among young men relative to young women. For example, young men’s overestimation of their overall SES is 0.19 standard deviations higher compared to young women (Columns 1 of Table 4). Columns 4 and 7 of Table 4 shows similar gender gap in overestimation when SES is sub-aggregated as awareness-related and management-related SES. Similarly, Columns 1 and 4 of Table A4 shows a similar gender gap in overestimation when SES is sub-aggregated as intrapersonal and interpersonal SES. These results hold for each disaggregated SES, except for self-control (Figures 3 and 4). Moreover, we find the

level of SES overestimation by both men and women positively correlate with their levels of education and cognition (Columns 2, 5 and 8).

Results in Table 4 (Columns 3, 6 and 9) and Table A4 (Columns 3 and 6) provide further evidence that gender norms are associated with skills (mis)perception. First, although young women are more likely to overestimate their SES when they exhibit higher social desirability, the magnitude of overestimation is significantly lower compared to men. Second, if women internalized the belief that modesty was a socially desirable skill for women, we would expect the relationship between social desirability and self-reported scores to be negative for women. However, in Table 3, women’s self-reported scores have a positive relationship with social desirability, though the magnitude is smaller than that among men. Third, the more young men endorse that men are more skilled than women at PSDM, the more they over-report their own skills conditional on a given level of behaviorally-measured skills and social desirability bias. Finally, unlike among men, gender attitudes among young women do not correlate with women’s own skills overestimation. These results provide further suggestive evidence that the gender gap in skills self-assessment reflects men’s overestimation of their own skills. This suggestive evidence of men’s overestimation of SES in Tanzania is in line with lab experiments in the United States. For example, [Bench et al. \(2015\)](#) and [Exley and Kessler \(2022\)](#) in their lab experiments find overconfidence among men in domains traditionally regarded as “male” (mathematics and science) from as early as middle school, in conformity with gender stereotypes.

## 4 Conclusion

Using innovative SES measures, based on self-reports, situational judgment tests and behavioral tasks, we provide novel evidence on gender differences in SES. We find that men report significantly higher levels than women on all 14 skills except the two negatively valenced scales — self-control and listening. The gender gap corresponds to a 0.20 standard deviation

unit difference on the all-SES aggregate measure.

Gender differences in self-reports are driven by (i) individuals at the higher end of the social desirability distribution — men’s SES self-reports correlating more with their tendency to align their responses to what they deem socially acceptable relative to women; (ii) individuals with lower levels of cognitive ability — the association between cognitive ability and SES being stronger among women; and (iii) men who hold the belief that men in general have stronger problem-solving and decision-making skills than women — suggesting that self-reports partly reflect biased beliefs on gendered abilities.

We further show that the male advantage in widely used self-reported SES measures disappears for the majority of skills when behavioral measures such as situational judgment tests and tasks are administered. Behavioral measures show weaker correlations with social desirability, suggesting that they may be less prone to biased reporting. Further, for men, holding regressive beliefs around men’s and women’s relative skills is linked with lower behavioral measures but higher self-reports.

Finally, the gap between self-reported and behavioral skill measures is significantly higher for young men than for young women (18 standard deviation units higher). While among women there is no correlation between the assessment gap and beliefs around men’s and women’s relative problem-solving abilities, this assessment gap is higher for men who hold regressive gender views. This may suggest that the skills assessment gap reflects men’s overestimation of their own skills due to their biased beliefs on gendered abilities, rather than women’s underestimation. In addition, the gender gap in skills overestimation is wider at high levels of social desirability, with men’s skills assessment gap more strongly correlated with their propensity to provide responses deemed socially acceptable. Further research may address the limitations of the present study by directly examining the gap between behavioral measures of SES and individual assessments of their score on the same measure, as an index of over-(under-)estimation.

These results have important implications for policies seeking to reduce the purported



gender gap in SES and equip women with the skills that may be critical for their employability and success in the labor market, especially in low-income settings. The results call for caution on several accounts: first, when measuring SES and gender gaps in SES, especially when using self-reports; second, when targeting women with SES trainings to address the gender gap in employability, since women may be as well equipped as men but less prone to skills' overestimation; and (iii) overinflated claims of gender differences may be costly in the workplace (Hyde, 2005). In contexts where the gap in skills' perceptions might be more a reflection of men overestimating their skills than of women lagging behind in actual skill levels, the gender gap in skills assessment may be more accurately addressed by updating men's beliefs about their individual abilities, as well as their social beliefs about men's and women's relative SES abilities. Interventions recalibrating misperceptions about others have indeed shown to have positive results in other settings (Bursztyn and Yang, 2022), as have interventions to reshape young men's (and women's) perceptions of gender attributes and regressive gender attitudes, for example via classroom discussions (Dhar et al., 2022) or video-based community sensitization (Bossuroy et al., 2022). Whether such interventions might affect individuals' perceptions of their own skills — and propensity e.g., to compete in labor market settings — remain questions for future research.

## References

- Agut, S., Martín-Hernández, P., Soto, G., and Arahuete, L. (2022). Understanding the relationships among self-ascribed gender traits, social desirability, and ambivalent sexism. *Current Psychology*.
- Ajayi, K., Das, S., Delavallade, C., Ketema, T., and Rouanet, L. M. (2022). Gender Differences in Socio-Emotional Skills and Economic Outcomes: New Evidence from 17 African Countries. *World Bank Policy Research Working Paper*, (10197).
- Alan, S., Boneva, T., and Ertac, S. (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics*, 134(3):1121–1162.
- Allemand, M., Kirchberger, M., Milusheva, S., Newman, C., Roberts, B., and Thorne, V. (2023). Conscientiousness and labor market returns.
- Ashraf, N., Bau, N., Low, C., and McGinn, K. (2020). Negotiating a better future: How interpersonal skills facilitate intergenerational investment. *The Quarterly Journal of Economics*, 135(2):1095–1151.
- Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I., and Sulaiman, M. (2020). Women’s empowerment in action: evidence from a randomized control trial in africa. *American Economic Journal: Applied Economics*, 12(1):210–259.
- Bench, S. W., Lench, H. C., Liew, J., Miner, K., and Flores, S. A. (2015). Gender gaps in overestimation of math performance. *Sex Roles*, 72:536–546.
- Bengtsson, C., Persson, M., and Willenhag, P. (2005). Gender and overconfidence. *Economics Letters*, 86(2):199–203.
- Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, 59(5):960.
- Boon-Falleur, M., Bouguen, A., Charpentier, A., Algan, Y., Huillery, É., and Chevallier, C. (2022). Simple questionnaires outperform behavioral tasks to measure socio-emotional skills in students. *Scientific reports*, 12(1):442.
- Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, 109(3):739–773.
- Borgonovi, F., Han, S. W., and Greiff, S. (2023). Gender differences in collaborative problem-solving skills in a cross-country perspective. *Journal of Educational Psychology*, 115(5):747–766.
- Bossuroy, T., Goldstein, M., Karimou, B., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Thomas, C. C., Udry, C., Vaillant, J., et al. (2022). Tackling psychosocial and capital constraints to alleviate poverty. *Nature*, 605(7909):291–297.

- Brar, R. K., Buehren, N., Papineni, S., and Sulaiman, M. (2023). Rebel with a cause: Effects of a gender norms intervention for adolescents in somalia. *Policy Research working paper*, WPS 10567.
- Bursztyjn, L. and Yang, D. Y. (2022). Misperceptions about others. *Annual Review of Economics*, 14:425–452.
- Cabrera, M. A. and Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, 9(1–2):103–113.
- Campos, F., Goldstein, M., McGorman, L., Munoz Boudet, A. M., and Pimhidzai, O. (2015). Breaking the metal ceiling: female entrepreneurs who succeed in male-dominated sectors. *World Bank Policy Research Working Paper*, (7503).
- Chioda, L., Contreras-Loya, D., Gertler, P., and Carney, D. (2021). Making entrepreneurs: Returns to training youth in hard versus soft business skills. Technical report, National Bureau of Economic Research.
- Chyung, S. Y., Barkin, J. R., and Shamsy, J. A. (2018). Evidence-based survey design: The use of negatively worded items in surveys. *Performance Improvement*, 57(3):16–25.
- Clerkin, A., Das, S., Delavallade, C., Gonzales, C., Jamison, J., and Rouanet, L. (2024). Socio-emotional skills in sub-saharan africa: Validating and comparing self-reported and behavioral measures. *mimeo*.
- Correll, S. J. (2004). Constraints into preferences: Gender, status, and emerging career aspirations. *American Sociological Review*, 69(1):93–113.
- Costa Jr, P. T., Terracciano, A., and McCrae, R. R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. *Journal of Personality and Social Psychology*, 81(2):322.
- Cunningham, W., Moroz, H., Muller, N., and Solatorio, A. (2022). The demand for digital and complementary skills in Southeast Asia.
- Danon, A., Das, J., De Barros, A., and Filmer, D. (2023). Cognitive and socioemotional skills in low-income countries: Measurement and associations with schooling and earnings. *Journal of Development Economics*, page 103132.
- Das, S., Koroknay-Palicz, T., Marsh, V., McDaniel, D., and Rouanet, L. (2024). Socioemotional skills in africa: Development and validation of 14 measures in sub-saharan africa with implications for economic outcomes. *mimeo*.
- De Martino, S., Farfán, G., Gayoso, L., and Osman, E. (2022). Socio-Emotional Drivers of Youth Unemployment: The Case of Higher Educated Youth in Sudan.
- Delavallade, C., Rouanet, L., and Das, S. (2020). Unpacking Socio-Emotional Skills for Women’s Economic Empowerment.

- Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 132(4):1593–1640.
- Dhar, D., Jain, T., and Jayachandran, S. (2022). Reshaping adolescents’ gender attitudes: Evidence from a school-based experiment in india. *American economic review*, 112(3):899–927.
- Di Maio, M. and Fiala, N. (2020). Be Wary of Those Who Ask: A Randomized Experiment on the Size and Determinants of the Enumerator Effect. *World Bank Economic Review*, 34(3):654–669.
- Dinarte, L. and Egana-delSol, P. (2019). Preventing violence in the most violent contexts: Behavioral and neurophysiological evidence. *World Bank Policy Research working paper*, (8862).
- Dinarte Diaz, L. I., Egana-delSol, P., and Martinez A, C. (2022). Socioemotional skills development in highly violent contexts.
- Duckworth, A. and Quinn, P. (2009). Development and validation of the short grit scale (grit-s). *Journal of personality assessment*, 91(2):166–174.
- Duckworth, A. L. and Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, 45(3):259–268.
- Duckworth, A. L. and Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4):237–251.
- Eagly, A. H. and Wood, W. (2012). Social role theory. *Handbook of theories of social psychology*, 2:458–476.
- Edmonds, E., Feigenberg, B., and Leight, J. (2021). Advancing the agency of adolescent girls. *Review of Economics and Statistics*, pages 1–46.
- Exley, C. L. and Kessler, J. B. (2022). The gender gap in self-promotion. *The Quarterly Journal of Economics*, 137(3):1345–1381.
- Feingold, A. (1994). Gender differences in personality: a meta-analysis. *Psychological Bulletin*, 116(3):429.
- Frese, M., Fay, D., Hilburger, T., Leng, K., and Tag, A. (1997). The concept of personal initiative: Operationalization, reliability and validity in two german samples. *Journal of occupational and organizational psychology*, 70(2):139–161.
- Ganimian, A., Barrera-Osorio, F., Biehl, M. L., and Cortelezzi, M. Á. (2020). Hard cash and soft skills: Experimental evidence on combining scholarships and mentoring in argentina. *Journal of Research on Educational Effectiveness*, 13(2):380–400.
- Hart, C. M., Ritchie, T. D., Hepper, E. G., and Gebauer, J. E. (2015). The balanced inventory of desirable responding short form (BIDR-16). *Sage Open*, 5(4):2158244015621113.

- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60(6):581.
- Hyde, J. S. and Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences*, 106(22):8801–8807.
- Jensen, A. R. (1999). The g Factor: The Science of Mental Ability. *Psychology*, 10(04):36–2443.
- Karkee, V. and Niall, O. (2023). African youth face challenges in the transition from school to work. Technical report, ILO.
- Laaajaj, R. and Macours, K. (2021). Measuring skills in developing countries. *Journal of Human Resources*, 56(4):1254–1295.
- Laaajaj, R., Macours, K., Pinzon Hernandez, D. A., Arias, O., Gosling, S. D., Potter, J., Rubio-Codina, M., and Vakis, R. (2019). Challenges to capture the big five personality traits in non-WEIRD populations. *Science Advances*, 5(7):eaaw5226.
- Lundeberg, M. A., Fox, P. W., and Punécohaí, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology*, 86(1):114.
- Napolitano, C., Molina, D. C., Johnson, H. C., Oswald, F., Hernandez, D. A. P., Tiwari, A., De Martino, S., and Trzesniewski, K. (2021). Are growth mindset, mastery orientation, and grit promising for promoting achievement in the Global South? Psychometric evaluations among Indonesian adolescents.
- Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.
- Prada, M. F., Rucci, G., and Urzúa, S. (2019). Training, soft skills and productivity: evidence from a Field Experiment in Retail. Technical report, IDB Working Paper Series.
- Raven, J. C. (1936). Mental tests used in genetic, the performance of related individuals on tests mainly educative and mainly reproductive.
- Reuben, E., Rey-Biel, P., Sapienza, P., and Zingales, L. (2012). The emergence of male leadership in competitive environments. *Journal of Economic Behavior & Organization*, 83(1):111–117.
- Reuben, E., Sapienza, P., and Zingales, L. (2014). How stereotypes impair women’s careers in science. *Proceedings of the National Academy of Sciences*, 111(12):4403–4408.
- Reuben, E., Wiswall, M., and Zafar, B. (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal*, 127(604):2153–2186.
- Rodriguez-Segura, D. and Schueler, B. E. (2023). Assessors influence results: Evidence on enumerator effects and educational impact evaluations. *Journal of Development Economics*, 163:103057.

- Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., and Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and individual differences*, 25(2):167–177.
- Schwarzer, R. and Jerusalem, M. (1995). Generalized self-efficacy scale. *Measures in health psychology: A user's portfolio. Causal and control beliefs*, pages 35–37.
- Selman, R. L., Beardslee, W., Schultz, L. H., Krupa, M., and Podorefsky, D. (1986). Assessing adolescent interpersonal negotiation strategies: Toward the integration of structural and functional models. *Developmental Psychology*, 22(4):450.
- Seron, C., Silbey, S. S., Cech, E., and Rubineau, B. (2016). Persistence is cultural: Professional socialization and the reproduction of sex segregation. *Work and Occupations*, 43(2):178–214.
- Soto, C. J., Napolitano, C. M., and Roberts, B. W. (2021). Taking skills seriously: Toward an integrative model and agenda for social, emotional, and behavioral skills. *Current Directions in Psychological Science*, 30(1):26–33.
- Valerio, A., Sanchez Puerta, M. L., Tognatta, N., and Monroy-Taborda, S. (2016). Are there skills payoffs in low-and middle-income countries? empirical evidence using step data. *World Bank Policy Research Working Paper*, (7879).
- Von Stumm, S., Chamorro-Premuzic, T., and Furnham, A. (2009). Decomposing self-estimates of intelligence: Structure and sex differences across 12 nations. *British Journal of Psychology*, 100(2):429–442.
- Wang, M.-T. and Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review*, 29:119–140.

# Tables

Table 1: Descriptive statistics - Control variables

Variable	Men (1)		Women (2)		t-test difference (1)-(2)	Min	Max
	N	Mean/SE	N	Mean/SE			
Dar es Salaam	2231	0.725 [0.009]	2228	0.725 [0.009]	-0.000	0	1
Dodoma	2231	0.200 [0.008]	2228	0.200 [0.008]	-0.000	0	1
Age in years	2231	21.080 [0.057]	2228	21.048 [0.061]	0.032	16	27
Years of education	2231	9.120 [0.069]	2228	8.996 [0.069]	0.124	0	14
Father's education	2231	8.071 [0.056]	2228	7.785 [0.050]	0.286***	0	15
Mother's education	2231	7.576 [0.054]	2228	7.426 [0.049]	0.150**	0	14
Cognitive Ability	2231	0.719 [0.005]	2228	0.686 [0.005]	0.033***	0	1
Social desirability index	2231	3.412 [0.010]	2228	3.442 [0.010]	-0.030**	1.50	5.00
Equitable beliefs regarding PSDM abilities	2231	2.353 [0.023]	2228	2.997 [0.025]	-0.644***	1.00	5.00

*Notes:* The values displayed for t-tests are the differences in the means across the groups. PSDM = Problem-solving and decision-making. \*\*\*  $p \leq 0.01$ , \*\*  $p \leq 0.05$ , \*  $p \leq 0.1$ .

Table 2: Descriptive statistics - Aggregate self-reported and behavioral measures

Variable	Men		Women		t-test difference (1)-(2)	Min	Max
	N	(1) Mean/SE	N	(2) Mean/SE			
Self-reported: All	2231	-0.000 [0.021]	2228	-0.194 [0.020]	0.194***	-6.894	3.164
Self-reported: Intra	2231	-0.000 [0.021]	2228	-0.177 [0.021]	0.177***	-6.522	3.153
Self-reported: Inter	2231	0.000 [0.021]	2228	-0.186 [0.020]	0.186***	-7.887	2.979
Self-reported: Awareness	2231	-0.000 [0.021]	2228	-0.159 [0.021]	0.159***	-7.995	2.890
Self-reported: Management	2231	0.000 [0.021]	2228	-0.195 [0.020]	0.195***	-7.281	3.200
Self-reported: Self-awareness	2231	0.000 [0.021]	2228	-0.161 [0.021]	0.161***	-8.372	2.417
Self-reported: Social Awareness	2231	-0.000 [0.021]	2228	-0.110 [0.021]	0.110***	-5.327	2.599
Self-reported: Self Management	2231	-0.000 [0.021]	2228	-0.168 [0.021]	0.168***	-6.755	3.209
Self-reported: Rel Management	2231	0.000 [0.021]	2228	-0.188 [0.020]	0.188***	-8.524	2.715
Behavioral: All	2231	-0.000 [0.021]	2228	-0.030 [0.021]	0.030	-4.573	3.511
Behavioral: Intra	2231	0.000 [0.021]	2228	-0.029 [0.021]	0.029	-6.783	3.028
Behavioral: Inter	2231	0.000 [0.021]	2228	-0.021 [0.021]	0.021	-3.397	4.145
Behavioral: Awareness	2231	-0.000 [0.021]	2228	-0.002 [0.020]	0.002	-5.074	2.060
Behavioral: Management	2231	-0.000 [0.021]	2228	-0.038 [0.021]	0.038	-3.672	3.808
Behavioral: Self-awareness	2231	0.000 [0.021]	2228	0.019 [0.021]	-0.019	-6.765	1.215
Behavioral: Social Awareness	2231	0.000 [0.021]	2228	-0.025 [0.020]	0.025	-4.286	2.003
Behavioral: Self Management	2231	-0.000 [0.021]	2228	-0.057 [0.022]	0.057*	-5.895	3.784
Behavioral: Rel Management	2231	-0.000 [0.021]	2228	-0.012 [0.021]	0.012	-2.921	3.800

Notes: The value displayed for t-tests are the differences in the means across the groups. \*\*\*  $p \leq 0.01$ , \*\*  $p \leq 0.05$ , \*  $p \leq 0.1$ .



Table 3: All/Awareness/Management - Self-reported and behavioral measures

	All						Awareness						Management					
	Self-reported			Behavioral			Self-reported			Behavioral			Self-reported			Behavioral		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Women	-0.20*** (0.03)	-0.17*** (0.03)	0.05 (0.24)	-0.04 (0.02)	-0.05** (0.02)	-0.02 (0.19)	-0.16*** (0.03)	-0.14*** (0.03)	-0.05 (0.24)	-0.03 (0.02)	-0.02 (0.02)	-0.10 (0.18)	-0.20*** (0.03)	-0.17*** (0.03)	0.09 (0.24)	-0.03 (0.02)	-0.05* (0.02)	0.03 (0.21)
Years of education		0.03*** (0.00)	0.03*** (0.01)		0.02*** (0.00)	0.02*** (0.01)		0.03*** (0.00)	0.03*** (0.01)		0.01** (0.00)	0.01 (0.01)		0.03*** (0.00)	0.03*** (0.01)		0.02*** (0.00)	0.02*** (0.01)
Years of education X Women			0.01 (0.01)		0.00 (0.01)	0.00 (0.01)		0.01 (0.01)	0.01 (0.01)		0.01 (0.01)	0.01 (0.01)		0.01 (0.01)	0.01 (0.01)		0.00 (0.01)	0.00 (0.01)
Cognitive Ability		0.40*** (0.07)	0.29*** (0.09)		0.40*** (0.06)	0.35*** (0.08)		0.34*** (0.07)	0.24*** (0.09)		0.18*** (0.06)	0.16** (0.07)		0.40*** (0.07)	0.29*** (0.09)		0.41*** (0.06)	0.35*** (0.08)
Cognitive Ability X Women			0.22** (0.11)		0.11 (0.10)	0.11 (0.10)		0.19* (0.11)	0.19* (0.11)		0.03 (0.09)	0.03 (0.09)		0.22** (0.11)	0.22** (0.11)		0.12 (0.10)	0.12 (0.10)
Social desirability index		0.49*** (0.04)	0.57*** (0.05)		0.19*** (0.03)	0.20*** (0.04)		0.45*** (0.04)	0.52*** (0.05)		0.02 (0.03)	0.01 (0.04)		0.47*** (0.04)	0.55*** (0.05)		0.23*** (0.03)	0.25*** (0.04)
Social desirability X Women			-0.16** (0.07)		-0.02 (0.05)	-0.02 (0.05)		-0.13** (0.07)	-0.13** (0.07)		0.02 (0.05)	0.02 (0.05)		-0.16** (0.07)	-0.16** (0.07)		-0.04 (0.05)	-0.04 (0.05)
Equitable beliefs regarding PSDM abilities		-0.03* (0.01)	-0.05** (0.02)		0.04*** (0.01)	0.05*** (0.02)		-0.02 (0.01)	-0.05** (0.02)		0.00 (0.01)	0.02 (0.01)		-0.03* (0.01)	-0.04** (0.02)		0.05*** (0.01)	0.06*** (0.02)
Equitable beliefs X Women			0.04* (0.03)		-0.03 (0.02)	-0.03 (0.02)		0.07** (0.03)	0.07** (0.03)		-0.03 (0.02)	-0.03 (0.02)		0.03 (0.03)	0.03 (0.03)		-0.02 (0.02)	-0.02 (0.02)
p(Edu. + Edu. X Women = 0)			0.00			0.00			0.00			0.02			0.00			0.00
p(CA + CA X Women = 0)			0.00			0.00			0.00			0.01			0.00			0.00
p(SD + SD X Women = 0)			0.00			0.00			0.00			0.37			0.00			0.00
p(Equit. + Equit. X Women = 0)			0.84			0.10			0.45			0.44			0.58			0.02
p(Women + Edu. X Women = 0)			0.65			0.98			0.99			0.77			0.55			0.85
p(Women + CA X Women = 0)			0.36			0.79			0.70			0.62			0.29			0.57
p(Women + SD X Women = 0)			0.00			0.46			0.00			0.82			0.00			0.42
p(Women + Equit. X Women = 0)			0.46			0.61			0.59			0.32			0.44			0.93
Observations	4459	4459	4459	4459	4459	4459	4459	4459	4459	4459	4459	4459	4459	4459	4459	4459	4459	4459
R-squared	0.12	0.20	0.20	0.40	0.42	0.42	0.11	0.17	0.17	0.46	0.46	0.46	0.12	0.19	0.19	0.35	0.38	0.38

Notes: Results presented are OLS estimates that include controls for enumerator, age, father's education, mother's education and city. Outcome measures are standardized naive scores. Robust standard errors in parentheses. PSDM = Problem-solving and decision-making. Edu. = Years of education. CA = Cognitive Ability. SD = Social Desirability. Equit. = Equitable beliefs. \*\*\*  $p \leq 0.01$ , \*\*  $p \leq 0.05$ , \*  $p \leq 0.1$ .

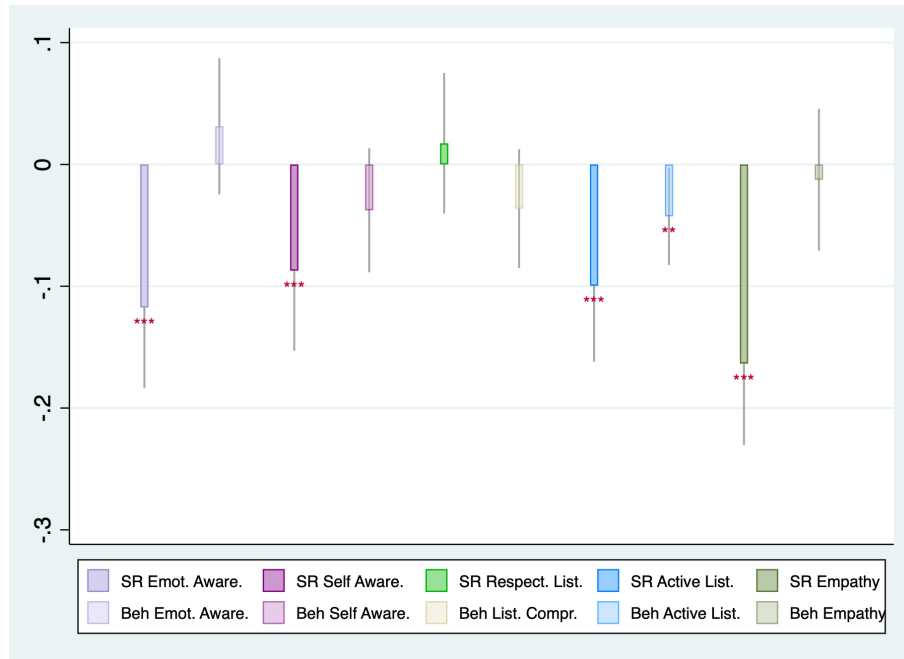
Table 4: All/Awareness/Management - Gap between self-reported and behavioral measures

	All SR-Behavioral gap			Awareness SR-Behavioral gap			Management SR-Behavioral gap		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Women	-0.19*** (0.03)	-0.16*** (0.03)	0.06 (0.23)	-0.16*** (0.03)	-0.14*** (0.03)	-0.04 (0.24)	-0.19*** (0.03)	-0.16*** (0.03)	0.09 (0.24)
Behavioral measure	-0.75*** (0.02)	-0.80*** (0.02)	-0.80*** (0.02)	-0.88*** (0.02)	-0.90*** (0.02)	-0.90*** (0.02)	-0.79*** (0.02)	-0.84*** (0.02)	-0.84*** (0.02)
Years of education		0.03*** (0.00)	0.02*** (0.01)		0.03*** (0.00)	0.03*** (0.01)		0.03*** (0.00)	0.02*** (0.01)
Years of education X Women			0.01 (0.01)			0.01 (0.01)			0.01 (0.01)
Cognitive Ability		0.32*** (0.07)	0.22** (0.09)		0.32*** (0.07)	0.22*** (0.09)		0.33*** (0.07)	0.23*** (0.09)
Cognitive Ability X Women			0.20* (0.10)			0.19* (0.11)			0.20* (0.11)
Social desirability index		0.45*** (0.04)	0.53*** (0.05)		0.45*** (0.04)	0.52*** (0.05)		0.43*** (0.04)	0.51*** (0.05)
Social desirability X Women			-0.16** (0.06)			-0.14** (0.07)			-0.16** (0.07)
Equitable beliefs regarding PSDM abilities		-0.03** (0.01)	-0.06*** (0.02)		-0.02 (0.01)	-0.05** (0.02)		-0.04** (0.01)	-0.05** (0.02)
Equitable beliefs X Women			0.05* (0.03)			0.07** (0.03)			0.04 (0.03)
p(Edu. + Edu. X Women = 0)			0.00			0.00			0.00
p(CA + CA X Women = 0)			0.00			0.00			0.00
p(SD + SD X Women = 0)			0.00			0.00			0.00
p(Equit. + Equit. X Women = 0)			0.64			0.42			0.37
p(Women + Edu. X Women = 0)			0.65			0.97			0.56
p(Women + CA X Women = 0)			0.37			0.67			0.32
p(Women + SD X Women = 0)			0.00			0.00			0.00
p(Women + Equit. X Women = 0)			0.40			0.53			0.43
Observations	4459	4459	4459	4459	4459	4459	4459	4459	4459
R-squared	0.43	0.47	0.47	0.46	0.50	0.50	0.47	0.51	0.51

Notes: Results presented are OLS estimates that include controls for enumerator, age, father's education, mother's education and city. Outcome measures are standardized naive scores. Robust standard errors in parentheses. SR = Self-reported. PSDM = Problem-solving and decision-making. Edu. = Years of education. CA = Cognitive Ability. SD = Social Desirability. Equit. = Equitable beliefs. \*\*\*  $p \leq 0.01$ , \*\*  $p \leq 0.05$ , \*  $p \leq 0.1$

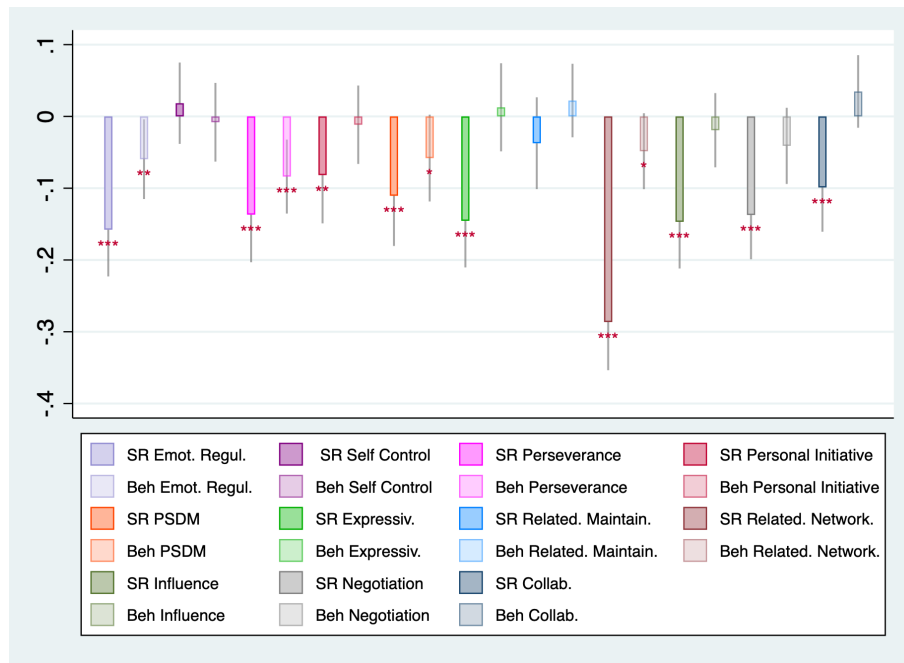
# Figures

Figure 1: Awareness - Female advantage on self-reported and behavioral measures



*Notes:* Results presented are OLS estimates that include controls for enumerator, age, father's education, mother's education and city. Emot. Aware. = Emotional Awareness; Self-Aware. = Self-awareness; Respect. List. = Respectful Listening; Active List. = Active Listening; List. Compr. = Listening Comprehension.  
 \*\*\*  $p \leq 0.01$ , \*\*  $p \leq 0.05$ , \*  $p \leq 0.1$

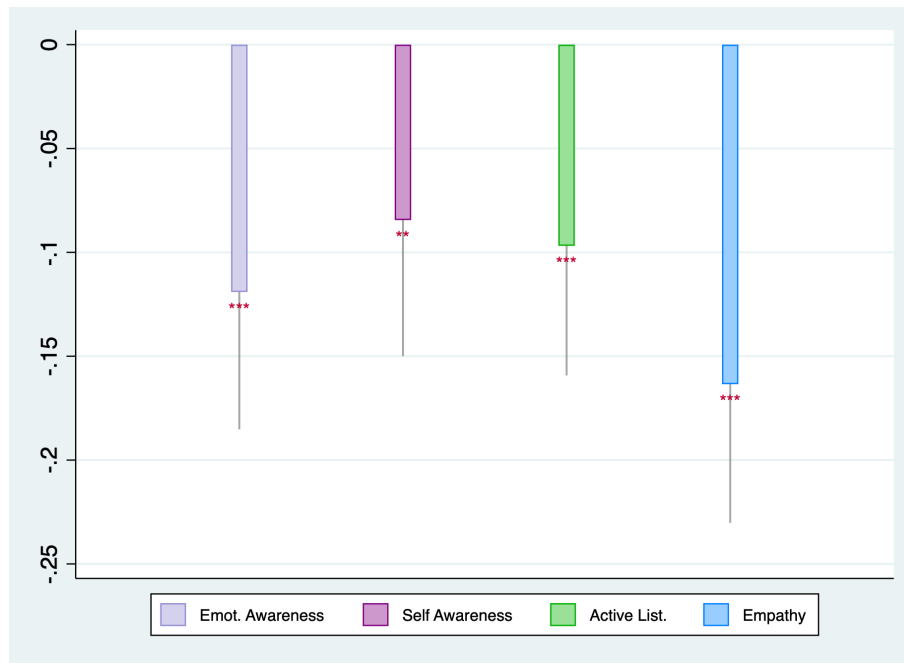
Figure 2: Management - Female advantage on self-reported and behavioral measures



*Notes:* Results presented are OLS estimates that include controls for enumerator, age, father's education, mother's education and city. Emot. Regul. = Emotional Regulation; PSDM = Problem-solving and Decision-making; Expressiv. = Expressiveness; Related. Network. = Relatedness - Networking; Related. Maintain. = Relatedness - Maintaining relationships; Collab. = Collaboration.

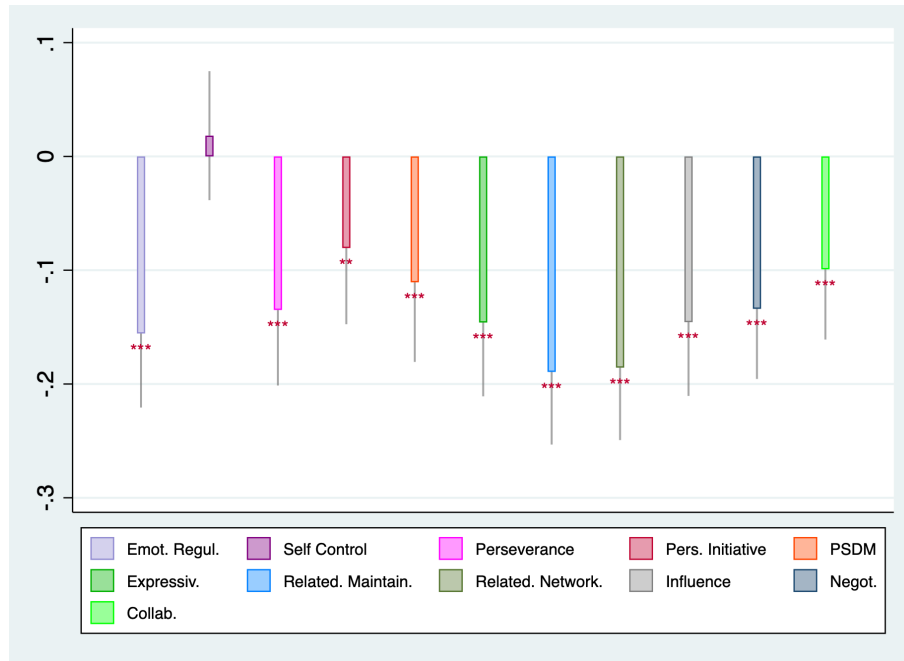
\*\*\*  $p \leq 0.01$ , \*\*  $p \leq 0.05$ , \*  $p \leq 0.1$

Figure 3: Awareness - Female advantage on gap between self-reported and behavioral measures



Notes: Results presented are OLS estimates that include controls for enumerator, age, father's education, mother's education and city. Emot. Awareness = Emotional Awareness. Active List. = Active Listening. \*\*\*  $p \leq 0.01$ , \*\*  $p \leq 0.05$ , \*  $p \leq 0.1$

Figure 4: Management - Female advantage on gap between self-reported and behavioral measures



*Notes:* Results presented are OLS estimates that include controls for enumerator, age, father's education, mother's education and city. Emot. Regul. = Emotional Regulation; Pers. Initiative = Personal Initiative; PSDM = Problem-solving and Decision-making; Expressiv. = Expressiveness; Collab. = Collaboration; Related. Network. = Relatedness - Networking; Related. Maintain. = Relatedness - Maintaining relationships; Negot. = Negotiation; Collab. = Collaboration.

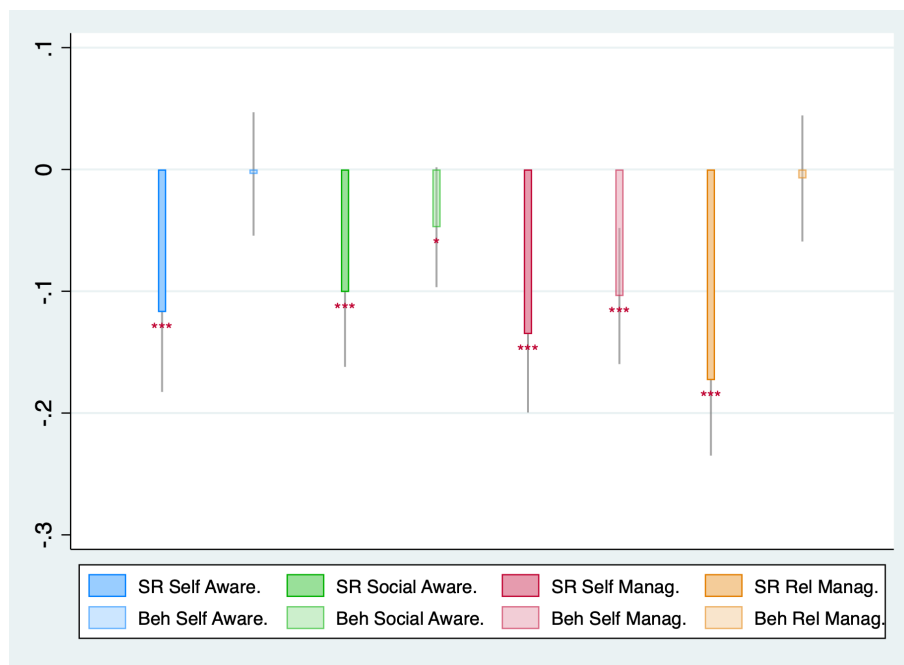
\*\*\*  $p \leq 0.01$ , \*\*  $p \leq 0.05$ , \*  $p \leq 0.1$

# Appendix

Figure A1: Skills definitions with levels of skills aggregation

	Intrapersonal	Interpersonal
Awareness	<p><b>Self-awareness</b>  <i>Emotional Awareness:</i> identifying and accepting one's emotions  <i>Self-awareness:</i> identifying and interpreting one's own thoughts and behaviors and to evaluate one's strengths and weaknesses</p>	<p><b>Social Awareness</b>  <i>Listening:</i> attending to what other people are saying, taking time to understand other's point of view, asking clarifying questions and not interrupting at inappropriate times  <i>Empathy:</i> understand another's viewpoint or thoughts and have emotional concern for another's situation or experience</p>
Management	<p><b>Self-management</b>  <i>Emotional Regulation:</i> maintaining or changing one's own emotions by controlling one's thoughts and behavioral responses  <i>Self-control:</i> focusing one's attention, staying on task, breaking habits, restraining impulses and keeping good self-discipline  <i>Personal Initiative:</i> developing long-term goal, to seek opportunities to improve one's self and to be motivated to put these plans and goals into action  <i>Perseverance:</i> sustaining effort despite setbacks  <i>Problem-solving and decision-making:</i> approaching a problem by gathering information, generating a number of solutions and evaluating the consequences of these solutions before acting</p>	<p><b>Relationship Management</b>  <i>Expressiveness:</i> explaining ideas in a way that others will understand and openly expressing one's opinion  <i>Interpersonal Relatedness:</i> taking actions intended to build trust and benefit others, initiating and maintaining relationships and being respectful, encouraging and caring for others  <i>Interpersonal Influence:</i> communicating in a manner that changes other's perspectives and adapting one's behavior in situationally appropriate ways to influence others  <i>Negotiation:</i> identifying one's own and others' interests and changing one's behaviors as a strategy for resolving interpersonal problems and achieving one's goals  <i>Collaboration:</i> considering different perspective, listening and communicating in groups of two or more people, identifying situations involving group problem-solving and decision-making, and organizing and coordinating team members to create shared plans and goals.</p>

Figure A2: Female advantage on self-reported and behavioral aggregate measures



*Notes:* Results presented are OLS estimates that include controls for enumerator, age, father's education, mother's education and city. Self-aware. = Self-awareness; Social Aware. = Social Awareness; Self Manag. = Self Management; Rel Manag. = Relative Management.

\*\*\*  $p \leq 0.01$ , \*\*  $p \leq 0.05$ , \*  $p \leq 0.1$



Table A1: Descriptive statistics - Disaggregated self-reported measures

Variable	Men		Women		t-test		
	N	(1) Mean/SE	N	(2) Mean/SE	difference (1)-(2)	Min	Max
Emotional Awareness	2231	-0.000 [0.021]	2228	-0.148 [0.021]	0.148***	-5.321	2.128
Self-awareness	2231	0.000 [0.021]	2228	-0.133 [0.020]	0.133***	-6.098	2.091
Emotional Regulation	2231	-0.000 [0.021]	2228	-0.167 [0.021]	0.167***	-7.585	2.290
Self-control	2231	0.000 [0.021]	2228	0.066 [0.020]	-0.066**	-2.352	2.041
Perseverance	2231	-0.000 [0.021]	2228	-0.165 [0.022]	0.165***	-5.549	2.165
Personal initiative	2231	-0.000 [0.021]	2228	-0.133 [0.022]	0.133***	-6.988	2.183
Problem-solving and Decision-making	2231	-0.000 [0.021]	2228	-0.183 [0.021]	0.183***	-7.835	2.436
Listening	2231	-0.000 [0.021]	2228	0.040 [0.021]	-0.040	-2.314	1.521
Listening 2	2231	-0.000 [0.021]	2228	-0.113 [0.021]	0.113***	-5.325	2.043
Empathy	2231	0.000 [0.021]	2228	-0.200 [0.021]	0.200***	-7.799	2.242
Expressiveness	2231	0.000 [0.021]	2228	-0.164 [0.021]	0.164***	-4.428	2.505
Relatedness	2231	0.000 [0.021]	2228	-0.198 [0.021]	0.198***	-5.481	2.216
Influence	2231	-0.000 [0.021]	2228	-0.175 [0.021]	0.175***	-6.429	2.198
Negotiation	2231	-0.000 [0.021]	2228	-0.124 [0.021]	0.124***	-5.083	2.207
Collaboration	2231	0.000 [0.021]	2228	-0.119 [0.020]	0.119***	-5.194	2.136
GSE	2231	-0.000 [0.021]	2228	-0.223 [0.021]	0.223***	-6.240	2.572

Notes: The value displayed for t-tests are the differences in the means across the groups. \*\*\*  $p \leq 0.01$ , \*\*  $p \leq 0.05$ , \*  $p \leq 0.1$ .

Table A2: Descriptive statistics - Disaggregated behavioral measures

Variable	Men		Women		t-test difference (1)-(2)	Min	Max
	N	(1) Mean/SE	N	(2) Mean/SE			
Emotional Awareness	2231	-0.000 [0.021]	2228	0.052 [0.021]	-0.052*	-4.613	1.109
Self-awareness	2231	0.000 [0.021]	2228	-0.020 [0.021]	0.020	-6.844	0.948
Emotional Regulation	2231	-0.000 [0.021]	2228	-0.047 [0.021]	0.047	-2.936	2.823
Self-control	2231	-0.000 [0.021]	2228	-0.022 [0.021]	0.022	-4	2
Perseverance	2231	-0.000 [0.021]	2228	-0.073 [0.020]	0.073**	-1.282	2.537
Personal Initiative	2231	0.000 [0.021]	2228	0.025 [0.021]	-0.025	-6.855	1.002
Problem-solving and Decision-making	2231	-0.000 [0.021]	2228	-0.023 [0.022]	0.023	-4.091	5.447
Listening	2231	-0.000 [0.021]	2228	-0.018 [0.021]	0.018	-2.904	1.041
Active Listening	2231	-0.000 [0.021]	2228	-0.005 [0.021]	0.005	-2.243	1.047
Listening Comprehension	2231	0.000 [0.021]	2228	-0.042 [0.021]	0.042	-5.164	0.813
Empathy	2231	0.000 [0.021]	2228	-0.018 [0.020]	0.018	-4.521	1.818
Expressiveness	2231	-0.000 [0.021]	2228	-0.001 [0.021]	0.001	-3	2
Relatedness	2231	-0.000 [0.021]	2228	0.012 [0.021]	-0.012	-4.270	2.424
Relatedness: Maintaining relationships	2231	-0.000 [0.021]	2228	0.021 [0.021]	-0.021	-3.092	2.500
Relatedness: Initiating Relationships	2231	0.000 [0.021]	2228	-0.010 [0.021]	0.010	-5.817	1.061
Influence	2231	-0.000 [0.021]	2228	-0.033 [0.021]	0.033	-3.258	2.481
Negotiation	2231	-0.000 [0.021]	2228	-0.042 [0.021]	0.042	-3.237	3.222
Collaboration	2231	-0.000 [0.021]	2228	0.030 [0.021]	-0.030	-1.836	1.170
GSE	2231	-0.000 [0.021]	2228	-0.007 [0.021]	0.007	-4	2

Notes: The value displayed for t-tests are the differences in the means across the groups. \*\*\*  $p \leq 0.01$ , \*\*  $p \leq 0.05$ , \*  $p \leq 0.1$ .

Table A3: Intrapersonal/Interpersonal - Self-reported and behavioral measures

	Intrapersonal						Interpersonal					
	Self-reported			Behavioral			Self-reported			Behavioral		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Women	-0.18***	-0.15***	-0.03	-0.04*	-0.05**	-0.19	-0.19***	-0.17***	0.13	-0.02	-0.03	0.13
	(0.03)	(0.03)	(0.24)	(0.03)	(0.03)	(0.21)	(0.03)	(0.03)	(0.24)	(0.02)	(0.02)	(0.19)
Years of education		0.03***	0.02***		0.02***	0.02***		0.03***	0.03***		0.01***	0.01**
		(0.00)	(0.01)		(0.00)	(0.01)		(0.00)	(0.01)		(0.00)	(0.00)
Years of education X Women			0.01			0.00			0.00			0.00
			(0.01)			(0.01)			(0.01)			(0.01)
Cognitive Ability		0.38***	0.27***		0.38***	0.31***		0.37***	0.27***		0.29***	0.26***
		(0.07)	(0.08)		(0.07)	(0.08)		(0.07)	(0.09)		(0.06)	(0.07)
Cognitive Ability X Women			0.23**			0.13			0.19*			0.05
			(0.11)			(0.11)			(0.11)			(0.09)
Social desirability index		0.43***	0.50***		0.12***	0.10***		0.49***	0.56***		0.19***	0.22***
		(0.04)	(0.05)		(0.03)	(0.04)		(0.04)	(0.05)		(0.03)	(0.04)
Social desirability X Women			-0.15**			0.05			-0.15**			-0.07
			(0.07)			(0.05)			(0.07)			(0.05)
Equitable beliefs regarding PSDM abilities		-0.03**	-0.06***		0.03***	0.07***		-0.02	-0.03		0.03**	0.02
		(0.02)	(0.02)		(0.01)	(0.02)		(0.01)	(0.02)		(0.01)	(0.02)
Equitable beliefs X Women			0.06**			-0.06***			0.02			0.01
			(0.03)			(0.02)			(0.03)			(0.02)
p(Edu. + Edu. X Women = 0)			0.00			0.00			0.00			0.01
p(CA + CA X Women = 0)			0.00			0.00			0.00			0.00
p(SD + SD X Women = 0)			0.00			0.00			0.00			0.00
p(Equit. + Equit. X Women = 0)			0.94			0.93			0.78			0.02
p(Women + Edu. X Women = 0)			0.86			0.45			0.51			0.45
p(Women + CA X Women = 0)			0.58			0.60			0.25			0.37
p(Women + SD X Women = 0)			0.00			0.92			0.00			0.28
p(Women + Equit. X Women = 0)			0.56			0.07			0.42			0.36
Observations	4459	4459	4459	4459	4459	4459	4459	4459	4459	4459	4459	4459
R-squared	0.10	0.16	0.16	0.33	0.34	0.34	0.12	0.19	0.19	0.41	0.42	0.42

*Notes:* Results presented are OLS estimates that include controls for enumerator, age, father's education, mother's education and city. Outcome measures are standardized naive scores. Robust standard errors in parentheses. PSDM = Problem-solving and decision-making. Edu. = Years of education. CA = Cognitive Ability. SD = Social Desirability. Equit. = Equitable beliefs. \*\*\*  $p \leq 0.01$ , \*\*  $p \leq 0.05$ , \*  $p \leq 0.1$ . Controls: Enumerator, age, father's education, mother's education and city.

Table A4: Intrapersonal/Interpersonal - Gap between self-reported and behavioral measures

	Intrapersonal SR-Behavioral gap			Interpersonal SR-Behavioral gap		
	(1)	(2)	(3)	(4)	(5)	(6)
Women	-0.17*** (0.03)	-0.15*** (0.03)	-0.01 (0.24)	-0.18*** (0.03)	-0.17*** (0.03)	0.11 (0.24)
Behavioral measure	-0.84*** (0.02)	-0.88*** (0.02)	-0.88*** (0.02)	-0.80*** (0.02)	-0.85*** (0.02)	-0.85*** (0.02)
Years of education		0.03*** (0.00)	0.02*** (0.01)		0.03*** (0.00)	0.03*** (0.01)
Years of education X Women			0.01 (0.01)			0.00 (0.01)
Cognitive Ability		0.34*** (0.07)	0.23*** (0.08)		0.32*** (0.07)	0.23** (0.09)
Cognitive Ability X Women			0.21** (0.11)			0.19* (0.11)
Social desirability index		0.41*** (0.04)	0.49*** (0.05)		0.46*** (0.04)	0.53*** (0.05)
Social desirability X Women			-0.16** (0.07)			-0.14** (0.06)
Equitable beliefs regarding PSDM abilities		-0.04** (0.01)	-0.07*** (0.02)		-0.02 (0.01)	-0.03 (0.02)
Equitable beliefs X Women			0.07** (0.03)			0.02 (0.03)
p(Edu. + Edu. X Women = 0)			0.00			0.00
p(CA + CA X Women = 0)			0.00			0.00
p(SD + SD X Women = 0)			0.00			0.00
p(Equit. + Equit. X Women = 0)			0.93			0.56
p(Women + Edu. X Women = 0)			0.79			0.57
p(Women + CA X Women = 0)			0.54			0.29
p(Women + SD X Women = 0)			0.00			0.00
p(Women + Equit. X Women = 0)			0.43			0.48
Observations	4459	4459	4459	4459	4459	4459
R-squared	0.47	0.50	0.50	0.45	0.49	0.49

*Notes:* Results presented are OLS estimates that include controls for enumerator, age, father's education, mother's education and city. Outcome measures are standardized naive scores. Robust standard errors in parentheses. SR = Self-reported. PSDM = Problem-solving and decision-making. Edu. = Years of education. CA = Cognitive Ability. SD = Social Desirability. Equit. = Equitable beliefs. \*\*\*  $p \leq 0.01$ , \*\*  $p \leq 0.05$ , \*  $p \leq 0.1$ .

Table A5: Psychometrics

	Self-reported					Behavioral				
	Cronbach's Alpha	CFI	TLI	RMSEA	SRMR	Cronbach's Alpha	CFI	TLI	RMSEA	SRMR
Emotional Awareness	0.700	0.998	0.997	0.034	0.019	0.830	1.000	0.990	0.050	0.030
Self-awareness	0.800	0.994	0.992	0.064	0.037	0.840	1.000	1.000	0.040	0.030
Emotional Regulation	0.800	0.997	0.995	0.042	0.025	0.702	0.998	0.997	0.037	0.024
Self-Control	0.880	0.994	0.987	0.085	0.034	No repeat measurement*				
Perseverance	0.750	0.992	0.987	0.054	0.034	No repeat measurement*				
Personal Initiative	0.820	0.992	0.989	0.063	0.042	0.890	0.990	0.990	0.070	0.050
PSDM	0.840	0.986	0.983	0.060	0.045	0.920	0.990	0.990	0.060	0.040
Listening (respectful)	0.610	Too few items in scale <sup>†</sup>				Not measured				
Listening (active listening)	0.770	0.980	0.973	0.062	0.048	0.884	Too few items in scale			
Listening (comprehension)		Not measured				0.755	Too few items in scale			
Empathy	0.710	0.999	0.996	0.051	0.017	0.555	0.983	0.948	0.112	0.047
Expressiveness	0.750	0.995	0.992	0.048	0.030	0.800	0.990	0.980	0.120	0.060
Relatedness (Maintaining)	0.810	0.999	0.998	0.034	0.017	0.678	0.995	0.990	0.071	0.039
Relatedness (Networking)	0.706	0.997	0.992	0.061	0.021	0.813	0.999	0.998	0.030	0.018
Influence	0.810	0.995	0.993	0.051	0.035	0.810	1.000	1.000	0.030	0.020
Negotiation	0.720	0.999	0.998	0.024	0.015	0.860	0.990	0.980	0.080	0.050
Collaboration	0.760	0.998	0.997	0.039	0.022	No repeat measurement				
GSE	0.750	0.993	0.989	0.078	0.040	0.742	1.000	0.999	0.017	0.009

Notes: CFI = Comparative fit index. TLI = Tucker-Lewis index. RMSEA = Root mean squared error of approximation. SRMR = Standardized root mean squared error. PSDM = Problem-solving and decision-making.

Tasks do not include repeat measurement allowing for psychometric analysis.

<sup>†</sup>Use of only three items does not allow for psychometric analysis.

Table A6: Examples of self-reported and behavioral measures

## **Emotional Awareness**

### *Examples of self-reported items*

“I know why my feelings change from one moment to another.”

“I recognize what I am feeling.”

“I can usually describe what I am feeling at the moment in great detail.”

“I try to notice my thoughts without judging them.”

“I am able to accept the thoughts and feelings I have.”

### *Example of a situational judgment test*

”You needed to complete a task for your boss, \$name1, but you were late! Your boss gets angry and says ””how can you be so irresponsible and stupid””?

b. How likely are you to: Notice how your boss’s words made you feel

c. How likely are you to: Notice whether your feelings have caused any physical sensation in your body

c2. How long are you likely to feel stressed or upset: Less than an hour, a few hours, the whole day, a few days, or longer

d. How likely are you to: Identify that you are feeling shame

e. How likely are you to: Reflect on other times that people’s words made you feel this way

## **Self-awareness**

### *Examples of self-reported items*

“I understand my own behaviors.”

“I am aware of my thoughts.”

“I monitor my thinking to ensure it is accurate.”

“I analyze my behavior after I make mistakes.”

### *Example of a situation judgment test*

"You like your job, and customers seem to love you. But your boss, \$name4. has criticized your performance at work. \$pronoun3 only gave you two out of five stars on your performance review."

a. How likely are you to: Stay confident in your abilities

c. How likely are you to: Take time to think about how you can improve

d. How likely are you to: Sit down and talk to \$name4 about why you received poor marks

hx. What skills and strengths do you have that will make you a good candidate for a new job in retail? Please list all of your SKILLS AND STRENGTHS. If you prefer, you can say "Don't know" or "None".

ix. What weaknesses would make you a poor candidate for a new job in retail? Please list all of your WEAKNESSES. If you prefer, you can say "Don't know" or "None".

## **Emotional Regulation**

### *Examples of self-reported items*

"When I feel nervous, I know what to do to feel more relaxed."

"When I feel sad, I know how to take my mind off my problems."

"When I am angry at someone, I can calm down before talking to them."

"When I'm faced with a stressful situation, I make myself think about it in a way that helps me stay calm."

### *Example of a situational judgment test*

You are in charge of the decorations for an annual meeting. Your employee, \$name3, was supposed to bring the flowers and they didn't reach on time for the meeting. The customer is angry at you and threatening to not work with you next year. You feel ashamed that you failed the customer.

"a. How likely are you to: Yell at your employee, \$name3"

b. How likely are you to: Talk to your employee immediately so they know how angry you are.

c. How likely are you to: Become so stressed that you get upset at others

e. How likely are you to: Take time to relax and calm down before you talk to your employee

f. How likely are you to: Discuss your stress with someone you trust

g. How likely are you to: Change how you think about the situation so you're less angry

d. How long are you likely to feel stressed or upset: Less than an hour, a few hours, the whole day, a few days, or longer

### **Self-Control**

*Examples of self-reported items*

“I say inappropriate things.”

“Pleasure and fun sometimes keep me from getting work done.”

“I do things that feel good in the moment, but I will regret later on.”

“Sometimes I can’t stop myself from doing something, even if I know it is wrong.”

*Examples of Enumerator post-survey questions:*

It was easy for respondent to focus on what he/she was doing.

Respondent rushed through the activities without being really attentive.

Task: Continuous Performance Task (CPT-X):

“In this task, you will be shown a list of letters, one by one. Your job here is, to figure out whether each letter is an X, or not an X. Each time you see an X. Do NOT touch the screen. If you are shown another letter, you answer by touching the screen quickly. Try and answer quickly while maintaining focus. Touch the screen when you are ready to start. You will start by doing some exercises as examples.”

### **Perseverance**

*Examples of self-reported items*

“I finish whatever I begin.”

“Setbacks don’t discourage me.”

“I am diligent.”

“When work is difficult, I keep up my effort.”

*Triangle Task*

After viewing example puzzles, Which version of the game do you want to play for the next: Easy or Difficult

You have 60 seconds to count the number of triangles in the figure.



Would you like to continue, or end the game?

two practice rounds, four test rounds

### **Personal Initiative**

*Examples of self-reported items*

“I actively tackle problems.”

“Whenever something goes wrong, I search for a solution immediately.”

“Whenever there is a chance to get actively involved, I take it.”

“I take action immediately even when others don’t.”

*Example of a situational judgment test*

SJT1. Imagine you want to open a clothing shop and you have some savings. Unfortunately, you know very little about the clothing business. You ask your friends or family, and they also do not know about the business.

b. How likely is it that you will: Do research on clothing shops online in your spare time

c. How likely is it that you will: Look for a training

You do not know any clothing shop owners,

d. How likely is it that you will: Find some clothing shop owners to ask for advice

f. How likely is it that you will: Open the shop and learn the business as you go.

### **Problem-solving and decision-making**

*Examples of self-reported items*

“I solve most problems if I put in the necessary effort.”

“I can find creative solutions to unplanned problems.”

“I can always solve difficult problems if I try hard enough.”

“If someone needs input on a problem, I can come up with many suggestions.”

*Example of a situational judgment test*

"You are part of a group organizing an annual festival for the surrounding five neighborhoods! \$name1 was in charge of publicizing the event, but you just found out that most don't know when the event is, some have never heard of it, and hardly anyone is planning to come! The event is in two days."

- b. How likely is it that you will: Contact \$name1 to ask what went wrong?
- c. How likely is it that you will: Contact \$name1 to ask what methods of advertising were used?
- d. How likely is it that you will: Think of as many ideas as possible for solving this problem.
- e. How likely is it that you will: Contact friends to ask for help coming up with as many ideas as possible.
- f. How likely is it that you will: Solve this problem and have high event attendance

## **Listening**

### *Examples of self-reported items*

"I ask questions to understand the other person's position on an issue."

"When I am listening to someone, I make sure they know I am interested in what they are saying."

"When I am listening to someone, I show them that I am open to their ideas."

"When I am listening to someone, I ask questions that show my understanding of what they are saying."

"I begin talking before the other person finishes talking."

"If I have something to say that is important, I will interrupt the other person." (reverse)

"I share my opinion without listening to others' opinions." (reverse)

### *Example of a situational judgment test*

"Imagine that I am your neighbor. I just found about a new business that you would like to learn about! Feel free to ask questions if you want to know more about the business. Ready?"

My friend, \$name5, just started a business where he processes rice and sells different products made of rice. They are making a lot of money: Tsh 45,000 per week. They attended a training for a few hours a day for two months. The training is held every six months in training centers all over our region. The best part is that little investment or equipment is required. Two other friends went into the same business- one made the same amount- the other made a bit less because they made some mistakes. Should you pursue this business?"

four active listening questions: e.g. Enumerator: as you were saying the story, did the respondent show they were listening, by using body language, e.g. nodding?

Enumerator: as you were saying the story, did the respondent show they were listening by making comments, e.g. “oh really” “yes” “mmhmm” etc. ?

4 Listening comprehension questions: e.g. What income did \$name5 make per week?

## **Empathy**

### *Examples of self-reported items*

“When I’m upset at someone, I usually try to imagine myself in their situation to better understand them.”

“Before judging somebody, I try to imagine how I would feel if I were in their place.”

“I ask questions to understand the other person’s position on a given issue.”

“I always try to understand the feelings of people I trust.”

“If someone is hurt, it makes me upset.”

*Task:* Rate level of pleasure and arousal for self and the other individual after hearing a list of scenarios

## **Expressiveness**

### *Examples of self-reported items*

“I ask for what I need when I need it.”

“I think it’s good to ask for what I want.”

“I find it easy to explain my perspective to others.”

### *Example of a situational judgment test*

Imagine you are attending a community meeting, and they are deciding whether to build a school, a clinic, or a road. The meeting has 30 men and 30 women, including your spouse.

How likely are you to: Stand up and share your opinion about the road

You are curious about how long each project will take: How likely are you to speak up and ask this question?

You have the idea that everyone should vote to decide which project to choose: How likely are you to: Discuss your idea with the person sitting next to you?

How likely are you to: Share your idea with the group without hesitation?

### **Relatedness: two dimensions**

#### *Examples of self-reported items*

“I listen patiently when people tell me their problems.”

“When I see that someone is going through a difficult time, I help out the best I can.”

“I give my friends and family encouragement when they need it.”

#### *Example of a situational judgment test*

A customer, \$name3, who you have seen before but don't know well comes to your shop. \$name3 really wants to buy rice but they have had troubles this week and they don't have enough money to pay this time. There are others in line and \$name3 is taking time.

Which picture best describes your tone?

Which picture best describes your tone?

How likely are you to

Dismiss \$name3

Tell \$name3 to return when they have money

Allow \$name3 to pay back later

Make sure \$name3 know you are assessing their trustworthiness

Encourage \$name3 to share why they cannot pay

Reassure \$name3 that things will get better

### **Influence**

#### *Examples of self-reported items*

“Other people do what I ask them to do.”

“When someone disagrees with me, I know how to adjust my argument to change their opinion.”

“I am good at getting people to help me when I need it.”

*Example of a situational judgment test*

You want to start a new business, making banana chips with a new method. To start the business, you need your family’s support because it will affect their financial situation. Currently your family does not want you to start the business.

How likely is it that you will: Try to convince your family to let you start the business

How likely is it that you will: Ask questions to understand why your family opposes you

How likely is it that you will: Analyze your family’s behavior carefully, to decide the best time to convince them

How likely is it that you will: Discuss the benefits and consequences of starting the business with them

Would you use any other methods to persuade your family?

Now imagine that your brother recently failed in his business. Would you use any other methods to persuade your family?

How likely is it that you will: Not be able to change your family’s perspective.

## **Negotiation**

*Examples of self-reported items*

“When I disagree with someone, I try to understand how that person feels.”

“When I disagree with someone, I am still able to listen to the other person’s perspective.”

“When I disagree with someone, I am able to give up some things I want to solve our disagreement.”

*Example of a situational judgment test*

Your work has become busier and you have less time for household responsibilities. If you have help at home, your income could increase! However, your 15 year old son does not want to help with cleaning or caring for the younger children. If he has extra time, he just wants to play football with his friends.

How likely is it that you will: Accept the situation and don’t say anything

How likely is it that you will: Tell him he has to do some household work and has no choice

How likely is it that you will: Explain that if he helps, the whole family will benefit

How likely is it that you will: Allow him to go play football if he completes his responsibilities

## **Collaboration**

*Examples of self-reported items*

“When I work with others, I tell others my ideas and ask for theirs in return.”

“I can tell when a problem should be solved by a team of many people instead of one person alone.”

“When I don’t know a solution to a problem, I can brainstorm with a group of people to get better ideas.”

*Task:* Simulated SMS conversation to find a market stand: “Looks like the group has sent you a message. Which of these responses, is most like how you would respond in this situation?”