



Cognitive and non-cognitive skills in developing countries

Anne Hilger

► To cite this version:

Anne Hilger. Cognitive and non-cognitive skills in developing countries. Economics and Finance. Université Paris sciences et lettres, 2018. English. NNT : 2018PSLEH077 . tel-03168267

HAL Id: tel-03168267

<https://theses.hal.science/tel-03168267>

Submitted on 12 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'Ecole des hautes études en sciences sociales

Cognitive and non-cognitive skills in developing countries

Ecole doctorale n°465

ECOLE DOCTORALE ECONOMIE PANTHEON SORBONNE

Spécialité Analyse et politiques économiques

Soutenue par Anne **HILGER**
le 03 juillet 2018

Dirigée par Flore **GUBERT**
Co-dirigée par **Christophe Jalil**
NORDMAN

COMPOSITION DU JURY :

Mme. HUILLERY Élise
Université Paris -Dauphine, Rapporteur

M. LANJOUW Peter
Vrije Universiteit Amsterdam, Rapporteur

M. BERNARD Tanguy
Université de Bordeaux, Membre du jury

M. SERNEELS Pieter
University of East Anglia, Membre du jury

Mme. GUBERT Flore
IRD-DIAL, Paris School of Economics,
Membre du jury

M. NORDMAN Christophe Jalil
IRD-DIAL, Membre du jury

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text. Chapter 2 of this thesis is co-authored with Dr. Christophe Jalil Nordman (IRD-DIAL, IFP, IZA) and Leopold R. Sarr (World Bank). I made substantial contributions to this chapter, including to the origin of the research question. I conducted all empirical analysis and most of the writing of the chapter. Chapter 3 of this thesis is co-authored with Dr. Christophe Jalil Nordman (IRD-DIAL, IFP, IZA). I made substantial contributions to the chapter, including to the research question, data collection, data cleaning, empirical analysis, and writing. This thesis has not been submitted for any other degree or professional qualification.

(Anne Hilger)

In Erinnerung an Richard Hilger

Acknowledgments

I am sincerely grateful to my supervisor, Christophe Jalil Nordman, for his friendship, guidance, and encouragement throughout my research. He has given me the chance to conduct field work in Tamil Nadu, India, and has been patient and encouraging at the same time. I am grateful for his tremendous feedback, support, and unwavering confidence in my learning. I am also grateful to my second supervisor, Flore Gubert, for her incredibly helpful guidance and support.

I would like to thank Élise Huillery and Peter Lanjouw for accepting to be my referees and for their helpful comments during my pre-defense. I would also like to thank David Margolis for providing feedback and support throughout the thesis as a member of my thesis committee. Finally, I would like to thank Tanguy Bernard and Pieter Serneels for being part of my thesis jury.

This thesis would not have been possible without the support of literally everyone at DIAL. I am grateful for positive encouragement, lunch breaks, and football matches. Being part of the DIAL family is truly an honor. Special thanks go out to Quynh for unwavering support and friendship, and for dealing with all of my empirical questions. Thanks also to Loïc and Anne for handling the administrative side of things and helping me manage my numerous trips.

I am also grateful to many of my fellow PhD students at PSE for their friendship and advice. Further, thanks to Sylvie Lambert for effectively and efficiently managing the Doctoral School and for helping me navigate multiple administrative hurdles; thanks to Véronique Guillotin for administrative support on the PSE side of things.

I would like to thank the Institut Français de Pondichéry (IFP) for hosting me for 4 months in the fall of 2016 to conduct field work in Tamil Nadu. I am really grateful to everybody I met at IFP, but especially Seb, Youna, Venkat, and Antoni. My field work would not have been possible without the financial support through a mobility grant from Paris School of Economics and a field work grant from the EHESS. Thanks also to Cepremap for providing financial support to transcribe the qualitative interviews.

I am grateful to the monitoring and evaluation team of Oxford Policy Management for involving me in the *MUVA* program and letting me use their data for Chapter 1. Special thanks to Jana for comments and suggestions, but also to Paul, Luize, and the Maputo-based *MUVA* staff.

Thanks also to the 2017-2018 LSE teaching crew, Geoff, Greta, and Moritz, for discussions about development and life. Sean, thanks for your glorious map-making skills. Aline, thanks for useful comments and excellent proofreading. Hannah, thank you for commiserating when necessary, encouraging at all other times, and providing much needed work-life balance.

Danke an meine Familie für eure außergewöhnliche Unterstützung nicht nur auf diesem, sondern allen meinen Wegen. Diese These wäre ohne euch und euren Glauben an meine Fähigkeiten nicht möglich gewesen.

Finally, James, thank you, for being my favorite editor, for opening my eyes to the beauty of the Oxford comma, for your unconditional support throughout the years, and for convincing me ever so often that starting a buffalo farm might not be a suitable alternative to a PhD.

Abstract

This dissertation examines the role that cognitive and non-cognitive skills play in developing countries along three axes: measurement of these skills, wage returns to them, and as determinants of levels of trust.

Chapter 1 provides a measurement perspective, contrasting skills as measured by self-assessments to those captured by observational exercises. Using panel data from two cohorts of a skills training program in Mozambique, I find that the self-assessment and the observational exercises measure different concepts; the former captures underlying personality traits, while the latter relates more closely to personality states, which are malleable as a result of the intervention. The paper thus highlights the importance of knowing exactly what is measured when assessing program impact.

Chapter 2 uses a novel matched employer-employee data set representing the formal sector in Bangladesh to provide descriptive evidence of both the relative importance of cognitive and non-cognitive skills in this part of the labor market and the interplay between skills and hiring channels in determining wages. While cognitive skills (literacy) do affect wages by enabling workers to use formal hiring channels, they have no additional wage return. Non-cognitive skills, on the other hand, do not affect hiring channels, but they do enjoy a positive wage return. This wage return differs by hiring channel: those hired through formal channels benefit from higher returns to openness to experience but lower returns to conscientiousness and hostile attribution bias; those hired through networks enjoy higher wages for higher levels of emotional stability, but they are also punished for higher hostile attribution bias. This is in line with occupational levels being hired predominantly through one channel or the other. We provide suggestive evidence that employers might use hiring channels differently, depending on what skill they deem important; employers valuing communication skills, which could arguably be observed during selection interviews, are associated with a larger within-firm wage gap between formal and network hires, while the importance of teamwork, a skill that is more difficult to observe at the hiring stage, is associated with a smaller wage gap.

Chapter 3 uses the 2016 demonetization policy in India, an unexpected and unforeseeable exogenous variation that had direct effects on network usage but not on interpersonal trust, to causally identify the effect of two measures of social networks on determining trust. It is thereby able to disentangle the relationship

between trust and social interactions, concepts which are inherently interrelated. We use first-hand quantitative and qualitative data from rural South India and control for a variety of individual characteristics that could influence network formation and trust, such as personality traits and cognitive ability. We find that social interactions only had a significant effect on levels of trust among men. Further, we find important differences along the lines of caste membership. Among lower castes, who live in homogeneous neighborhoods and rely predominantly on their neighbors and employers to cope with the shock, making use of one's network more intensely increases levels of trust placed in neighbors. Among middle castes, who live in more heterogeneous neighborhoods and rely largely on members of their own caste to cope, a larger network size leads to higher levels of trust placed in kin among employees but lower levels of trust in neighbors (who tend to be more dissimilar). This paper thus shows that social interactions can foster trust, though this is dependent on the type of interaction occurring.

Keywords: Development Economics, Labor Economics, Cognitive and non-cognitive skills

Résumé

Cette thèse examine le rôle joué par les compétences cognitives et non cognitives dans les pays en développement, selon trois axes : la mesure de ces compétences, leurs rendements salariaux et les déterminants de la confiance interpersonnelle. Le premier chapitre fournit une perspective sur la mesure. Il fait la comparaison entre deux types de mesures de ces compétences : auto-évaluations et observations d'exercices. En utilisant des données de panel provenant de deux cohortes d'individus récipiendaires d'un programme de formation des compétences au Mozambique, je trouve que l'auto-évaluation et les observations d'exercices mesurent des concepts différents : le premier type de mesure capture les traits de personnalité ; le deuxième semble mieux adapté aux évaluations de programme car il reflète des compétences qui seraient malléables à la suite d'une intervention. Le chapitre souligne ainsi l'importance de déterminer exactement ce qui doit être mesuré lors de l'évaluation de l'impact d'un programme.

Le deuxième chapitre tire profit d'une nouvelle base de données appariées employeurs-employés, représentant le secteur formel au Bangladesh. Le chapitre fournit une analyse de l'importance relative des compétences cognitives et non cognitives dans ce marché du travail et de l'interaction entre ces compétences et la méthode d'embauche (formelle ou informelle, c'est-à-dire par le réseau social) pour la détermination des salaires. D'une part, les compétences cognitives (le fait de savoir lire et écrire) sont positivement corrélées aux salaires de manière indirecte, car elles permettent aux travailleurs d'accéder aux méthodes formelles d'embauche ; les travailleurs n'en tirent par la suite aucune rémunération supplémentaire. D'autre part, les résultats révèlent des rendements positifs des compétences non cognitives sur les salaires, qui varient selon la méthode d'embauche utilisée : ceux qui sont embauchés par des voies formelles bénéficient de rendements plus élevés de l'ouverture à l'expérience, mais un rendement moindre des traits « conscientiousness » et « hostile attribution bias ». Ceux qui sont embauchés à travers les réseaux bénéficient de salaires plus élevés pour des niveaux plus élevés de stabilité émotionnelle, mais ils sont également punis pour un biais d'attribution hostile plus élevé. Cela s'explique par le fait que les catégories socio-professionnelles sont recrutées principalement par une voie plutôt qu'une autre. Nous montrons que les employeurs utilisent les méthodes d'embauche différemment, en fonction des compétences qu'ils jugent importantes : les employeurs valorisant les compétences

telles que la capacité de communiquer, une compétence qui pourrait être observée lors des entretiens de sélection, sont associés à des entreprises produisant un écart salarial intra-entreprise plus large entre travailleurs embauchés par voies formelles et ceux embauchés par le réseau. La compétence du travail en équipe en revanche, une compétence plus difficile à observer au moment de l'embauche, est associée à un écart salarial plus faible.

Le troisième chapitre utilise la politique de démonétisation en Inde, un choc exogène inattendu et imprévisible, pour identifier d'une manière causale l'effet des réseaux sociaux sur la détermination de la confiance interpersonnelle. Ce choc a eu des effets directs sur l'usage des réseaux interpersonnels, mais pas sur la confiance interpersonnelle. En utilisant ce choc, nous tentons de démêler les mécanismes de la confiance et de la formation et l'usage des réseaux sociaux, des concepts qui sont intimement liés. Nous avons recours à des données quantitatives et qualitatives nouvellement collectées dans une zone rurale de l'Inde du Sud. Les données permettent de contrôler un ensemble de caractéristiques individuelles, en particulier les traits de personnalité et les capacités cognitives, qui sont susceptibles d'influencer la formation et l'usage des réseaux sociaux et la confiance. Les résultats montrent que les interactions sociales déterminent la confiance, en particulier pour les hommes. De plus, des différences importantes apparaissent entre castes. Parmi les basses castes, qui vivent dans des quartiers homogènes socialement et qui dépendent de leur voisinage et de leurs employeurs pour faire face à un choc, une utilisation plus intensive de leur réseau a eu pour conséquence d'augmenter le niveau de confiance qu'ils placent en leurs voisins. Pour les castes intermédiaires, qui vivent dans des quartiers plus hétérogènes et dépendent principalement des autres membres de leur caste, une taille plus grande du réseau est liée à un niveau plus élevé de confiance témoigné entre employés d'une même communauté ou famille. Cet article montre ainsi que les interactions sociales sont en mesure de promouvoir la confiance, et souligne la nécessité d'une prise en compte fine de la stratification sociale sur un tel sujet.

Mots clés : Économie du développement, Économie du travail, Compétences cognitives et non cognitives

Contents

Introduction	1
Appendix for the Introduction	11
1 Using self-assessments and observations to capture non-cognitive skills: Insights from a skills training program in Mozambique	13
1.1 Introduction	14
1.2 Context	18
1.2.1 Mozambique	18
1.2.2 <i>MUVA</i> Attitude program	19
1.3 Data and descriptive statistics	19
1.3.1 Data	19
1.3.2 Descriptive statistics of the panel sample	21
1.4 Methodology	21
1.4.1 Observational exercises	22
1.4.2 Self-assessments	24
1.5 Measuring non-cognitive skills – application	29
1.5.1 Correcting for acquiescence	29
1.5.2 Aggregating through EFA	31
1.5.3 Choosing a self-assessment measure	32
1.5.4 Inter-rater reliability in observational exercises	35
1.5.5 Aggregating the observational exercises	37
1.6 Comparing measures of non-cognitive skills	38
1.6.1 Correlations	38
1.6.2 Rank order changes	41
1.7 Discussion and conclusion	44
Appendix for Chapter 1	46
2 Cognitive and non-cognitive skills, hiring channels, and wages in Bangladesh	55
2.1 Introduction	56
2.2 Literature and conceptual framework	58
2.2.1 Literature review	58
2.2.2 Conceptual framework	61

2.3	Data	63
2.4	Methodology	65
2.4.1	Returns to skills	65
2.4.2	Skills and the type of hiring channel	67
2.5	Results	70
2.5.1	Descriptive statistics	70
2.5.2	Wage returns to different types of skills	71
2.5.3	Choice of hiring channel	75
2.5.4	Skills and endogenous hiring channel	77
2.5.5	Starting wages and wage growth	81
2.5.6	Determinants of the within-firm wage gap between formal and network hires	84
2.6	Conclusion	88
	Appendix for Chapter 2	91
3	The determinants of trust: Evidence from rural South India	101
3.1	Introduction	102
3.2	Background	105
3.2.1	Tamil Nadu	105
3.2.2	Demonetization	106
3.3	Conceptual framework	107
3.4	Data and descriptive statistics	109
3.4.1	Description of the survey	109
3.4.2	Construction of the social network variables	111
3.4.3	Measuring trust	113
3.4.4	Descriptive statistics	115
3.5	Empirical strategy	116
3.5.1	OLS	116
3.5.2	Instrumental variables	116
3.6	Results	122
3.6.1	OLS estimates of the determinants of trust	122
3.6.2	First stage results: determinants of network size and density	123
3.6.3	Second stage results: the causal determinants of trust	125
3.6.4	Heterogeneity analysis	129
3.6.5	Robustness checks	136
3.7	Conclusion	140
	Appendix for Chapter 3	142
	Conclusion	153

Introduction

Motivation

Several decades of research in economics have highlighted the importance of skills for life and labor market outcomes. Starting from the seminal work of Becker (1964) and Mincer (1974), human capital, the abilities and qualities that make people economically productive, has been at the center of the economic literature in labor and education. Traditionally, human capital has simply been approximated by levels of education, but it has become more and more obvious over time that this approach ignores the innate multidimensionality of human capital, which holistically refers not only to (technical) knowledge, as might be approximated by levels of schooling and (work) experience, but also to other dimensions, such as general intelligence, motivation, the ability to work diligently (or show up in the first place), and even an individual's health. Lack of data meant that empirical studies have oftentimes ignored this multidimensionality, with cognitive and non-cognitive skills simply part of the 'unobservables'.

This thesis embraces recent extensions of human capital to a more holistic picture and aims to empirically understand the role that both cognitive and non-cognitive skills play in developing countries. Addressing the measurement and role of these skills in a developing country context is important for two reasons. First, it is unclear that these skills should be rewarded similarly in developed and developing countries. For example, jobs in developing countries often have a different task-structure, labor market segmentation is stricter, informal jobs are more prevalent (and often more common than formal forms of employment), and a larger share of the labor market tends to be self-employed instead of engaging in wage labor. Employers might thus reward skills that enable the precise execution of tasks more than skills that deal with intellectual curiosity and independent working. In addition, other factors, such as social hierarchies, might guide wage setting. Second, it is unclear how these skills can be developed in a setting characterized by low levels of quality education. As such, it is crucial to extend the evidence base on the importance of non-cognitive skills to developing countries, which is the goal of this thesis.

What are cognitive skills?

Cognitive skills can be defined as the “ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought” (Neisser et al., 1996, p.1). This definition includes both facets of cognition: fluid intelligence (the rate at which people learn) and crystallized intelligence (knowledge learned). Cognitive skills have regularly been approximated via intelligence tests in developed countries, often non-verbal tests such as Raven’s Progressive Matrices, which capture ideas of fluid intelligence, or standardized test scores, such as scores on college admission tests, which relate more to crystallized intelligence.

Measurement and effects on outcomes

Traditional intelligence tests aim to approximate the measurement of g , a general factor driving human intelligence. Given a lack of these tests in many developing countries, cognitive skills have often been approximated via numeracy and literacy tests, which refer more closely to crystallized than fluid intelligence.

In developed countries, cognitive skills have been associated with higher labor earnings (e.g. Hanushek et al., 2015; Heckman et al., 2006; Lindqvist and Vestman, 2011; Vignoles and McIntosh, 2001). Hanushek and Woessmann (2008) provide a global overview of the role of cognitive skills in developing countries. The authors conclude that it is the possession of cognitive skills, rather than mere school attainment, that is most powerfully related to individual earnings. Most of their data indeed rely on literacy and numeracy tests to approximate cognitive skills. This is potentially dangerous, as literacy and numeracy are learning outcomes that capture not just a person’s general intelligence but also facets of motivation, or quality of education received. Still, given the lack of more appropriate data, it is a common proxy.

What are non-cognitive skills?

Non-cognitive skills have received even more attention in recent years, as a quasi-panacea crucial not only for labor but for life outcomes in general. They have appeared under a variety of names, from ‘soft skills’ to ‘21st century skills’ to ‘socio-emotional skills’. ‘Non-cognitive skills’ then are thought to capture skills, potentially changeable characteristics, as well as personality traits, considered to be rather stable over the life course. Distinguishing this strictly between cognitive and non-cognitive skills is of course a simplification, as non-cognitive skills are also cognitive in the sense that information processing underlies many personality traits (Bandura, 1999; Mischel and Shoda, 1999). Further, ample evidence shows that cognitive and non-cognitive skills interact in producing ‘cognitive’ outcomes, such as school performance or performance on intelligence tests (Borghans et al., 2008b;

Duckworth et al., 2011; Duckworth and Seligman, 2005). Indeed, non-cognitive skills are harder to define than those that are considered cognitive. Messick (1978, p.2) notes that “once the term cognitive is appropriated to refer to intellectual abilities and subject-matter achievement in conventional school areas [...] the term noncognitive comes to the fore by default to describe everything else”. More recent work has attempted to distinguish between skills, the changeable part of non-cognitive skills, and personality traits, being defined as “enduring patterns of thoughts, feelings, and behaviors that reflect the tendency to respond in certain ways under certain circumstances” (Roberts, 2009, p.140).

Measurement and effects on outcomes

The first method of measuring non-cognitive skills relies on *self-assessments*. It stems from psychology and the so-called lexical hypothesis: traits which are important in people’s lives tend to be captured in language (Golsteyn and Schildberg-Hörisch, 2017). The Big Five personality test, usually attributed to Allport and Odbert (1936), is an example of one of the most widely accepted self-assessment instruments and has been replicated across cultures (John and Srivastava, 1999) and developmental stages of the life course (Soto et al., 2008). This test, widely used among psychologists, consists of five dimensions: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism (or its inverse, emotional stability) oftentimes summarized in the acronym OCEAN, under which all more narrowly defined traits could be classified (Costa and McCrae, 1992). Table A0.1 in the Appendix provides an overview of the five traits and their underlying characteristics. Though self-assessments are the most commonly used form of measurement, they are inherently subject to bias, which refers to measuring the true skills with error. One cause of such bias is the reliance of self-assessments on a “correct” assessment of oneself, making the information relatively subjective (McConaughy and Ritter, 1995). Further, information captured by a self-assessment is usually retrospective (Shapiro and Kratochwill, 2000) and specific to the reference group implied when assessing oneself. In addition, self-assessments can be prone to response bias factors, such as faking, a tendency to agree or disagree no matter the question (acquiescence), or social desirability bias (Merrell, 2003).

The second method of assessment, *observable behaviors*, mostly looks at (usually adverse) behaviors, such as delinquent behavior, teenage pregnancy, or smoking among teenagers, and then uses these observable behaviors as a proxy for the level of non-cognitive skills, as negative behaviors and skills are thought (and have been shown) to be negatively correlated (Elkins et al., 2006). This method thus implicitly makes the assumption that individuals engage in negative behaviors due to a lack of certain non-cognitive skills.

Lastly, *teacher and parent evaluations*, such as rating scales, have been used extensively in the education literature (Crowe et al., 2011; Humphrey et al., 2011).

Ideally, these rely on multiple observations over an extended period of time and are thought to be an efficient tool for observing other people’s behavior, but observations can be expensive in terms of time needed; Doll and Elliott (1994) use classroom observations during free play periods and find that at least five observations of thirty minutes each across several weeks are needed for an accurate assessment. While observations are not subject to response bias from participants, they are prone to observer bias. Achenbach et al. (1987) show that agreement among raters is higher among similar raters (i.e. a pair of teachers) than among dissimilar raters.

The fascination with non-cognitive skills stems in large part from an understanding that they have been associated with a number of positive outcomes in developed countries. These include a positive effect on educational attainment (Duckworth et al., 2007), a positive effect on wages beyond that of cognitive ability (Heckman and Rubinstein, 2001; Heineck and Anger, 2010) and as a general predictor of labor market outcomes (e.g. Almlund et al., 2011; Borghans et al., 2008b), such as occupational choice (Cobb-Clark and Tan, 2011) and job search methods (Caliendo et al., 2015). The reader may refer to Kautz et al. (2014) for a recent overview.

The literature looking at the association of levels of non-cognitive skills with various outcomes in developing countries is small in comparison. Blom and Saeki (2011) find evidence that employers of engineers in India stress interpersonal skills such as reliability and willingness to learn above cognitive skills such as literacy and numeracy. In Peru, Díaz et al. (2013) find that returns to perseverance are as high as returns to average cognitive ability. Other papers have found rather mixed evidence: Glewwe et al. (2017) show that, in China, both cognitive and non-cognitive skills are important for the school-to-work transition, but they do not predict wages. Cunningham et al. (2016) use data for four Latin American countries (Bolivia, Colombia, El Salvador, and Peru) to show that non-cognitive skills are more important than cognitive skills in determining labor force participation, though some non-cognitive skills are also correlated with labor earnings in some countries; and Acosta et al. (2015) find a larger impact of cognitive than non-cognitive skills for labor market outcomes in Colombia. In developing countries, non-cognitive skills have further been related to a higher rate of technology adoption among farmers (Abay et al., 2017).

While research on the effects of non-cognitive skills on life outcomes has largely focused on developed countries, so have policy makers attempting to foster these skills. Sanchez Puerta et al. (2016) take stock of programs worldwide aiming to build non-cognitive skills. Of the 86 programs included in the overview, only about 30 percent were from the developing world, with the majority of programs located in the United States. Further, most developing world programs focused on those already out of school, at which point malleability of skills is less likely.

Contribution

This thesis contributes to the small but growing literature on non-cognitive skills in developing countries in a number of ways. The first chapter provides a measurement perspective. Using the monitoring and evaluation data from a program among youth in Mozambique aimed to foster non-cognitive skills, I contrast two different methods of capturing these skills: observational exercises (individual and group exercises during which the participants were observed and rated on their performances on a number of different skills) and self-assessments of (sometimes overlapping) dimensions by the same individuals. Given the lack of data capturing non-cognitive skills, a measurement paper trying to better understand what is being measured and if traditional forms, such as self-assessments, can work in the context of a low-literacy population that might not be prone to the type of self-reflection required for such an exercise, adds considerable value to the field. Having found that self-assessments can work given a few response bias adjustments, the second chapter provides a more classical approach, by estimating the wage returns to non-cognitive skills in Bangladesh. Little is known so far regarding wage returns to non-cognitive skills (in our case the Big Five personality traits, grit, and hostile attribution bias) in developing countries. Taking into account a potentially mitigating factor, the channel through which the individual was hired, as well as firm-specific heterogeneity in wage setting through firm fixed effects, we find that non-cognitive skills carry a wage return in Bangladesh. The third chapter builds on a newly collected Indian data set. I had the chance to contribute to the questionnaire and add a longer non-cognitive skills module, capturing the Big Five but with seven instead of three questions per dimension to improve the internal validity of the dimensions. The data further include a real cognitive assessment, Raven's Coloured Matrices, a non-verbal test assessing fluid intelligence, in addition to numeracy and literacy tests, assessing crystallized intelligence. This chapter does not use the non-cognitive skill measures as the covariates of interest but focuses instead on a natural experiment to causally identify the effect of social interactions on different measures of trust. At the same time, the inclusion of the cognitive and non-cognitive skill variables allow me to control for individual heterogeneity that might affect trust formation.

Data

Any empirical work on cognitive and non-cognitive skills in developing countries faces the challenge of finding appropriate data sources capturing these skills. This thesis relies on a number of new data sets, some collected as part of this research. The following section briefly describes those data sets and the measures of cognitive and non-cognitive skills included in them.

Chapter 1: Mozambique ‘*MUVA* Attitude’ monitoring data

Chapter 1 is based on a unique data set collected as part of the monitoring and evaluation system of an intervention implemented in urban Mozambique (Maputo and Beira) aimed at improving the employability skills of young people. The data are not representative of the Mozambican population, but do allow me to contrast different measures of non-cognitive skills for the same individual. Data were collected at three points in time: before the start of the program (baseline), after intensive non-cognitive skills training (midline), and after an additional technical training component (endline). I combine the first two cycles of the monitoring and evaluation data, providing panel data for 354 individuals.¹

The data set includes two non-cognitive skills assessments. The first type of instrument are observational exercises (applied to individuals and in groups). The group exercise consisted of a short task that participants were asked to complete in groups of six or seven participants. Three raters assessed the individual members’ performance on the following skills: listening, negotiation, motivation to work with others, creativity, flexibility, and personal motivation. The individual exercise consisted of an individual presentation in front of the raters, in which raters judged the participants in terms of display of body language, logical argumentation, professional attitude, and speaking with confidence. The second type of instrument, a self-assessment, consisted of a questionnaire (40 items) about participants’ attitudes and behaviors in the form of degree of agreement or disagreement with a statement. The self-assessment was thought to capture the following skills: communication skills, teamwork skills, work ethic (being organized and focused), empathy, aspirations, ability to deal with criticism, and grit. The skills chosen to be assessed via the self-assessment do not follow a particular trait hypothesis, such as the Big Five, but were based on employers’ needs as identified through employer surveys.

Chapter 2: Bangladesh Enterprise-Based Skills Survey

Chapter 2 is based on the 2012 Bangladesh Enterprise-Based Skills Survey (ESS), a matched employer-employee survey. The survey covers formal sector firms in the industrial and manufacturing sectors, sampling 500 firms and 6,981 individuals, stratified by economic sector and firm size. Despite its limitation to five sectors (manufacturing, commerce, finance, education, and public administration), the survey is quite representative of the formal sector in Bangladesh (Nomura et al., 2013). Matched employer-employee data sets are particularly useful in considering returns to cognitive and non-cognitive skills, as they allows us to take into account firm-specific characteristics such as differences in wage-setting or sorting of workers with a particular skill set into firms that pay different wages.

¹The program was not rolled out in a randomized way, nor does any possible ‘control’ data set exist, making a formal evaluation unfeasible.

The survey consists of two modules, one each for employees and employers. The employee portion of the survey contains detailed information of each individual’s background, educational attainment, and numeracy and literacy skills, as well as their personality traits. Measures for numeracy and literacy assess primary school knowledge. These tests serve as our proxy for cognitive ability, taking into account the shortcomings of such a measure highlighted earlier. The measures of non-cognitive skills are based on the short Big Five Inventory (BFI-S) as well as questions about grit and hostile attribution bias. In our survey, each Big Five dimension (openness to experience, conscientiousness, agreeableness, extraversion, and emotional stability or neuroticism) is based on three questions, as is grit, while hostile attribution bias is based on only two. The questionnaire was phrased in terms of short questions (e.g. “Are you outgoing and sociable? For example, do you make friends very easily?”), departing from the original Big Five questionnaire, which consists of statements. However, at the time of questionnaire development, questions were deemed to be more appropriate for the population surveyed.

The survey further elicits responses from business owners and high-level managers for the employer module dealing with the importance of certain selection criteria in hiring potential employees (such as academic performance, skills, or affiliation with an informal network) and the importance employers place on types of skills in their workforce (such as problem-solving, motivation, or an ability to work as part of a team).

Chapter 3: India NEEMSIS data

Chapter 3 is based on a novel data set from rural Tamil Nadu, India, entitled Networks, Employment, Debt, Mobilities, and Skills in India Survey (NEEMSIS). The survey data were collected over two periods: first from August 2016 to early November 2016 and then from January to March 2017 in 10 villages in the Cuddalore and Villupuram districts of Tamil Nadu. The survey uses a stratified sampling framework according to, first, agro-ecological considerations (dry/irrigated agriculture in villages), then urban proximity, and lastly social groups (caste representation).

The NEEMSIS consists of comprehensive household and individual level modules, completed by both the household head, and a randomly chosen younger member of the household (older than 18 and younger than 35). The total sample size of the individual survey is 952 individuals. This individual-level survey provides information on labor force participation, labor outcomes, and social networks, alongside a cognitive and a non-cognitive skills assessment. The cognitive skills assessment includes Raven’s Colored Matrices, a cognitive, visual, non-verbal test that captures concepts of fluid intelligence. The survey further includes a simple literacy test (four questions) and a numeracy test (four questions). The non-cognitive skills assessment consists of a Big Five questionnaire and questions

about grit. The questionnaire was phrased in the form of questions, similar to those used in the 2012 Bangladesh ESS; the length of the questionnaire was extended to include seven questions per Big Five dimension and six questions for grit, however. The extension of the questionnaire reflects the rather low internal validity of the Big Five dimensions in the ESS. The language in the question set was adjusted to accommodate a low-literacy population, and a careful translation to local Tamil was developed after numerous discussions and tests among the survey team, which included local enumerators.

Outline of the thesis

This thesis consists of three chapters that can be read independently of each other but also form a coherent piece of work.

Chapter 1 provides a measurement perspective. This paper contrasts the traditional self-assessment with another form of measurement: observational exercises. Using panel data from two cohorts of a skills training program, I find that the self-assessment and the observational exercises measure different concepts: the former captures underlying personality traits, while the latter relates more closely to personality states, which are malleable as a result of the intervention. The paper thus highlights the importance of knowing exactly what is measured when assessing program impact.

Chapter 2 uses a novel matched employer-employee data set representing the formal sector in Bangladesh to provide descriptive evidence of both the relative importance of cognitive and non-cognitive skills in this part of the labor market and the interplay between skills and hiring channels in determining wages. While cognitive skills (literacy, a learning outcome) affect wages only by enabling workers to use formal hiring channels, they have no additional wage return. Non-cognitive skills, on the other hand, do not affect hiring channels, but they do enjoy a positive wage return. This wage return differs by hiring channel: those hired through formal channels benefit from higher returns to openness to experience but lower returns to conscientiousness and hostile attribution bias. Those hired through networks enjoy higher wages for higher levels of emotional stability, but they are also punished for higher hostile attribution bias. This is in line with different occupational levels being hired predominantly through one channel or the other. We provide suggestive evidence that employers might use hiring channels differently, depending on what skill they deem important; employers valuing communication skills, a skill potentially observable during selection interviews, are associated with a larger within-firm wage gap between formal and network hires, while the importance of teamwork, a skill that is more difficult to observe at the hiring stage, is associated with a smaller wage gap.

Chapter 3 uses the demonetization policy in India, an unexpected and unfore-

seeable exogenous variation that had direct effects on network usage but not on interpersonal trust, to causally identify the effect of social networks in determining trust. It is thereby able to disentangle trust and participation in social networks, concepts which are inherently interrelated. We use first-hand quantitative and qualitative data from rural South India and control for a variety of individual characteristics that could influence network formation and trust, such as personality traits and cognitive ability. We find that social interactions only had a significant effect on levels of trust among men. Further, we find important differences along the lines of caste membership. Among lower castes, who live in homogeneous neighborhoods and relied on their neighbors and employers to cope with the shock, making use of one's network more intensely increases levels of trust placed in neighbors. Among middle castes, who live in more heterogeneous neighborhoods and relied predominantly on other caste members to cope, a larger network size leads to higher levels of trust placed in kin among employees but lower levels of trust in neighbors (who tend to be more dissimilar). This paper thus shows that social interactions can foster trust, though this is dependent on the type of interaction occurring. The paper also demonstrates the importance of having clearly defined in- and out-groups in trust measures, given the highly segregated nature of social interactions in rural South India.

First and foremost, the thesis as a whole highlights the importance of measurement and obtaining reliable constructs before engaging in any deeper analysis. For any empirical work, the construction of valid constructs is crucial to be able to draw valid conclusions. This is particularly true for dimensions such as personality traits, which, most commonly, rely on self-assessments that are prone to a number of response biases, as outlined in the first chapter. The response bias corrections described in the first chapter then prove useful in the remaining chapters, in which correcting for one type of response bias, acquiescence, vastly improves the internal validity of dimensions derived from self-assessments. In all chapters, response bias is stronger for the less educated, making this a particularly important point in developing countries, which often have rather low levels of educational attainment.

Second, the thesis brings into focus the importance of taking into account country characteristics when looking at the importance of skills in a developing country context. The second chapter takes into account the hiring channel when estimating returns to cognitive and non-cognitive skills and finds that these returns do indeed differ by hiring channel, providing suggestive evidence that employers might use channels differently depending on what type of skill they value. Ignoring selection into hiring channels and estimating returns on the overall sample would have led us to the conclusion that non-cognitive skills carry no wage return in a developing country (Bangladesh), when in fact different non-cognitive skills do carry wage returns in different hiring channels. The third chapter highlights the importance of taking into account the structure of the local society—in our case,

caste membership. While demonetization affected all individuals equally in theory (by simply reducing the overall money supply), the effect of the policy and an individual's type of interactions as a result was strongly dependent on gender and caste membership in practice. As different caste groups interacted with different peers to cope with the crisis, this then led to differential effects on trust measures. Again, this is particularly important in developing countries, which often feature more stringent gender roles and social hierarchies that must be taken into account.

Taken together, the chapters of this thesis seek to provide evidence of the roles that cognitive and non-cognitive skills can play in developing countries, a context broadly ignored by the literature but in which they might be most relevant.

Appendix for the Introduction

Table A0.1 – Description of the Big Five personality model

Big Five Personality dimension	American Psychology Association Dictionary Description	Facets (and correlated skill adjective)
Openness to Experience	The tendency to be open to new aesthetic, cultural, or intellectual experiences	Fantasy (imaginative), Aesthetic (artistic), Feelings (excitable), Actions (wide interests), Ideas (curious), and Values (unconventional)
Conscientiousness	The tendency to be organized, responsible, and hardworking	Competence (efficient), Order (organized), Dutifulness (not careless), Achievement striving (ambitious), Self-discipline (not lazy), and Deliberation (not impulsive)
Extraversion	An orientation of one's interests and energies toward the outer world of people and things rather than the inner world of subjective experience; characterized by positive affect and sociability	Warmth (friendly), Gregariousness (sociable), Assertiveness (self-confident), Activity (energetic), Excitement seeking (adventurous), and Positive emotions (enthusiastic)
Agreeableness	The tendency to act in a cooperative, unselfish manner	Trust (forgiving), Straight-forwardness (not demanding), Altruism (warm), Compliance (not stubborn), Modesty (not show-off), and Tender-mindedness (sympathetic)
Neuroticism	A chronic level of emotional instability and proneness to psychological distress	Anxiety (worrying), Hostility (irritable), Depression (not contented), Self-consciousness (shy), Impulsiveness (moody), Vulnerability to stress (not self-confident)

Notes: This table has been adapted from John and Srivastava (1999) and is also shown in Kautz et al. (2014).

Chapter 1

Using self-assessments and observations to capture non-cognitive skills: Insights from a skills training program in Mozambique

Abstract

The importance of non-cognitive skills for educational and labor market outcomes has gained a great deal of attention in the policy sphere recently. Most of the available measures of non-cognitive skills, in both developed and developing countries, rely on self-assessments. However, it is unclear whether these are valid measures, especially among young and vulnerable populations. This paper contrasts the traditional self-assessment with another form of measurement: observational exercises. Using panel data from two cohorts of a skills training program, I find that the self-assessment and the observational exercises measure different concepts: the former captures underlying personality traits; the latter relates more closely to personality states, which are malleable as a result of the intervention. The paper thus highlights the importance of knowing exactly what is measured when assessing program impact.

1.1 Introduction

Skills have been a prominent component of the public debate surrounding education and labor in recent years as jobs evolve and work becomes ever more complex. Non-cognitive skills in particular have captured the attention of researchers and policy makers, due to both their positive effects on educational attainment, labor market success, health, and criminality on the one hand, and their malleability over the life cycle on the other. In a developing country context, non-cognitive skills have been found to have greater predictive power than cognitive skills in the school-to-work transition in rural China (Glewwe et al., 2017).

In developed countries, the evidence of the positive effect of non-cognitive skills¹ is more developed. Almlund et al. (2011) and Kautz et al. (2014) provide a good overview for the predictive power of non-cognitive skills, finding evidence that non-cognitive skills have been related to years of schooling with similar predictive power as cognitive skills (Almlund et al., 2011), to job performance and wages (Nyhus and Pons, 2005), to occupational attainment (Cobb-Clark and Tan, 2011), and to educational attainment and grade point average (Duckworth et al., 2007). The returns to social skills have further been increasing, at least in the US labor market (Deming, 2017). The importance of non-cognitive skills has also been stressed by evaluation of programs, such as the General Educational Development (GED) program in the US, a second-chance program for students who dropped out of high school. It consists of passing a series of cognitive tests, resulting in credentials that are generally considered equivalent to a high school degree. Despite similar cognitive skills between normal high school graduates and those who completed the GED program, however, the latter have been shown to have worse life outcomes in terms of employment (shorter spells of employment), health (worse health), and social outcomes (higher rates of divorce and higher probability of incarceration), illustrating the role that non-cognitive skills can play (Heckman and Rubinstein, 2001).

Despite the evidence that non-cognitive skills have an important influence on life outcomes, much less is known with regard to how to build these skills, especially in developing countries. In general, unlike cognitive skills, non-cognitive skills are thought to be malleable throughout adolescence and into young adulthood (Cunha and Heckman, 2008; Kautz et al., 2014), though they are understood to be rather stable during adulthood, apart from the effect of major life events (Elkins et al.,

¹Non-cognitive skills in this paper refer to traits that are not captured by standard intelligence or achievement tests. These traits are also oftentimes called soft skills, 21st century skills, or socio-emotional skills. Distinguishing this strictly between cognitive and non-cognitive skills is of course a simplification, as first, non-cognitive skills are also cognitive in the sense that information processing underlies many personality traits (Bandura, 1999; Mischel and Shoda, 1999) and second, ample evidence shows that cognitive and non-cognitive skills interact in producing ‘cognitive’ outcomes, such as school performance or performance on intelligence tests (Borghans et al., 2008b; Duckworth et al., 2011; Duckworth and Seligman, 2005).

2017).²

Programs explicitly aimed at fostering non-cognitive skills have shown mixed outcomes so far. The evaluation of a youth training program in the Dominican Republic shows a positive impact from the skills training program on formality and earnings, though the program did not affect rates of employment (Card et al., 2011; Ibararán et al., 2015). Preliminary evidence of an in-school non-cognitive skills training program in France shows a positive effect of the program on skills, but the result is not robust to the measurement method (Algan et al., 2016).

The caveat surrounding the latter finding above is important and forms the basis for this paper. Indeed, the way the non-cognitive measure is constructed matters greatly for the conclusions reached about the impact of these skills on a variety of outcomes (Humphries and Kosse, 2017). Most notably, three main types of measurement are typically used to assess non-cognitive skills: self-assessment questionnaires, sometimes using the terminology of the Big Five; inference based on observable behaviors, such as delinquent behavior; and evaluations by teachers or peers.

The first method of assessment stems from psychology and the so-called lexical hypothesis: traits that are important in people’s lives tend to be captured in language (Golsteyn and Schildberg-Hörisch, 2017). The Big Five personality test, usually attributed to Allport and Odbert (1936) is an example of one of the most widely accepted self-assessment instruments and has been replicated across cultures (John and Srivastava, 1999) and developmental stages of the life course (Soto et al., 2008). Self-assessments are inherently subject to bias, which refers to measuring the true skills with error. One cause of such bias is the reliance of self-assessments on a “correct” assessment of oneself, making the information relatively subjective (McConaughy and Ritter, 1995). Further, information captured by a self-assessment is usually retrospective (Shapiro and Kratochwill, 2000) and specific to the reference group implied when assessing oneself. West et al. (2016) provide an example of the latter. The authors look at a large set of non-cognitive skills among eighth graders and find that children who attended charter schools had generally better results in terms of educational attainment and attendance but simultaneously rated themselves as worse than children attending regular public schools. West et al. (2016) attribute this to reference bias: children in charter schools compared themselves to other high-achieving children in charter schools and therefore rated themselves more critically. Anchoring vignettes have been used to overcome the problem of different reference groups in non-cognitive skills measurement, for example in Brazil (Primi et al., 2016).³ In addition, self-assessments can be prone

²Non-cognitive skills have been used as a rather broad term, being used interchangeably for personality traits, thought to be stable from young adulthood onwards, and other ‘soft’ skills, which are more malleable. Throughout this paper, non-cognitive skills can mean either of the two concepts.

³Anchoring vignettes have been used in many different settings that are prone to subjectivity (Hopkins and King, 2010), such as assessments of health (Salomon et al., 2004) or well-being (Ravallion, 2012).

to response bias factors, such as faking, yay-saying, or social desirability bias (Merrell, 2003).

The second method of assessment, observable behaviors, mostly looks at (usually adverse) behaviors, such as delinquent behavior, teenage pregnancy, or smoking among teenagers and then uses these observable behaviors as a proxy for the level of non-cognitive skills, as negative behaviors and skills are thought (and have been shown) to be negatively correlated (Elkins et al., 2006). This method thus implicitly makes the assumptions that negative behaviors are due to a lack of certain non-cognitive skills. This way of measuring non-cognitive skills might not be particularly helpful when attempting to measure the effect of skills training programs, however, as they only focus on a small subset of (rather strong) behaviors, which presumably take a while to change. Focusing on behaviors would thus disregard any small change that resulted from a program.

Lastly, teacher and parent evaluations, such as rating scales, have been used extensively in the educational literature (Crowe et al., 2011; Humphrey et al., 2011). Ideally, these rely on multiple observations over an extended period of time and are thought to be an efficient tool for observing other people's behavior, but observations can be expensive in terms of time needed; Doll and Elliott (1994) use classroom observations during free play periods and find that at least five observations of thirty minutes each across several weeks are needed for an accurate assessment. Further, in a meta-analysis of 199 studies, Achenbach et al. (1987) show that agreement among raters about the score is higher among similar raters (i.e. a pair of teachers) than among dissimilar raters, illustrating the role that observer bias can play. Indeed, observational ratings are still subject to bias, this time driven by the observers and the behavioral priors they have towards other people. For example, Uher and Asendorpf (2008) show that human raters asked to judge the behavior of crab-eating macaques rated younger monkeys as more curious and impulsive than older ones and females as cleaner than males, even though when purely looking at executed behaviors, no actual differences were visible. Applied to the case at hand, this phenomenon might lead, for example, to raters rating similar performances among boys and girls higher among boys due to underlying stereotypes. Observational exercises scores would then be different from their true underlying scores.

This paper contributes to the non-cognitive skills measurement literature by explicitly contrasting two types of measurement of non-cognitive skills among the same population and three rounds of surveys in a developing country. Given the importance of non-cognitive skills for socio-economic outcomes, ever more programs aim to enhance or augment non-cognitive skills, and non-cognitive instruments are more commonly included in standard individual level surveys, typically as short self-assessments. It remains unclear, however, what is actually measured by these short self-assessments and to what extent they can inform program evaluations. It

further remains unclear to what extent these skills can be measured reliably in a developing country context given educational and cultural backgrounds distinct from those of developed nations.

We use a unique data set collected as part of the monitoring and evaluation system of an intervention implemented in urban Mozambique (Maputo and Beira) aimed at improving the employability skills of young people from disadvantaged backgrounds by providing capacity-building in non-cognitive skills and vocational training.⁴ The data collected include a non-cognitive skills self-assessment and a non-traditional assessment tool, observational exercises (applied to individuals and in groups), both for the same individuals. Focusing on a vulnerable, low-literacy population is crucial, as this is arguably the population that can benefit most from improvements in non-cognitive skills. At the same time, self-assessments, the most common and cheapest way of assessing skills, could be more difficult for this population, which might not be used to practicing self-reflection. Indeed, the only other paper explicitly addressing the reliability of non-cognitive skills measurement through self-assessments in a low-literacy developing country context finds that the reliability of non-cognitive skills measures is rather low for a sample of farmers in rural Kenya (Laajaj and Macours, 2017). We contribute to the literature by not only providing more evidence on the reliability of skills measurement in developing countries, but by doing so for a different population (young adults in urban areas).

Further, it remains unclear which is the appropriate instrument for capturing changes in non-cognitive skills as a result of a skills training program. Algan et al. (2016) contrast different measures of non-cognitive skills among the same French high school population. They show that pupils did not perceive themselves any differently prior to and after the intervention, while teachers observed a substantial improvement among pupils. Beyond our exploration of an urban cohort, we contribute to the literature by providing another set of comparisons between self-assessment and observations. However, our observations do not rely on a teacher’s assessment, which might be subject to more inherent bias, but on observed behavior during standardized tasks. This paper has thus a clear measurement focus, contrasting different types of non-cognitive skills measurement for the same Mozambican young adult. It provides evidence as to what extent self-assessments, relying on self-perception, correspond to other (comparatively objective) forms of assessments that rely on demonstrated behaviors in the context of a developing country.

Our results are as follows. First, looking at the reliability of measures, we find that the self-assessments are subject to response bias. Further, not all dimensions as conceived theoretically are confirmed by the structure of the data. The observational exercises are internally valid, which is partly by design, and inter-rater reliability

⁴Unfortunately, the program was not rolled out in a way that would allow us to provide causal estimates. In particular, randomization was not feasible.

is high. Second, both measurement types sought to capture communication, teamwork, and personal motivation. Comparing the different measures for the conceptually same construct shows very low correlations during all three time periods. Observational exercise scores improved over the duration of the program, especially for those who attended more training sessions. Self-assessment scores, in comparison, remained rather constant. This provides evidence that the two measures capture fundamentally different concepts. The observational exercises seem to capture personality states which are more temporary behaviors or, within the concept of non-cognitive skills, the “skills” part, (i.e. part that is actually malleable). This is distinct from the concept of personality traits, which are relatively stable over the life course and, in our case, seem to be more closely captured by the self-assessment. Choosing the type of measurement thus has important implications for our ability to assess the impact of programs that aim to enhance non-cognitive skills.

The remainder of this paper is organized as follows: section 1.2 describes the Mozambican context and the *MUVA* Atitude program; section 1.3 presents the data and descriptive statistics; section 1.4 describes the methodology for the aggregation of non-cognitive skills dimensions; section 1.5 applies these concepts to our data; section 1.6 compares the constructs based on the self-assessment and the observational exercises; and section 1.7 concludes.

1.2 Context

1.2.1 Mozambique

Mozambique has so far been unable to translate recent economic growth into widespread poverty reduction and employment. The majority of economic growth has been driven by large, capital-intensive projects, which have generally benefited only a few people in urban areas (World Bank, 2017). In rural areas, subsistence farming is widespread, while in urban areas, urban poverty and low-productivity informal jobs are prevalent. Jobs have the potential to turn this one-sided growth into a broader, more participatory development path.

One particular challenge in getting people into employment seems to be a lack of skills. In a 2016 survey among the urban private sector, enterprises identified a lack of skills, most notably a lack of socio-emotional skills, as a challenge to employability, especially of youth. The skills most stressed by employers were responsibility, respect, honesty, motivation, and punctuality (AVSI, 2016). This is in line with global employer surveys, which show that employers perceive the greatest skills gaps to be in non-cognitive skills (Cunningham and Villaseñor, 2016).

1.2.2 *MUVA* Attitude program

MUVA Attitude is part of the wider *MUVA* framework focusing on female economic empowerment in Mozambique through education, work, and targeting social and economic barriers that prevent adolescent girls and women from succeeding⁵. It stems from the understanding that even where technical skills training programs exist in the country, women oftentimes lack the self-confidence to complete this training and find a job. Skills training that addresses more than the technical side of skills needed for a particular job can thus be especially effective in getting women into employment (Acevedo et al., 2017). Further, employers to a larger extent stress the importance and lack of non-cognitive skills as a decisive factor for the lack of youth employment (AVSI, 2016). The program thus seeks explicitly to address these internal constraints to employment by focusing on fostering non-cognitive skills.

The program is currently implemented in the urban areas of Maputo and Beira. Young women (and men) from disadvantaged backgrounds are invited to participate in a program that identifies their natural strengths and interests before nurturing them through practical exercises and instructions, combining elements of technical skills training with training of social and emotional competences (non-cognitive skills). *MUVA* Attitude consists of two complementary components: 1) two months of intensive non-cognitive skills training (four times per week), followed by 2) six months of technical skills training combined with a light-touch non-cognitive skills training (once a week). The technical skills training is composed of technical and vocational education and training courses (TVET) as well as an internship. The novel elements of *MUVA* Attitude are the strong emphasis put on soft skills training and methods and surveys developed at different points in time to monitor the development of skills.

1.3 Data and descriptive statistics

1.3.1 Data

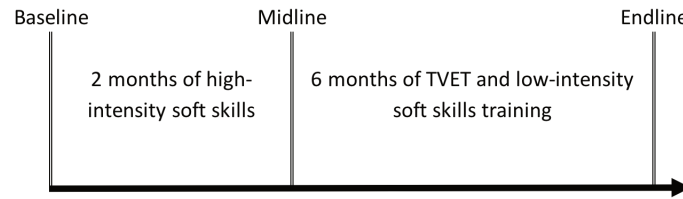
Data was collected from participants at three points: baseline data before the start of the intensive non-cognitive skills training, midline data after the two months intensive non-cognitive skills training, and an endline survey after six months of technical and ‘light touch’ non-cognitive skills training (once per week). Figure 1.1 illustrates the sequence of data collection.⁶

As of March 2018, the full sequence of data collection is available for two cohorts (cycles 1 and 2). Cycle 1 started the program in December 2016 and finished in August 2017; cycle 2 started in April 2017 and finished in December 2017. The

⁵The *MUVA* program is implemented by Oxford Policy Management (OPM) and funded by DFID.

⁶Data collected so far does not allow us to look into potential long-term effects, especially with regards to labor market outcomes.

Figure 1.1 – Sequence of *MUVA* Attitude program and data collection



program was similar in content in each cycle, though selection into the program differed slightly. In both cycles, selection was based on a geographical filter, a gender ratio, an age-filter (between 15 and 30 years old), being in neither full-time education or employment, and having a poverty score below the cutoff.⁷

In the second cycle, those interested were also required to pass a very simple literacy test to be eligible to participate. During the implementation period of the first cycle, it became clear that despite having completed at least the 7th grade, many participants were not able to read and were therefore unable to fully follow the technical training offered. Despite this difference in selection into the program, in practice and as will be described in Section 1.3.2, once we focus on those who are present in all rounds of the survey, differences in terms of observable characteristics both at the individual level and household level are small. Therefore, this paper combines cycles.

The first and second cycle of the program started with 309 participants and with 331 participants, respectively. In both cases, significant shares of the initial sample did not complete the program or surveys. Table 1.1 depicts the sample composition over both cycles and periods of data collection. While the large amount of people dropping out could certainly introduce problems of selection bias, this should not matter for the purpose of this paper, as we do not aim to assess the impact of the program on improving skills, but instead focus on contrasting different measurements of these skills.⁸

⁷The poverty score takes into account socio-economic circumstances of the young people's households. It was collected for all those interested in taking part in the program. However, the detailed individual information, including data on skills, was only collected for those who were eventually selected to take part (i.e. only those at the poverty threshold and below). This was mainly done for cost purposes but prohibits us from using the poverty cutoff to establish causal estimates of the effect of the program on skill improvements.

⁸Attrition could still be an issue if, for example, those who leave the program are from the higher end of the ability distribution. If that was the case and if we then did not find any correspondence between the different instruments, this could simply reflect the inadequate nature of the instruments for the lowest educated. The instruments were designed for a population with rather low levels of education, but require some literacy and text understanding skills. However, those who leave the program score on average worse at baseline on all measures of non-cognitive skills than those who stay. Those who leave are also less educated than those who stay. This raises confidence that the instruments are appropriate for the population surveyed.

Table 1.1 – Sample composition

	Cycle 1	Cycle 2
Baseline only	93	57
Two waves*	54	82
Panel with non-missing skills	162	192
Original sample at baseline	309	331

Notes: *Also includes 5 individuals from cycle 1 and 12 individuals from cycle 2 who, despite being present in all three waves, have missing values for some (or all) of the observational exercises.

1.3.2 Descriptive statistics of the panel sample

The following section briefly provides descriptive statistics for those young adults of both cycles who are present in all three waves of the surveys. This constitutes the sample of interest for our analysis. Table A1.1 presents descriptive statistics for the panel sample for both cycles. A little over 60 percent of participants are women, who are on average almost 24 years old at the beginning of cycle 1 and about 22 years old at the beginning of cycle 2. About a third of the sample are married and about half already have children, reflecting the high rates of early pregnancy in Mozambique. Participants in cycle 1 have on average more children, which goes hand in hand with their higher average age. Participants have on average obtained a little more than 9 years of education. About 20-30 percent have been economically active (this could be in any type of work) prior to completing the survey, and about 20 percent have also gathered previous experience in an internship or a training course.

In terms of household characteristics, according to the constructed poverty score, households in the first cycle are more likely to be poor than those in the second.⁹ Large households are very common (the average household size is 7.2 members in cycle 1, 6.5 in cycle 2). Finally, Table A1.1 reports on the *bairros* (neighborhoods) within Maputo, the capital, and Beira, the third largest city in Mozambique. In both cycles, about one third of participants come from Beira, while the rest live in the capital.

Indeed, this description illustrates that the composition of the two cycles is very similar among those present in the panel, which justifies combining them for the remainder of the analysis. This leaves us with a total sample size of 354 individuals present across all panel waves.

1.4 Methodology

Two different measures of non-cognitive skills were captured during all three rounds (baseline, midline, endline) of the survey: observational exercises and

⁹The poverty score captures a household's likelihood of being below a certain poverty line, taking into account characteristics of the dwelling as well as asset ownership. A lower score means that the household has a higher probability of being poor.

self-assessments. The skills assessed via either instrument are not fully identical, but they are partially overlapping. The following section describes each instrument in more detail.

1.4.1 Observational exercises

Observational exercises are the first type of measurement of non-cognitive skills. These are task-based tests which observe respondents ‘in action’ during group and individual exercises. In both cases, participants were judged by a group of (ideally) three people: a facilitator from a community-based organization in each *bairro*, a person from the *MUVA* Atitude project team, and a member of the *MUVA* MEL team.¹⁰ In practice, the number of raters varied. While three raters were present in most cases, a few had either two or four people. Raters had a list of definitions of what the ‘best’ or ‘worst’ performance of each skill looked like. Table A1.2 displays the assessment criteria for a good performance within each non-cognitive skill dimension. Two types of observational exercise were used: a group-based exercise and an individual exercise.

Group exercise

The group exercise consisted of a short task that participants were asked to complete in groups of six or seven participants, all randomly allocated. Despite this random allocation, though, it is still likely that participants knew at least some of their fellow group members.¹¹ The groups for the group exercise were randomly re-drawn at each round of data collection to ensure that participants met new group members at each turn.

The groups were allocated about 45 minutes to complete a given task. The content of exercises differed between the different periods of data collection to ensure that differences in scores did not stem from participants having familiarized themselves with the exercise. At the same time the exercises were similar enough to ensure comparability. As an example, one group exercise asked participants to jointly build an animal from recycled material. They had to decide on a name, the characteristics, and the appearance of the animal, and present it to the other groups. Raters then observed the groups without interacting with them, observing group members’ participation individually and assessed the individual members’ performance on the following skills: listening, negotiation, motivation to work with others, creativity, flexibility, and personal motivation. Each rater was then asked

¹⁰Examiners participated in training prior to the data collection and judged one pilot group exercise during the training. Before each wave of data collection a refresher training was given to the raters. MEL stands for Monitoring, Evaluation, and Learning.

¹¹The *MUVA* Atitude program was implemented at the *bairro* level. In each *bairro*, the course is implemented twice per day, once in the morning and once in the afternoon. For the observational group exercise, the morning and afternoon classes were combined, and the smaller groups of six to seven members were randomly drawn from the combined morning/afternoon sample. Thus, it is quite likely that participants knew some but not necessarily all members of their group exercise group.

to give each student a grade for the display of each skill, ranging from 1 to 5, with 1 the lowest. Raters would then discuss their scores with each other and jointly decide on a composite score. For a good rating on the skill ‘listening’, for example, the participant would need to show that she does not interrupt her group members, but instead listens to their arguments and tries to take their opinions into account. This could also include asking questions to other group members to understand their arguments better. Table A1.2 displays the criteria for a good performance.

The raters were not directly involved in the exercise and did not interact with participants. The group exercise thus simply relies on observing participants, who oftentimes were not even aware after a certain period of time that they were being observed. Direct observation is considered the gold standard for non-cognitive skills assessments in the educational literature (Elliott et al., 2015). Unlike the group exercise the gold standard of observation does not rely on specific tasks, but instead observes students behavior unprompted, such as during free play periods (Doll and Elliott, 1994). The literature recommends repeated observations over a period of time. While the group exercises as used in this setting are thus not exactly equivalent to the ‘gold standard’ used in the literature, they do come close.

Individual exercise

The individual exercise consisted of an individual presentation in front of the raters. Participants were given preparation time before delivering a three-minute presentation about themselves, their experience in *MUVA* Atitude, and their professional goals with only the examiners present. Raters then judged the participants in terms of display of body language, logical argumentation, professional attitude, and speaking with confidence. For example, a good performance for the dimension ‘body language’ means that the participant has a confident posture and a calm voice when speaking and makes eye contact with the raters. Raters again rated participants individually on all four dimensions and then jointly decided on a composite score. The exact prompt for the individual presentation changed between each round of data collection but remained broadly comparable. The individual presentation and the rating criteria aim to mimic professional situations in which students could find themselves while employed. Table A1.2 displays the criteria for a good performance.

Aggregating the observational exercises

The skills as assessed by both types of observational exercises form sub-scales of larger dimensions, which we aggregate for comparison with the self-assessment (described in the next section). For aggregation, we first aggregate the ten skills into three larger dimensions motivated by our theoretical understanding of how these skills should map by taking the simple average across items within a dimension. The three dimensions that we look at are communication skills (made up of body

language, logical argumentation, professional attitude, and speaking confidence), teamwork (listening, negotiation, and motivation of others), and personal motivation (creativity, flexibility, and personal motivation). We then use exploratory factor analysis to see whether the data supports this distinction.

Inter-rater reliability

To assess the internal validity of the observational exercises, we look at the inter-rater reliability of scores. At each point in time, each participant was assessed by (ideally) three different raters, who would each assign individual scores and then jointly decide on a composite score. We first assess the joint probability of agreement, i.e. the percentage of time that raters individually assign the same score. As this could also happen purely by circumstance, we then consider Fleiss' κ , which factors out the agreement due to chance (Fleiss, 1971).¹²

1.4.2 Self-assessments

The second type of instrument, a non-cognitive skills self-assessment, represents a more traditional form of measurement. Participants were given a questionnaire and asked to answer questions about their own attitudes and behaviors in the form of degree of agreement or disagreement with a statement on a 1-5 Likert scale. In total, the self-assessment questionnaire consisted of 40 questions, which can then be aggregated into larger dimensions. The self-assessment was thought to capture the following skills: communication skills, teamwork skills, work ethic (being organized and focused), empathy, aspirations, ability to deal with criticism, and grit (the perseverance and passion for long term goals). The skills chosen to be assessed via the self-assessment do not follow a particular trait hypothesis, such as the Big Five, but were based on employers' needs identified through employer surveys. Questions were drawn from a variety of sources, including other studies conducted in developing countries and new questions designed by the team to capture the local context.

Unlike the observational exercises, self-assessments do not rely on shown behavior, but on a subjective and retrospective assessment of one's attitudes and behaviors. As such, they require individuals not only to be able to reflect on one's personality and behavior, but also to be able to understand the question asked. Self-assessment can thus be subject to a number of problems, ranging from dependence on the reference group implied to response biases, in the form of faking, yay-saying, and answering in accordance with perceived social norms and attitudes (social desirability bias).

It is therefore crucial to test the reliability and validity of the self-assessment skill measures before using them for a comparison with other forms of measurement.

¹²Fleiss' κ is an extension of Cohen's κ to more than two raters.

Indeed, the literature has shown that these biases are not only valid in theory, but also occur in practice. Laajaj and Macours (2017) investigate the measurement of non-cognitive skills in rural Kenya among a sample of farmers, who are on average 46 years old with six years of education and find that the self-assessment used is subject to measurement error in the form of acquiescence bias.¹³ Rammstedt and Farmer (2013) use a sample of German adults and show that response biases can even invalidate the correlational structure of a commonly used and validated assessment of personality traits, the Big Five inventory. They find that the response bias is stronger for those with low/medium levels of education. The authors then extend this analysis to 18 countries in Rammstedt et al. (2013), with similar findings, though the educational bias seems less pronounced in non-individualistic cultures.¹⁴ In the following, we describe the procedure to correct for one of the possible biases, acquiescence bias, following the procedure described in Rammstedt and Farmer (2013).¹⁵

Acquiescence bias

Acquiescence bias is the tendency to answer statements in an affirmatively, even if this leads to contradictory statements. For example, a respondent who agrees with the statement *In a discussion, I have the ability to express my opinions with clarity* should disagree with the statement *Sometimes, I have difficulties to find the right words for what I want to say*. If the respondent agrees with both statements, this suggests acquiescence bias is present. We follow the approach described in the literature (Laajaj and Macours, 2017; Rammstedt et al., 2013; Soto et al., 2008) and determine matched pairs that are logically opposite (such as the pair describe above). We then calculate the acquiescence score, which is the mean of all opposed pairs by participant. Given our five-level Likert scale, an acquiescence score > 3 is an indication of acquiescence bias, as it means that participants are more likely to answer affirmatively. The acquiescence score can then be used to ‘correct’ the items as described below to enhance internal validity.

¹³The sample used in Laajaj and Macours (2017) is notably older and less educated than the sample used in this paper. The sample used here is on average 23 years old across both cycles and has 9.56 years of education. Still, Laajaj and Macours (2017) provide a good benchmark for the type of response bias that we can expect.

¹⁴The countries considered are the USA, Germany, Ireland, New Zealand, France, Denmark, Switzerland, Belgium (Flanders), Latvia, Russia, Czech Republic, Israel, Japan, Philippines, Mexico, Taiwan, South Korea, and the Dominican Republic.

¹⁵Unfortunately, the structure of the data does not allow us to correct for other forms of bias, such as social desirability bias or reference bias, i.e. having different forms of reference groups in mind when answering the questionnaire. The latter could be corrected for through anchoring vignettes which impose a comparable reference group on participants. Anchoring vignettes describe a hypothetical person and then ask the respondent to rank the person in terms of skills shown. They are placed before questions examining these skills and therefore ‘anchor’ individuals’ answers.

Aggregation of the self-assessment

Having addressed acquiescence bias in the sample at hand, this section describes how to aggregate the 40 self-assessment statements into larger dimensions. We make use of two types of aggregation: one based on our theoretical understanding of how the statements should map into dimensions and a second using a data-driven approach. In both cases, we start from the uncorrected statements and then repeat the approach after correcting for acquiescence bias using the acquiescence score described above.

Aggregation based on pre-defined dimensions

One of the most commonly used methods of aggregating non-cognitive skills indices is taking simple averages across all items that are supposed to map into a pre-defined dimension. In the case at hand, seven dimensions were pre-defined (communication, teamwork, work ethic, empathy, aspiration, ability to deal with criticism, and grit). In designing questionnaires, researchers oftentimes come up with a specific set of questions that are thought to cover different elements of one larger dimension (this was also the case in the design of the self-assessment instrument underlying this paper). Simple averages are the suggested form of aggregation in, for example, the World Bank STEP surveys (Pierre et al., 2014) and are used in the cases of both German and Australian data (see Heineck and Anger (2010) and Cobb-Clark and Tan (2011), respectively). Simple averages over pre-defined dimensions have the advantage of being transparent and easy to understand. Laajaj and Macours (2017) call this approach ‘naive’, however, as simple averages do not take into account potential acquiescence bias and rely fully on a ‘correct’ mapping of items to dimensions, which is up to the judgment of the researcher aggregating the items. It thus does not take into account any patterns driven by the data.

Aggregation using a data-driven approach

A second approach is simply to ignore any pre-defined mapping and instead rely fully on underlying patterns in the data. This approach explicitly recognizes that answers to the questionnaire are only imperfect proxies of the true underlying latent traits and relies on the data to uncover those latent traits. To do so, we employ exploratory factor analysis (EFA). EFA is appropriate since we aim to discover the underlying latent variables that guide the factor structure.

The first step in aggregating the skills dimensions via EFA is to extract the factors. Then, we have to determine how many different dimensions are actually present in the data (how many factors to retain). In making this decision, we consider three of the most widely used criteria (Hayton et al., 2004):

1. *Kaiser's criterion* keeps all factors with eigenvalues ≥ 1 (Kaiser, 1958). This is the most commonly used technique (Hair et al., 1998). The eigenvalue represents the amount of variance accounted for by the factor. An average single item contributes 1 to the total eigenvalue, while a factor with an eigenvalue < 1 would account for less information than a single item. The Kaiser criterion is oftentimes criticized for retaining too many factors, but it can be appropriate if based on 20-50 items (Hair et al., 1998);
2. *Cattell's scree plot* suggests to keep all factors before the distribution of factor 'breaks' off, i.e. before a discontinuity in the plot (Cattell, 1966). The scree plots depict the eigenvalues on the y-axis in descending order against their factor numbers on the x-axis. The scree plots then rely on visual inspections to determine the point at which the slope levels off, which indicates the number of factors to retain;
3. *Horn's Parallel Analysis* (Horn, 1965) is one of the most reliable ways to determine the number of factors (Hayton et al., 2004). Horn's parallel analysis is in effect a sample-based alternative to the Kaiser criterion, taking into account that under Kaiser's criterion, some factors might have an eigenvalue ≥ 1 simply due to sampling error (Hayton et al., 2004). It is the current 'gold standard' for dimensionality assessment (Braeken and Van Assen, 2017).

The exact number of factors to retain should then rely on a combination of these techniques. Following the retention of factors, we rotate the factors to improve their meaningfulness and reliability. The purpose of the rotation is to achieve a simpler structure, rotating the reference axes of the factors to better fit the data structure. Following Attanasio et al. (2015), we use an oblique, quartimin rotation. An oblique rotation assumes that the factors can be correlated (unlike an orthogonal rotation which assumes the factors are independent) (Hair et al., 1998) which is conceptually more meaningful in our case, i.e. it is likely that the dimensions of non-cognitive skills are not fully independent. For example, an individual with good communication skills might also have good teamwork skills. The quartimin procedure ensures that the items mostly load on one factor. For aggregation, we then use the rotated factor loadings, with items assigned to the factor on which their load is highest with all other loadings set to 0 (Attanasio et al., 2015; Laajaj and Macours, 2017). The strength of loadings varies and requires assessing whether the loadings are worth considering. No real rules exist for what constitutes a meaningful loading. Ford et al. (1986) suggest that any loading > 0.4 should be considered meaningful. Hair et al. (1998) provide simulations for sample sizes needed to make various loadings significant, suggesting that at a sample size of $n = 350$, loadings of $+/- 0.3$ meet the minimum standards to be considered meaningful. Since our sample size is more or less exactly at this cutoff (with $n = 354$ per round), we follow Hair et al.'s (1998) guidelines.

Taking into account acquiescence bias

To improve the uncorrected scores, we correct items for acquiescence bias by subtracting the acquiescence score from the uncorrected items, i.e. we have an item k for an individual i :

$$kcorrected_i = k_i - acquiscore_i \quad (1.1)$$

We can then aggregate the corrected items either through simple means (the ‘naive’ method) or through exploratory factor analysis. In total, we thus have four different ways of aggregating the self-assessment raw data:

1. ‘Naive’ scores, the simple average of uncorrected items over pre-defined dimensions;
2. ‘Corrected’ scores, the simple average of acquiescence-corrected items within pre-defined dimensions;
3. ‘EFA’ scores, using EFA to determine the mapping of items to dimensions and their loadings based on uncorrected items;
4. ‘Corrected EFA scores’, using the same method as in 3. but based on acquiescence-corrected items.

Assessing internal validity and reliability

We use Cronbach’s alpha as our measure of validity and reliability of the aggregated dimensions. Cronbach’s alpha, first described in Cronbach (1951), is one of the most commonly used measures of internal consistency. It is defined as follows: Suppose we have a measure X , in our case a specific non-cognitive skill, that we measure through a number of sub-components k :

$$X = Y_i + Y_{i+1} + \dots + Y_k \quad (1.2)$$

Cronbach’s alpha is then defined as:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_{Y_i}^2}{\sigma_X^2} \right); \quad (i = 1, 2, \dots, k) \quad (1.3)$$

where σ_X^2 is the variance of the measure X , the non-cognitive skill, and $\sigma_{Y_i}^2$ is the variance of item i in the sample at hand.

Cronbach’s alpha can be used as a measure of reliability and validity. Alpha captures the lower bound of reliability of a test, because if all items were measuring exactly the same thing, alpha would purely be influenced by the degree to which each item is measured with noise. If items differ, Cronbach’s alpha also acts as a measure of validity because it captures both the extent to which different items are measuring the same (latent) construct and any measurement error involved. A low alpha can thus come about due to either the items of a construct not measuring

the same latent trait or items measuring the same trait but with high measurement error. A threshold of $\alpha \geq 0.7$ is usually deemed sufficiently high for good internal consistency of the construct captured.

1.5 Measuring non-cognitive skills – application

Having described the concepts central to aggregating the different forms of non-cognitive skill measures in the previous section, this section applies these concepts to the data at hand. We start with a description of the self-assessment, followed by the observational exercises, which require less correction in terms of response biases.

1.5.1 Correcting for acquiescence

We start the correction for acquiescence bias with a first examination of whether yay-saying is an issue in the data at hand. We matched pairs to the extent possible and came up with nine opposing pairs, one for each domain and two each for communication and empathy.¹⁶ We then calculate the mean acquiescence scores for each individual, averaging across the nine opposing pairs. Using matched pairs to calculate acquiescence response bias is useful, as participants are answering questions that are similar in content but opposite in scoring. We record answers on a Likert scale ranging from 1-5. Acquiescence bias (yay-saying) is then evident if the mean of the matched pairs is larger than the middle response option (in our case, 3); nay-saying is present if the score is below 3 (Rammstedt et al., 2013).

Table 1.2 displays the acquiescence score for both cycles combined, by period. The average value in the sample is above 3 in all three data rounds, thus providing evidence that yay-saying (acceptance acquiescence) is present in the sample. The values are all statistically significantly different from 3 and from each other. Acquiescence significantly reduces over the time frame of the program and reduces more for participants of the first cycle than for those of the second. This might suggest that the issue could be partly driven by participants not fully understanding questions in the first data collection round (baseline) and therefore tending to agree more. The score further reduces more for women, who start from a higher initial average value, than men.

Figure 1.2 shows the acquiescence score over the age range and years of schooling range of the sample. Panel a) illustrates a U-shaped relationship between the acquiescence score and age, with younger and older participants being more prone to yay-saying. Panel b) illustrates that the score decreases with additional years

¹⁶The matching is not as clean as in Soto et al. (2008), who use the Big Five Inventory, because our questionnaire uses qualifying instead of simple statements. More context-specific statements were thought to make it easier for our sample to answer the questions. However, this leads to matches such as “I prefer working alone to working in a group” with “When I work in a group I can achieve better results”. Of course, these are not perfect matches, as one could, for example, prefer to work alone, but still concede that working in a group leads to better results.

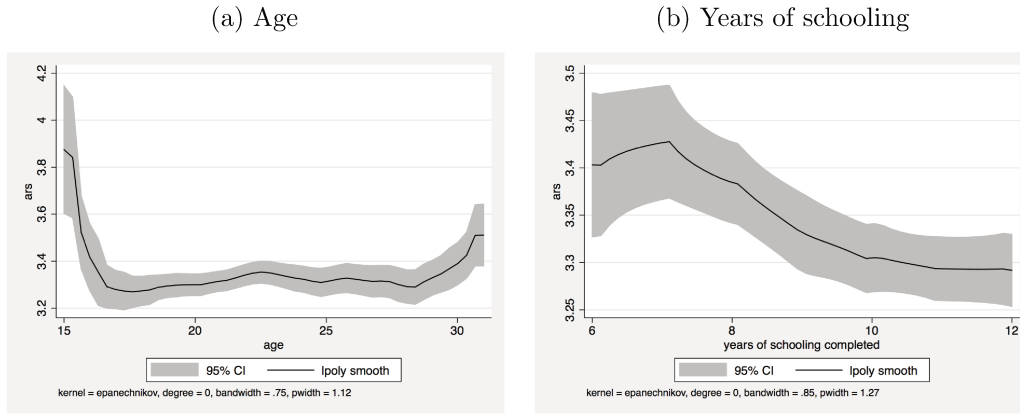
Table 1.2 – Mean acquiescence scores by cycle, period, and gender

Combined	Baseline	Midline	Endline
Acquiescence score (N=354)	3.38 (0.42)	3.32 (0.42)	3.27 (0.38)
Women (N=230)	3.45 (0.40)	3.33 (0.44)	3.28 (0.40)
Men (N=124)	3.27 (0.43)	3.30 (0.36)	3.24 (0.35)
Cycle 1			
Acquiescence score (N=162)	3.43	3.37	3.26
Women (N=101)	3.53	3.38	3.26
Men (N=61)	3.28	3.35	3.27
Cycle 2			
Acquiescence score (N=192)	3.34	3.28	3.27
Women (N=129)	3.38	3.29	3.30
Men (N=63)	3.26	3.24	3.21

Notes: The sample only contains those who are present in all three waves. Standard deviation in parentheses for combined sample.

of education, which is in line with the literature (Laajaj and Macours, 2017; Rammstedt and Farmer, 2013). Table 1.2 illustrates that acquiescence is present in our sample and that correcting for it is therefore crucial.

Figure 1.2 – Acquiescence scores over the age range and years of schooling



Notes: Sample size: 354 individuals over 3 waves of data, thus 1062 individual data points.

We further assess to what extent participants might be unwilling to agree or disagree with the statement and choose the middle response option instead. Figure A1.1 depicts the number of times that individuals chose the middle response option per 40 items. The maximum number of middle response options (out of 40) chosen is 20, which was chosen by only a single individual. In fact, only 24 people over the entire three periods answered the questionnaire with more than 10 middle response options (1062 total data points). It thus seems that participants were very willing to opt for either the disagree or agree option, making the construction of the non-cognitive skills dimension more meaningful. In the aggregation, we first recode reversely coded items and then use the acquiescence score to correct the

items (as described in equation 1.1).

1.5.2 Aggregating through EFA

As described in section 1.4, in addition to using simple averages over pre-defined dimensions for aggregation, we also use a data-driven approach, namely exploratory factor analysis (EFA), separately for uncorrected and acquiescence-corrected items. Having extracted the factor, the first step is to apply the different criteria—the Kaiser criterion, Cattell’s scree plots, and Horn’s Parallel Analysis—to determine how many factors to retain. Based on our preconceptions, we would expect about seven factors to be retained, which is the number of pre-defined dimensions that questions were mapped to in the questionnaire.

Figure A1.2 depicts the Cattell scree plots for the uncorrected items and figure A1.3 shows the scree plots for acquiescence-corrected items, displayed separately by data collection round.¹⁷ The visual inspection reveals that the number of factors to be retained is rather similar, with three factors standing out before a visual break in the plot line (apart from maybe the midline for corrected items, panel b) in Figure A1.3, where retaining 2 factors seems more appropriate). Table 1.3 summarizes the suggestions of number of factors to be retained according to the three different criteria used and for the three waves of data collection, where t_0 is baseline, t_1 is midline, and t_2 is endline. The table shows that all methods suggest a similar number of factors to be retained, which does not change much from round to round of data collection. In general, the scree test and Horn’s Parallel Analysis are considered more reliable (Ford et al., 1986). We therefore decide to retain three factors; this is significantly below the number of factors we would have expected given our seven pre-defined dimensions, suggesting that those dimensions are not reflected cleanly in the data.

Table 1.3 – Number of factors to retain according to different criteria

	Kaiser criterion	Catell’s scree plot	Horn’s PA
Naive SA t_0	3	3	4
Naive SA t_1	3	3	4
Naive SA t_2	3	3	3
Acquiescence SA t_0	2	3	3
Acquiescence SA t_1	2	2	2
Acquiescence SA t_2	3	3	3

Notes: The sample only contains those who are present in all three waves. $t_0 = \text{baseline}$, $t_1 = \text{midline}$, $t_2 = \text{endline}$; $N = 354$.

The next step to construct the non-cognitive skills dimensions through EFA is to decide which items map on which factor and whether their loadings are sufficiently high to be considered for the construction of the factor. We use an

¹⁷We separate the analysis by data collection round as the factor structure might have changed between the different rounds.

oblique, quartimin rotation so that factors can be correlated and items mostly load on one dominant factor. We follow Hair et al.’s (1998) guidelines for what constitutes a sufficiently high loading. Given that our sample size is 354 individuals per round, a loading of $+/- 0.3$ is considered appropriate. We further follow Attanasio et al. (2015) in only considering items that load clearly on one factor. This means that items have to both load around the threshold of 0.3 and load twice as strongly on that factor as on any other factor to be considered. As we have three rounds of data available, we use the loadings from the first round (baseline) and apply them to all subsequent rounds, making the dimensions comparable over time (again following the suggestions in Attanasio et al. (2015)). Table 1.4 depicts the rotated factor loadings at baseline based on the corrected and uncorrected items. Items are sorted in descending order based on the factor on which they load highest. Those marked in bold satisfy our criteria for meaningful loading and are considered for the aggregation of the three dimensions. Tables A1.3 and A1.4 depict the factor loadings at midline and endline for the corrected and uncorrected items and show that, in practice, factor loadings do not vary much between the different rounds of data collection.

Focusing on the factor loadings at baseline indicates that several of the items either do not load strongly (below 0.3) or do not load clearly on a single factor. Based on the retained items, the dominant interpretation for factor 1 is communication skills, based on 6 retained items for the uncorrected items and 13 retained items for the corrected items. Factor 2 seems to cover goals and aspirations. Factor 3 is a little less clear in its dominant interpretation, however. Questions related to it are mostly related to difficulties faced and the inability to overcome them. It is based on 5 items for the factor analysis based on uncorrected items, but only 2 items load clearly on it for the corrected items. As many of the items loading highly for the uncorrected items are reversely coded, factor 3 could thus also simply cover acquiescence response bias. We retain it for the remainder of the paper, with the caveat that its interpretation is somewhat unclear.

1.5.3 Choosing a self-assessment measure

Following the aggregation of the self-assessment measure according to pre-defined dimensions and using the data-driven approach of factor analysis, Table 1.5 displays our measure of internal validity, Cronbach’s α , for all four types of aggregated skills measures. A value of $\alpha = 0.7$ is generally considered acceptable. Looking at the uncorrected items and pre-defined dimensions first, Table 1.5 shows that several of the pre-defined dimensions do not meet that threshold if based on uncorrected items. Most notably, among seven pre-defined dimensions, only one dimension (communication) could be considered internally valid, with alphas ranging from 0.6 at baseline to 0.7 at endline. Several other dimensions, such as grit and empathy, but especially work ethic, have extremely low alphas. Work ethic only sports

Table 1.4 – EFA - Rotated factor loadings at baseline

(a) Uncorrected items				(b) Corrected items			
	F1	F2	F3		F1	F2	F3
abc_111	0.57			cr_com_139	0.67		-0.11
com_139	0.57			cr_com_109	0.65	-0.11	
com_109	0.45			cr_abc_111	0.63	0.13	-0.19
com_130	0.41			cr_com_100	0.60		
com_100	0.39			cr_em_129	0.58	0.10	0.18
com_119	0.38		0.16	cr_em_105	0.56		0.14
com_141	0.25		0.16	cr_com_130	0.48	0.18	
em_105	0.25		0.18	cr_abc_125	0.48		0.30
abc_117	0.24			cr_com_119	0.48	0.13	
asp_116	-0.11	0.52		cr_com_141	0.43	0.10	
asp_137		0.52		cr_com_124	0.42		0.13
tw_135		0.48	0.12	cr_gr_122	0.41	0.31	
tw_149	0.15	0.44		cr_gr_147	0.41	0.19	0.17
asp_106		0.43		cr_abc_117	0.40		
asp_107		0.38		cr_abc_138	0.38	0.12	0.19
em_108		0.36		cr_abc_150	0.26	0.16	-0.34
gr_133		0.36		cr_gr_112	0.26	0.16	
we_128	-0.11	0.35	0.20	cr_we_151	0.25	0.23	0.23
abc_150	0.25	0.35	-0.16	cr_em_108	0.23	0.13	
asp_144		0.34		cr_em_123	0.20		
com_132	0.18	0.32	-0.10	cr_em_118	0.19		-0.33
gr_115		0.31	0.20	cr_asp_116		0.59	
em_118	0.17	0.31	-0.24	cr_asp_137		0.55	
gr_147		0.29	0.12	cr_tw_135		0.54	
tw_146		0.25	0.18	cr_asp_106		0.48	
gr_112	0.12	0.22		cr_tw_149	0.12	0.47	-0.14
em_123		0.16	-0.12	cr_asp_107		0.46	
asp_120			0.55	cr_asp_144	0.10	0.36	
abc_125	0.12	-0.17	0.43	cr_we_128	0.21	0.35	0.23
we_151		0.13	0.41	cr_gr_133		0.34	
tw_127			0.38	cr_gr_115	0.20	0.28	
em_129	0.33		0.37	cr_tw_146	0.25	0.26	0.17
gr_122	0.31	0.10	0.32	cr_we_103		0.21	0.18
abc_138	0.14		0.29	cr_we_126	-0.21		0.44
com_140	-0.31	0.11	0.21	cr_asp_120	0.36	0.18	0.40
we_126	-0.50		0.18	cr_com_140			0.37
com_124	0.15		0.13	cr_tw_127	0.30	0.10	0.36
com_143	-0.39		-0.18	cr_com_143		-0.33	0.29
we_103				cr_com_132	0.19	0.17	-0.27

Notes: blanks represent $abs(\text{loading}) < 0.1$. F stands for factor. Items are labeled according to pre-defined categories. These are: com = communication, tw = teamwork, we = work ethic, em = empathy, gr = grit, abc_ri = ability to deal with criticism, asp = aspirations. Items in bold are items that should be retained for the construction of the factor.

an alpha of 0.18, making this an extremely unreliable dimension. Correcting for acquiescence bias improves the internal validity even among the pre-defined dimensions, though. The alphas of teamwork, aspiration, ability to deal with criticism, and grit all improve after the correction and can be considered to almost reach the threshold or, in the case of aspiration, even slightly surpass it. The weakest dimension, work ethic, improves as well, but it remains far from the threshold of 0.7.

Table 1.5 – Comparing different aggregations of the self-assessment - Cronbach’s α

Pre-defined dimensions	# items	Uncorrected items			# items	Corrected items		
		Baseline	Midline	Endline		Baseline	Midline	Endline
Communication	10	0.57	0.71	0.70	10	0.70	0.81	0.79
Teamwork	4	0.37	0.40	0.49	4	0.59	0.62	0.69
Work ethic	4	0.17	0.19	0.18	4	0.34	0.43	0.32
Empathy	5	0.18	0.40	0.45	5	0.46	0.58	0.61
Aspiration	6	0.51	0.52	0.57	6	0.64	0.67	0.71
Ability criticism	5	0.35	0.44	0.48	5	0.58	0.65	0.65
Grit	5	0.32	0.37	0.33	5	0.57	0.60	0.56
EFA main loadings	# items	Uncorrected items			# items	Corrected items		
		Baseline	Midline	Endline		Baseline	Midline	Endline
F1: Communication	6	0.61	0.74	0.72	13	0.84	0.89	0.88
F2: Aspirations, goals	11	0.64	0.67	0.68	8	0.68	0.70	0.76
F3: Difficulties	5	0.53	0.47	0.46	2	0.33	0.20	0.15

Notes: EFA is based on loadings at baseline. For the uncorrected items, 18 items did not load strongly enough or did not load clearly on only one factor. For the corrected items, 17 items did not load well either. Those items were not considered for the aggregation of the factors (see Table 1.4 for an indication of which items did not load well).

The lower part of Table 1.5 displays Cronbach’s alpha for the three factors identified through factor analysis and labeled with their dominant interpretation. For both the corrected and uncorrected items, despite being based on fewer items, the communication dimension as identified through factor analysis performs better in terms of internal consistency than the pre-defined communication dimension. The aspirations and goals dimension also depicts higher alphas, though this could in part be mechanical, as it is based on more individual items, which leads to an increase in alpha. The last dimension, called difficulties here, has low alphas, especially for the corrected items (though it is only based on two items in this case). Again, the interpretation of this dimension is not entirely clear and could also simply cover bias due to yay-saying.

Table 1.5 shows that only two internally valid non-cognitive constructs can be identified in the data: communication skills and aspirations/goals. It is also clear that correcting for acquiescence bias improves the internal consistency even among pre-defined dimensions and should therefore be applied. It is therefore suggested that the comparison strategy between the self-assessment and the observational exercise should focus on the acquiescence-corrected items and potentially place a larger focus on the dimensions communication and aspirations as those have been identified both by theory (i.e. researchers mapped those as pre-defined dimensions)

and as latent variables confirmed by the data.

1.5.4 Inter-rater reliability in observational exercises

At each point in time, each participant was supposed to be assessed by two different raters for the group exercise and by three different raters for the individual exercise, who would then agree on a joint score. In this section, we assess the extent to which raters' assessments were similar and therefore more reliable. Table A1.5 shows the number of raters per period. Most participants were indeed judged by two raters during the group exercise (about 75 percent), though a significant number was assessed by more raters, especially at baseline (about 50 percent). This is due to the fact that sometimes additional raters participated as a type of training. For the individual exercise the vast majority of participants (especially after the baseline) was indeed rated by three raters (between 65 percent at baseline and 89 percent at midline).

As few students were rated by more than three raters, we focus on those rated by two or three raters for this section. We consider the group exercise and the individual exercise separately, as they were assessed separately. Table 1.6 displays the percentage agreement as well as Fleiss' κ for the ten different skills assessed by the observational exercises, be they a group or individual exercise. In addition to raw scores, Table 1.6 also shows weighted scores, using quadratic weights. Weighted scores are generally preferred (Cohen, 1968), as otherwise all disagreement is treated equally. This means that if scores are not weighted, if rater R1 assigns a score of 1 (the worst) and rater R2 a score of 5 (the best), this disagreement is treated the same as R1 assigning a score of 4 and R2 assigning a score of 5. Using quadratic weights instead means that if raters' disagreement is close (for example, they are only one category apart), this disagreement is punished less harshly than if raters differ by several categories. Landis and Koch (1977) provide guidelines for interpreting κ , according to which a value of $\kappa < 0.2$ signifies slight agreement, $0.21 < \kappa < 0.4$ relates to fair agreement, $0.41 < \kappa < 0.6$ stands for moderate agreement, and $0.61 < \kappa < 0.8$ means substantial agreement, with anything larger being almost perfect agreement.

Looking at Table 1.6 shows that even without weights, the percentage of times that raters agree is non negligible. For the group exercise, two raters agree on average about 30 percent of the time, while percentage agreement among raters for the individual exercise is even higher at close to 40 percent. This makes intuitive sense: in the group exercise, raters were essentially asked to multi-task; they had to assess individuals' performance within the group with other group members present. Raters might have shifted their focus from one group member to the next and might therefore not have perceived the exact same behavior. The individual exercise only had one participant present at a time, making the rating task much easier. Despite the decent percentage agreements, looking at the chance-corrected

Table 1.6 – Inter-rater reliability for observational exercises

Group exercise # raters	Unweighted		Quadratic weights	
	% Agreement	Fleiss' κ	% Agreement	Fleiss' κ
<u>2 raters</u>				
Listening	0.30	0.15	0.96	0.63
Negotiation	0.30	0.19	0.96	0.69
Mot others	0.32	0.21	0.96	0.67
Creativity	0.32	0.21	0.96	0.74
Flexibility	0.31	0.21	0.96	0.72
Personal Mot	0.33	0.21	0.96	0.71
<u>3 raters</u>				
Listening	0.32	0.17	0.96	0.60
Negotiation	0.26	0.14	0.96	0.68
Mot others	0.26	0.13	0.96	0.63
Creativity	0.31	0.20	0.96	0.72
Flexibility	0.31	0.20	0.96	0.70
Personal Mot	0.29	0.17	0.96	0.72
Individual exercise # raters	Unweighted		Quadratic weights	
	% Agreement	Fleiss' κ	% Agreement	Fleiss' κ
<u>2 raters</u>				
Body language	0.37	0.24	0.98	0.72
Logical argum	0.38	0.25	0.97	0.70
Professional attitude	0.40	0.25	0.98	0.69
Speaking confidence	0.38	0.26	0.97	0.74
<u>3 raters</u>				
Body language	0.35	0.22	0.97	0.69
Logical argum	0.36	0.24	0.97	0.76
Professional attitude	0.38	0.25	0.98	0.70
Speaking confidence	0.39	0.28	0.98	0.79

Fleiss' κ 's shows a different picture, with most skill dimensions scoring in the range of only slight agreement (for group exercise dimensions) and fair agreement (for individual exercise dimensions). Table 1.6 shows the importance of using weights. Once we introduce quadratic weights (the last two columns), most dimensions now score quite highly on κ , achieving scores in the 60s and 70s, which refers to substantial agreement between raters. This illustrates that while the raters in our sample disagree, they tend to assign similar scores. The vast improvement of scores when weighted shows that raters might, for example, score an individual with $R1 = 4$ and $R2 = 5$, but are unlikely to offer scores of $R1 = 1$ and $R2 = 5$ for the same individual. The composite scores from the observational exercises used throughout this paper can thus be considered reliable as they are based on a high inter-reliability of raters.

In addition to testing inter-rater reliability, raters could exhibit bias in assigning the final, composite score per individual, which is used throughout this paper. Raters were asked to decide on a single score per dimension, per individual jointly. The simplest way of doing so would be to simply assign each individual the mean of individual raters' scores. However, raters could only provide scores round to

the 0.5 decimals. If a simple mean was different from 0.5, raters had to decide to either round the composite score up or down. This could introduce bias, if raters are more likely to round up for some people than for others. We thus looked at whether we can find predictors of a deviation from the mean, using the following simple regression:

$$Dev_i = \beta_0 + \beta_1 * male + \beta_2 * agegroup + \beta_3 * period \quad (1.4)$$

where Dev_i is defined as the deviation of the precise arithmetic mean from the composite score raters decided on for individual i . Thus, a positive value for Dev_i means that raters are more lenient, as they are rounding up. *Male* is a gender dummy taking the value of 1 for males and 0 for females, *age group* are dummies for four different age groups, and we also include period controls for the three time periods in our data. Results of a simple OLS specification are displayed in Table A1.6. While it is important not to draw too many conclusions from this simple estimation, it indicates that some bias might be present in the data. The coefficient for male is significant for some dimensions (creativity, flexibility, body language, and logical argumentation) meaning that in those cases, raters were more likely to round up. The male dummy has no significantly negative effects in our specifications. Further, compared to the youngest age group (15-19 years old), raters were more likely to round up for other age groups with regards to the professional attitude dimension. This could suggest that despite *MUVA* Attitude explicitly focusing on female empowerment, raters might still be subject to inherent gender-based biases that could influence their judgment.

1.5.5 Aggregating the observational exercises

In order to be able to compare the observational exercises to the self-assessment, we also aggregate the observational exercises into larger dimensions. Given the skills measured, three larger dimensions make intuitive sense: communication, teamwork, and personal motivation. As described previously, the observational exercises consist of two different tasks: an individual exercise (a presentation) and a group exercise. Given this method of data collection, it seems intuitive that items collected as part of the individual exercise would have a high correlation among themselves, as would items collected as part of the group exercise. This can be attributed to the fact that while raters were asked to assess individual skills, it is likely that raters are influenced by a latent dimension. For example, if raters see a good performance on one specific skill during the exercise, they might rate other skills assessed during that exercise higher as well. Also, it is likely that the different skills are correlated to form an overall good performance. Somebody with good listening skills, for example, might also be good at negotiation. It therefore seems likely that the data would only support two larger dimensions,

grouping skills assessed through the group exercise and skills assessed through the individual exercise. Indeed, exploratory factor analysis confirms this suspicion.¹⁸ To enable a comparison between the self-assessment and the observational exercises in more than two dimensions, though, we stick with the three dimensions that make intuitive sense: communication, teamwork, and personal motivation (splitting the six items from the group exercise into two dimensions, teamwork and personal motivation).¹⁹ Table 1.7 provides the Cronbach alphas for these three dimensions by data collection round. All alphas in all rounds are very high and significantly above the 0.7 threshold usually considered appropriate. The observational exercises thus seem to measure the three different skill dimensions quite appropriately and can be considered internally valid.

Table 1.7 – Cronbach’s α for the combined observational exercises

Pre-defined dimensions	# items	Baseline	Midline	Endline
Communication	4	0.89	0.90	0.88
Teamwork	3	0.93	0.92	0.93
Personal motivation	3	0.89	0.91	0.92

Notes: N = 354 per round.

1.6 Comparing measures of non-cognitive skills

This section describes the comparison strategy employed to contrast the different skills measures: observational exercises and self-assessment. We rely on the aggregated broader dimensions which have been determined best for this exercise, which are the acquiescence-corrected simple average and factor-based aggregations for the self-assessment and the simple averages for the observational exercises. We first look at correlations between the different types of measurement, then at rank changes.

1.6.1 Correlations

Table 1.8 depicts pairwise correlations between the non-cognitive skill measures. It consists of comparing three different constructs: non-cognitive skills measured by the observational exercises, summarized into communication, teamwork, and personal motivation, the self-assessment, captured by the socio-emotional constructs corrected for acquiescence bias, but keeping the originally conceived constructs (seven non-cognitive skills), and the self-assessment combined into only three skills,

¹⁸Results not reported but available upon request.

¹⁹This distinction is also in line with rater guidelines. During the group exercise, raters were asked to consider an individual’s behavior towards the group to assess listening skills, negotiation skills, and motivation of others, all of which are combined in the teamwork dimensions. Raters were then asked to assess creativity, flexibility, and personal motivation by focusing fully on the individual and disregarding their relationship to the group. These three items then make up the second dimension, personal motivation.

as evident from the exploratory factor analysis.²⁰ Table 1.8 is useful to inquire whether the dimensions constructed from the observational exercises and those constructed from the self-assessment capture similar underlying concepts, in which case correlations between them should be high. This should especially be the case for the dimensions communication and teamwork which are included in both the self-assessment and the observational exercises.

In fact, correlations between the dimensions derived from the self-assessment and the observational exercises are rather weak, but they are highly statistically significant. For the communication dimension, observational exercises and the pre-defined self-assessed communication dimension correlate at 0.24. The correlation is marginally higher if we look at the self-assessment as aggregated through factor analysis (F1:communication), but even then it only reaches 0.27. The same pattern holds true for the teamwork dimensions, with a correlation of 0.27. In fact, the correlations are highest between dimensions that we would consider similar, but even these only reach a maximum of around 0.27. In addition to examining dimensions that are thought to capture similar underlying traits, we can also look at dimensions that could capture related traits such as aspiration, grit, and personal motivation.²¹ This does not provide a clearer picture, though, as correlations remain rather low, at for example 0.22 for a correlation between aspirations and personal motivation and 0.26 for factor aspirations and personal motivation.

Table 1.8 – Correlations between self-assessment and observations

Self-assessment	Teamwork	Observational exercise	
		Personal motiv	Communication
Communication	0.21***	0.24***	0.25***
Teamwork	0.27***	0.27***	0.17***
Empathy	0.23***	0.23***	0.23***
Aspirations	0.22***	0.25***	0.21***
Ability criticism	0.18***	0.21***	0.23***
Grit	0.16***	0.18***	0.19***
F1: Communication	0.25***	0.27***	0.26***
F2: Aspiration	0.25***	0.26***	0.20***
F3: Difficulties	0.05	0.05*	0.04

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Correlations reported here are over all three time periods. Correlations by time period are similar and available upon request.

Table 1.8 suggests suggests that both types of measurement are only weakly correlated, even for dimensions which are meant to capture similar underlying traits. This could be the case for a number of reasons. First, the observational exercises and the self-assessment could be subject to different types of measurement error, which could weaken the correlation between them. This seems unlikely, however, given that the previous section has shown that the constructs are internally

²⁰Table A1.7 depicts the correlations between all skills constructs, not just between the observational exercises and the self-assessment.

²¹We do not include work ethic in any of our considerations here due to the weak properties of this dimension, as evident from Table 1.5.

valid. Further, while the correlations depicted in Table 1.8 are small, they are statistically significant. Second, it is possible that a single latent trait, such as the ability to perform in a stressful situation could underlie the observational exercise measurement. Students could have perceived the observational exercises as more of a “test” situation than the self-assessment questionnaire, as raters were present in the classroom. Still, we would then expect all observational exercise items to only load on one single factor (the latent test performance factor), which is not confirmed by the data. Further, conversations with raters suggest that the classroom atmosphere during the exercises was friendly and students seemed to become increasingly unaware that they were being observed as they became more engaged in the tasks. Third, it is possible that the observational exercises and the self-assessment simply capture different concepts. The self-assessment could be thought of as capturing underlying personality traits, while the observational exercises might instead refer to personality states. Traits are thought to represent underlying characteristics that are rather stable and difficult to change, while states refer to more temporary beliefs and behaviors (Eysenck et al., 1987). The two concepts are related, but they do not perfectly correspond to each other. This explanation seems most likely.

Having correlated the self-assessment and observational exercises dimensions with each other, we now consider how both correlate over the three rounds of data collection. Considering self-assessment dimensions derived from factors instead of pre-defined dimensions does not seem to change the interpretations; for simplicity, we only consider the pre-defined dimensions for now and focus on those that have the most support from the data, which are communication and aspirations. Table 1.9 depicts the correlations between the different skills dimensions over time (that is, the correlation of a given dimension at baseline with that same dimension at endline).

Table 1.9 – Correlations between skills over time

Corrected self-assessment	Baseline/ Midline	Midline/ Endline	Baseline/ Endline
Communication	0.45	0.49	0.58
Teamwork	0.40	0.53	0.56
Work ethic	0.42	0.39	0.40
Empathy	0.44	0.53	0.61
Aspiration	0.45	0.46	0.50
Ability criticism	0.51	0.53	0.56
Grit	0.45	0.47	0.44
Combined observational exercises			
Communication	0.47	0.38	0.51
Teamwork	0.35	0.42	0.37
Personal motiv	0.32	0.48	0.43

Notes: Sample size is 354 individuals in each round. Sa refers to self-assessment. Obs refers to observational exercises. All correlations shown are significant at the 1% level.

The stability correlations have been used to assess whether or not a person’s personality is stable over time, usually relying on larger time intervals (multiple years passing in between measurements). It shows that correlations over time are high for both the self-assessment and observational exercises, especially between the baseline and endline. Our correlations are high when compared to the literature, such as Roberts and DelVecchio (2000) who find correlations of about 0.3 for young children to 0.7 for older adults. Our sample is rather young, but the time passed between measurements is also rather short, explaining the high correlations. We can also compare our results to Laajaj and Macours (2017), who provide test-retest correlations for a sample of low-educated rural farmers. Their test-retest correlations mostly range from 0.2 to 0.3, with a test-retest period of three weeks. The fact that our correlations between time periods are much higher, even over a longer period of time, provides evidence that self-assessment can work among poor and vulnerable populations, though a minimum level of education might be required: Laajaj and Macours’ (2017) sample have on average 6 years of education, while our sample has about 9.5 years. More information on the quality of education obtained in each sample would be necessary to assess what is necessary to obtain reliable self-assessments.

Table 1.9 suggests that both types of measurement consistently capture their dimensions over time, though the overtime correlations are slightly higher for the self-assessment, providing further evidence that the self-assessment could be capturing underlying traits while the observations could capture a more changeable construct. This interpretation is also in line with the literature, which oftentimes struggles to find effects of skill training interventions on non-cognitive skills as measured by self-assessments (e.g. Ganimian et al., 2018).

1.6.2 Rank order changes

Lastly, we look at whether each type of measurement ranks individuals in a stable and similar way. Table 1.10 depicts the average ranking of skills per period, presenting intra-individual rankings (i.e. for each individual, on which skill do they perform best and on which worst). As there are seven pre-defined dimensions, a ranking of seven is considered best, and a ranking of one is considered worst. For the combined observational exercises, a ranking of three would be the best and a ranking of one worst.

Table 1.10 shows that for the self-assessment, most participants score highest on the aspiration dimension and lowest on the communication dimension at baseline. This does not change substantially over the three periods of data collection; while some dimensions improve or decline their ranking slightly, on average over the entire sample, the ranking stays exactly the same (from worst-to-best ranked dimension: communication, ability to deal with criticism, empathy, work ethic, grit, teamwork, aspiration). At baseline, about 38 percent of the sample have aspiration

Table 1.10 – Average rank of skills

Corrected self-assessment	Baseline	Midline	Endline
Communication	1.93	2.04	2.03
Teamwork	5.25	5.26	5.60
Work ethic	3.83	3.66	3.54
Empathy	3.47	3.55	3.51
Aspiration	5.84	6.00	6.09
Ability to deal with criticism	2.66	2.71	2.61
Grit	5.01	4.77	4.62
Combined observational exercises			
Communication	2.35	2.31	2.00
Teamwork	1.96	1.75	1.93
Personal motiv	1.70	1.94	2.06

Notes: Sample size is 354 individuals in each round.

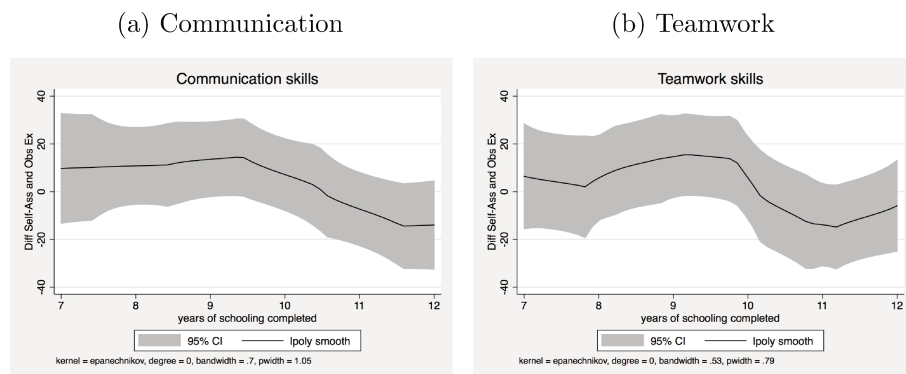
as their best ranked dimension, which increases to 46 percent at endline. For the lowest ranked dimension, at baseline, 39 percent score lowest on the dimension communication, which slightly decreases to 36 percent at endline. 35 percent of individuals keep the same dimension as their highest rank dimension between baseline and endline, and 44 percent keep their lowest rank dimension. In general, this seems to suggest that the ranking of the self-assessment remains quite stable over time over the sample population.

The observational exercises only consist of three dimensions. At baseline, communication ranks highest, followed by teamwork and personal motivation. At endline, this ranking is reversed, with personal motivation ranking highest, followed by communication and teamwork. Among the observational exercises, a rank reversal thus occurs, with participants vastly improving their personal motivation skills relative to their communication and teamwork skills. In fact, Table 1.10 shows that the ranking of the self-assessment is the same in every time period. The ranking of the observational exercise changes from than once, with the personal motivation dimension moving from last place to second place to first place. This could provide further evidence that the observational exercises measure something that is changeable over the period of time of the program, while the self-assessments seem to capture a more stable concept.

In addition to the global stability of intra-individual skills shown in Table 1.10, we also look at how individuals ranked on specific skills within the sample, i.e. inter-individual rankings. Here, we focus on communication and teamwork skills, for which measures are available in both instruments. Correlations between the within-sample rank between the different time periods is high for both communication and teamwork skills, and for both the self-assessment and the observational exercises. It is slightly stronger for the self-assessment (ranging from 0.4 - 0.54) than the observational exercises (from 0.3 - 0.49), showing again the higher stability of the self-assessment over time (as also visible in Table 1.9). More interestingly, we can

look at the differences in ranking between the observational exercises. For this, we took the difference in within-sample ranking by time period per skill (for example $rankingSA_{it_0} - rankingOBS_{it_0}$) then averaged this over all three time periods, thus computing the average distance between the within-sample ranking of the self-assessment and the observational exercise for the same skill. A score above zero thus means that within the sample, the individual ranked higher on the self-assessment score than on the score given by raters for the individual exercise. Figure 1.3 displays this difference in ranking by years of schooling completed. The left panel depicts the difference in ranking for communication skills and shows that those with lower education tend to, on average, receive a higher within sample ranking on the self-assessment than on the observational exercise; the opposite is true for those with more years of education. In both cases, a 95 percent confidence interval shows that we cannot significantly distinguish the difference in ranking from zero. Figure 1.3 seems to suggest that those with lower education might overestimate their true skill level, while those with higher education might underestimate it. This suggests that reference bias could be an issue, as individuals have different levels of “good” skills in mind when scoring themselves on the self-assessment. West et al. (2016) find a similar phenomenon among US high school students. A possible solution could be anchoring vignettes, which explicitly “anchor” the reference group individuals have in mind, as they elicit information on what individuals consider a “good” level of the skill prior to rating themselves on that same skill. Including anchoring vignettes in future surveys could further improve the quality of self-assessments, especially among diverse populations; they have in fact been implemented successfully in a developing country context (Primi et al., 2016).

Figure 1.3 – Difference between ranking by self-assessment and observational exercise



Lastly, Table A1.8 shows correlations between the number of training sessions a participant missed and his or her skills score, indicating that while a higher number of sessions missed is associated with a smaller increase in skills scores, this association is only statistically significant for the observational exercises. Those who participated in more sessions are thus associated with an increase in

communication and teamwork skills as measured by the observational exercises, but not as measured by the self-assessment. The skills program was explicitly geared to improve skills such as communication and teamwork. The fact that attending the sessions does not show a significant correlation with the self-assessment measures seems to suggest that the self-assessment measures something akin to personality traits, which are rather stable over time. The observational exercises do positively correlate with the number of sessions attended, though. This suggests that the observational exercises are better able to capture the part of non-cognitive skills that can be changed by a skills training program such as *MUVA* Attitude.

1.7 Discussion and conclusion

This paper contrasts two types of measurement of non-cognitive skills over three periods of time for the same individuals: young adults participating in a skills training program in urban Mozambique. Non-cognitive skills training programs have received continued attention over the last several years, due to the positive effect that these skills have been shown to exhibit on employment and life outcomes. Most of the time, these skills have been captured by self-assessments. Contrasting this traditional type of measurement with novel observational exercises provides useful information for both researchers and policy makers to better understand what is measured by the different instruments.

Results show that the self-assessment used in the case at hand suffers from response bias in the form of yag-saying: participants are more prone to agree with a statement than disagree. The items are therefore measured with error, which leads to less stable constructs but correcting for acquiescence bias improves the reliability of pre-defined dimensions. Further, while seven non-cognitive skill dimensions were thought to be captured by the self-assessment, exploratory factor analysis could only confirm two dimensions (communication and aspirations), with a third dimension being extracted as well; this last factor is unclear however. Information on individual raters shows that raters' individual scores are quite similar, with most of the disagreement stemming from closer disagreement.

Results further show that the self-assessment and observational exercises seem to capture different underlying concepts: the self-assessment seems to be closer to capturing personality traits, which are rather stable, while the observational exercises cover personality states, which are more temporary behaviors and feelings. This is evident from the low correlations between the communication and teamwork dimensions, which were measured with both instruments, but only correlate lowly. Further, the individual ranking of the self-assessment hardly changes, i.e. participants who scored highest on one trait at baseline are also likely to score highest on this trait at endline, whereas the ranking of the dimensions from the observational exercises sees changes, with more participants having personal motivation as their

highest ranking at endline compared to baseline. This is in line with data on training sessions missed. We cannot find any significant correlations between sessions missed and the self-assessment dimensions, but correlations between sessions missed and the observational exercises are significant. This means that those who missed more sessions tend to have lower scores on the observational exercise dimensions communication and teamwork.

This paper illustrates that it is important to understand what is measured by any given measurement tool. It suggests that if the purpose of measurement is to assess changes in states as a result of a skills training program, observational exercises can be a good solution. If the purpose of measurement is to determine the impact of stable traits on life or employment outcomes, self-assessments are appropriate. In the case of self-assessments, it should be assessed whether acquiescence bias is a problem, and if so, corrections should be undertaken. This means that in the design of the questionnaire, opposing pairs of questions should be included. Further, in the design and testing of self-assessment survey instruments, it is important to assure that the questionnaire is consistent and reliable. In our case, the seven pre-defined dimensions supposed to be covered by the self-assessment were not represented well by the data. An additional idea could be to include anchoring vignettes, to be able to correct for other forms of response bias.

Appendix for Chapter 1: Tables and Figures

Table A1.1 – Descriptive statistics for panel sample, cycle 1 and cycle 2

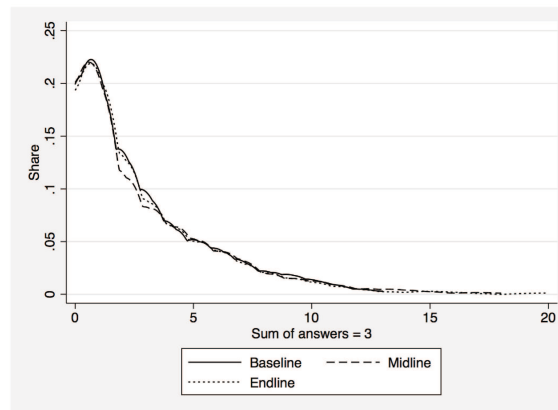
	Cycle 1					Cycle 2					Difference	
	N	Mean	Std. Dev.	Min	Max	N	Mean	Std. Dev.	Min	Max	Diff means	Std Err
Female	162	0.62	0.49	0	1	192	0.67	0.47	0	1	-0.05	0.05
Age	147	23.71	3.65	15	31	178	22.43	3.38	17	30	1.28	0.39
Married	152	0.33	0.47	0	1	182	0.29	0.46	0	1	0.04	0.05
Has one child	162	0.31	0.47	0	1	192	0.25	0.43	0	1	0.06	0.05
Has 2+ children	162	0.27	0.44	0	1	192	0.18	0.38	0	1	0.09	0.04
Average years of schooling	162	9.42	2.14	1	12	163	9.71	2.22	2	12	-0.29	0.24
Worked in last 30 days	152	0.31	0.46	0	1	185	0.22	0.42	0	1	0.09	0.05
Earnings for those who did	43	1829.88	1728.44	30	7000	39	2562.18	6365.69	0	40000	-732.30	1008.89
Has done internship before	162	0.20	0.40	0	1	192	0.20	0.40	0	1	0.01	0.04
Has done prof training course	162	0.17	0.37	0	1	192	0.22	0.42	0	1	-0.06	0.04
Portuguese speaking	152	2.95	0.21	2	3	185	2.86	0.34	2	3	0.09	0.03
Portuguese reading	152	2.95	0.22	2	3	184	2.96	0.19	2	3	-0.01	0.02
Portuguese writing	152	2.93	0.27	1	3	183	2.93	0.27	1	3	0.00	0.03
English speaking	148	1.45	0.56	1	3	184	1.49	0.56	1	3	-0.05	0.06
English reading	151	1.62	0.61	1	3	185	1.62	0.62	1	3	0.00	0.07
English writing	151	1.51	0.58	1	3	185	1.46	0.60	1	3	0.05	0.06
ICT skills	150	1.43	0.69	1	3	185	1.54	0.74	1	3	-0.11	0.08
<u>HH variables</u>												
Poverty score	150	43.02	10.50	9	68	184	48.95	12.92	9	82	-5.93	1.31
Dependency ratio	150	0.70	0.57	0	2.67	185	0.75	0.88	0	8	-0.04	0.08
HH head female	149	0.40	0.49	0	1	184	0.29	0.45	0	1	0.11	0.05
<u>Maputo areas</u>												
Polana Canico	162	0.20	0.40	0	1	192	0.15	0.36	0	1	0.05	0.04
Aeroporto	162	0.18	0.38	0	1	192	0.17	0.38	0	1	0.01	0.04
Inhagoia	162	0.15	0.36	0	1	192	0.15	0.36	0	1	0.00	0.04
Chamanculo	162	0.11	0.32	0	1	192	0.13	0.34	0	1	-0.02	0.03
<u>Beira areas</u>												
Munhava	162	0.15	0.36	0	1	192	0.18	0.38	0	1	-0.03	0.04
Chipangara	162	0.22	0.41	0	1	192	0.22	0.41	0	1	0.00	0.04

Notes: This table only takes the panel sample into account, i.e. those who are present in all three survey rounds and have non-missing skills data.

Table A1.2 – Observational exercise assessment criteria

<u>Group exercise</u>	
Listening	Does not interrupt others. Listens to the arguments of others and makes an effort to understand them and take their opinions into account. Asks questions to understand what others want to say.
Negotiation	Actively works together with others towards the common goal and participates in solving the problem.
Motivation (others)	Helps to create a positive and energetic atmosphere in the group and values the work/opinions of others.
Creativity	Tries to find innovative ideas and materializes them in a creative way (uses various materials).
Flexibility	Reacts positively to new and unexpected challenges and is open to new ideas.
Personal motivation	Approaches the task with a motivated and positive attitude.
<u>Individual exercise</u>	
Body language	Confident posture and voice when speaking, calm attitude, not afraid to make eye contact with others when speaking.
Logical argumentation	Logical arguments which are well organized and relevant. Can formulate questions and structure ideas.
Professional attitude	Speaks in a polite way and does not use inappropriate (foul) vocabulary. Is on time and does not use phone during the exercise.
Speaking confidence	Defends ideas in front of others and is not afraid to speak up or disagree.

Figure A1.1 – Share of people neither agreeing nor disagreeing with a statement



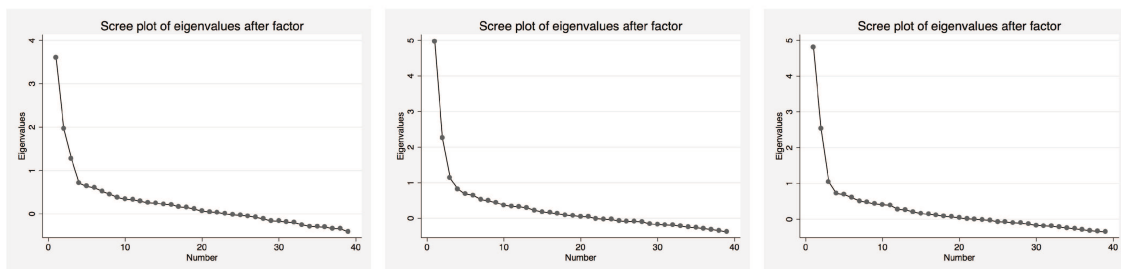
Notes: The figure shows the number of times that people answered 3 (neither agree nor disagree) for all self-assessment questions (40 questions in total) per survey round.

Figure A1.2 – Scree plots for naive self-assessment

(a) Baseline

(b) Midline

(c) Endline



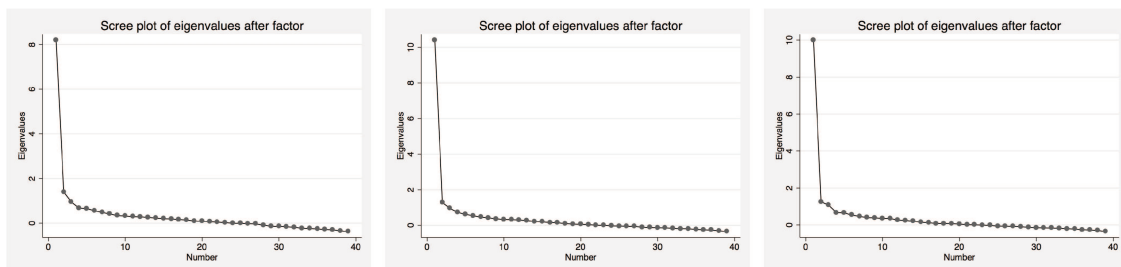
Notes: Sample size: 354 individuals

Figure A1.3 – Scree plots for acquiescence corrected self-assessment

(a) Baseline

(b) Midline

(c) Endline



Notes: Sample size: 354 individuals

Table A1.3 – Exploratory factor analysis -
Rotated factor loadings at Midline

(a) Uncorrected items				(b) Corrected items			
	F1	F2	F3		F1	F2	F3
com_139	0.65			cr_com_109	0.76	-0.13	-0.11
com_119	0.60			cr_com_139	0.76		
abc_111	0.60			cr_com_119	0.68		
com_109	0.58		-0.14	cr_com_130	0.66		
com_130	0.53		-0.10	cr_abc_111	0.64	0.13	-0.11
com_141	0.52		0.12	cr_com_141	0.62		
com_100	0.49			cr_com_100	0.62		
em_129	0.44	0.17	0.13	cr_abc_138	0.61		0.22
gr_122	0.43		0.21	cr_we_151	0.58		
abc_138	0.43		0.29	cr_gr_122	0.56	0.12	0.13
we_151	0.42	0.10	0.18	cr_com_124	0.51	-0.11	0.30
em_105	0.34	0.12		cr_abc_117	0.49		
com_132	0.29	0.28	-0.31	cr_asp_120	0.43	0.22	0.25
abc_117	0.26			cr_em_129	0.43	0.33	
we_103	0.24			cr_com_132	0.42		-0.36
com_143	-0.60			cr_gr_147	0.41	0.31	
asp_116		0.57		cr_em_105	0.41	0.23	
em_108		0.54	-0.17	cr_we_103	0.39	0.12	0.11
asp_137		0.53	0.22	cr_gr_112	0.37	0.12	
gr_115	0.18	0.48		cr_abc_125	0.35		0.25
gr_133		0.45		cr_abc_150	0.32		-0.18
tw_149	0.12	0.45	-0.11	cr_em_123	0.26		
gr_147	0.20	0.41		cr_asp_144	0.25	0.16	
tw_135	0.22	0.39		cr_em_118	0.21	0.12	-0.24
asp_107		0.39	0.11	cr_asp_116		0.65	
we_128		0.34	0.28	cr_asp_137		0.62	
tw_146		0.34	0.13	cr_gr_133		0.45	-0.10
abc_150	0.16	0.30	-0.11	cr_em_108		0.45	-0.23
em_118		0.29	-0.20	cr_tw_149	0.20	0.43	-0.14
com_140	-0.18	0.28		cr_we_128	0.14	0.42	0.26
asp_106		0.23	0.14	cr_gr_115	0.23	0.41	-0.12
gr_112	0.19	0.19		cr_asp_107		0.40	
asp_144		0.18		cr_tw_135	0.29	0.39	-0.12
tw_127			0.43	cr_tw_146	0.17	0.37	0.13
asp_120	0.27		0.34	cr_com_140		0.32	0.10
abc_125	0.16		0.30	cr_asp_106	0.22	0.21	0.15
com_124	0.22		0.26	cr_tw_127	0.21	0.34	0.40
we_126	-0.32	0.12	0.17	cr_com_143	-0.28	-0.19	0.39
em_123			-0.14	cr_we_126		0.15	0.31

Notes: blanks represent $abs(\text{loading}) < 0.1$. F stands for factor. Items are labeled according to pre-defined categories. These are: com = communication, tw = teamwork, we = work ethic, em = empathy, gr = grit, abc_ri = ability to deal with criticism, asp = aspirations. Items in bold are items that should be retained for the construction of the factor.

Table A1.4 – Exploratory factor analysis -
Rotated factor loadings at Endline

(a) Uncorrected items				(b) Corrected items			
	F1	F2	F3		F1	F2	F3
com_130	0.60			cr_com_109	0.71		
com_119	0.56		0.19	cr_com_119	0.68		
com_109	0.56			cr_com_130	0.65		
com_139	0.54			cr_abcri_111	0.61	0.19	
abcri_111	0.53	0.15		cr_com_139	0.60	0.11	
com_100	0.41		0.15	cr_com_141	0.59		0.20
com_141	0.40		0.22	cr_com_100	0.55		0.10
com_132	0.32	0.27	-0.12	cr_com_132	0.52		-0.18
we_103	0.25	-0.17		cr_abcri_117	0.43		0.15
abcri_117	0.22		0.16	cr_abcri_150	0.37	0.19	-0.26
asp_116	-0.11	0.66		cr_we_151	0.36		0.25
tw_135	0.16	0.57		cr_tw_149	0.36	0.31	
asp_137		0.56		cr_we_103	0.35		0.25
asp_107	-0.14	0.49	0.13	cr_com_124	0.31		0.29
gr_115		0.48	0.26	cr_em_118	0.29		-0.19
we_128		0.44		cr_gr_112	0.25	0.13	
asp_106		0.44		cr_tw_146	0.28	0.16	0.24
gr_147		0.42	0.27	cr_asp_116	-0.16	0.76	
tw_149	0.22	0.39		cr_tw_135	0.14	0.65	
gr_133		0.39	-0.23	cr_asp_137		0.60	
abcri_150	0.26	0.35	-0.12	cr_asp_107		0.58	0.11
em_108		0.34	0.15	cr_gr_115		0.54	
em_123	0.12	0.32	-0.16	cr_asp_106		0.47	0.13
asp_144		0.32		cr_we_128	0.14	0.45	
em_118	0.11	0.28	-0.13	cr_gr_147	0.15	0.41	0.28
tw_146		0.27	0.14	cr_gr_133		0.39	-0.16
gr_112	0.14	0.20	-0.11	cr_em_108		0.39	
com_140	-0.14	0.15		cr_asp_144		0.38	
we_126	-0.34	0.14		cr_em_129	0.29	0.37	0.22
abcri_138	0.14		0.46	cr_gr_122	0.29	0.33	0.17
em_105	0.11	0.10	0.45	cr_em_105	0.21	0.32	0.30
em_129	0.25	0.15	0.35	cr_em_123	0.17	0.23	-0.13
we_151	0.18		0.34	cr_com_140		0.13	0.12
tw_127			0.34	cr_tw_127		0.27	0.41
abcri_125			0.33	cr_abcri_138	0.29	0.13	0.36
asp_120			0.29	cr_asp_120	0.16	0.24	0.35
gr_122	0.27		0.29	cr_com_143	-0.46		0.34
com_124	0.17	-0.12	0.19	cr_abcri_125	0.30		0.32
com_143	-0.54			cr_we_126	-0.22	0.12	0.17

Notes: blanks represent $abs(\text{loading}) < 0.1$. F stands for factor. Items are labeled according to pre-defined categories. These are: com = communication, tw = teamwork, we = work ethic, em = empathy, gr = grit, abcri = ability to deal with criticism, asp = aspirations. Items in bold are items that should be retained for the construction of the factor.

Table A1.5 – Number of raters per period, % of sample

Group exercise			
# of raters	Baseline	Midline	Endline
2	50.63	75.83	74.5
3	39.62	24.17	21.78
4	8.18	0	3.72
5	1.57	0	0
Individual exercise			
# of raters	Baseline	Midline	Endline
2	23.27	9.67	6.3
3	65.09	89.12	85.39
4	10.69	1.21	8.31
5	0.94	0	0

Notes: N = 354 per round.

Table A1.6 – Deviation from mean in observational scores

	(1) Lis	(2) Neg	(3) MotOth	(4) Crea	(5) Flex	(6) PersMot	(7) BodyL	(8) LogArg	(9) ProfAtt	(10) Conf
20-14yo	0.01 (0.03)	-0.02 (0.03)	0.04 (0.03)	-0.05* (0.03)	0.05* (0.03)	0.03 (0.03)	0.02 (0.02)	-0.03 (0.02)	0.06** (0.02)	-0.04* (0.02)
25-30yo	0.01 (0.03)	-0.03 (0.03)	0.02 (0.03)	-0.04 (0.03)	0.00 (0.03)	0.02 (0.03)	0.05** (0.03)	-0.02 (0.03)	0.07*** (0.03)	-0.03 (0.03)
>30yo	0.06 (0.16)	0.22 (0.19)	0.31 (0.19)	-0.02 (0.17)	0.17 (0.18)	0.09 (0.18)	-0.12 (0.15)	0.30** (0.15)	0.28* (0.16)	0.02 (0.15)
Male	-0.00 (0.02)	-0.00 (0.02)	0.02 (0.02)	0.04** (0.02)	0.04** (0.02)	0.03 (0.02)	0.04** (0.02)	0.03** (0.02)	0.00 (0.02)	0.01 (0.02)
Midline	0.04* (0.02)	0.04* (0.03)	0.03 (0.03)	0.05** (0.02)	-0.01 (0.02)	0.01 (0.02)	0.02 (0.02)	0.00 (0.02)	-0.01 (0.02)	0.01 (0.02)
Endline	0.02 (0.02)	0.02 (0.03)	0.02 (0.03)	0.03 (0.02)	-0.02 (0.02)	-0.01 (0.02)	0.01 (0.02)	-0.04** (0.02)	0.02 (0.02)	-0.04** (0.02)
N	993	993	993	993	993	993	993	993	993	993
R^2	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1; Base category for age group is 15-19 years old. Sample size smaller (993 instead of 1062) due to missing age for some individuals.

Table A1.7 – Correlations between different measures of non-cognitive skills

	Obs: tw	Obs: pm	Obs: com	Sa: com	Sa: tw	Sa: we	Sa : em	Sa: asp	Sa: abcri	Sa: gr	F1: com	F2: asp
Obs: pm	0.85											
Obs: com	0.42	0.46										
Sa: com	0.21	0.24	0.25									
Sa: tw	0.27	0.27	0.17	0.60								
Sa: we	0.12	0.15	0.11	0.51	0.49							
Sa : em	0.23	0.23	0.23	0.60	0.53	0.39						
Sa: asp	0.22	0.25	0.21	0.54	0.60	0.47	0.50					
Sa: abcri	0.18	0.21	0.23	0.65	0.52	0.45	0.53	0.52				
Sa: gr	0.16	0.18	0.19	0.61	0.59	0.51	0.58	0.58	0.56			
F1: com	0.25	0.27	0.26	0.93	0.66	0.56	0.71	0.62	0.80	0.69		
F2: asp	0.25	0.26	0.20	0.56	0.74	0.48	0.54	0.90	0.53	0.65	0.64	
F3: diff	0.05	0.05	0.04	0.17	0.18	0.41	0.09	0.13	0.08	0.12	0.08	0.13

Notes: N = 1062, which is 354 individuals over 3 periods. The table depicts all periods combined. Obs stands for observational exercises. Sa stands for self-assessment, these are the pre-defined dimensions based on acquiescence corrected items. F stands for factor, as extracted through EFA from the self-assessment based on acquiescence corrected items. The dimensions, either pre-defined or confirmed by the data are: com = communication, tw = teamwork, we = work ethic, em = empathy, gr = grit, abcri = ability to deal with criticism, asp = aspirations.

Table A1.8 – Correlations of skills with number of training sessions missed

	# of days missed
Sa: com	-0.05 (0.24)
Sa: tw	-0.01 (0.79)
Sa: we	-0.04 (0.43)
Sa: em	-0.05 (0.22)
Sa: asp	-0.05 (0.26)
Sa: abcri	-0.02 (0.59)
Sa: gr	-0.07 (0.1)
Obs: tw	-0.12* (0.00)
Obs: pm	-0.06 (0.15)
Obs: com	-0.25* (0.00)

Notes: Sample size is 701 individuals since we combined midline and endline data but have some missing data for the number of sessions attended. Sa refers to self-assessment. Obs refers to observational exercises. Significance level for correlations in brackets. Correlations significant at 5% level marked with *.

Chapter 2

Cognitive and non-cognitive skills, hiring channels, and wages in Bangladesh

Abstract¹

This paper uses a novel matched employer-employee data set representing the formal sector in Bangladesh to provide descriptive evidence of both the relative importance of cognitive and non-cognitive skills in this part of the labor market and the interplay between skills and hiring channels in determining wages. While cognitive skills (literacy, a learning outcome) affect wages only by enabling workers to use formal hiring channels, they have no additional wage return. Non-cognitive skills, on the other hand, do not affect hiring channels, but they do enjoy a positive wage return. This wage return differs by hiring channel: those hired through formal channels benefit from higher returns to openness to experience, but lower returns to conscientiousness and hostile attribution bias. Those hired through networks enjoy higher wages for higher levels of emotional stability, but they are also punished for higher hostile attribution bias. This is in line with different occupational levels being hired predominantly through one channel or the other. We provide suggestive evidence that employers might use hiring channels differently, depending on what skill they deem important: employers valuing communication skills, a skill that could arguably be observed during selection interviews, are associated with a larger within-firm wage gap between formal and network hires, while the importance of teamwork, a skill that is more difficult to observe at the hiring stage, is associated with a smaller wage gap.

¹This chapter is co-authored with Christophe Jalil Nordman and Leopold R. Sarr.

2.1 Introduction

Human capital has long been recognized as a predictor of wages and other labor market outcomes. While the literature has traditionally focused on the effect of classical human capital components such as education and experience on the labor market, non-cognitive skills like persistence, motivation, and communication have more recently been found to have an effect on wages and labor market outcomes that extend beyond cognitive ability.

Non-cognitive skills are likely to have both direct and indirect effects on wages. The direct effect stems from personality being considered as part of a worker’s endowment. The indirect effect comes from personality affecting, for instance, occupational choice (Cobb-Clark and Tan, 2011), educational attainment (Heckman et al., 2006), or job search methods (Caliendo et al., 2015). Literature so far has focused on developed countries and oftentimes on either direct or indirect effects alone. Little research has been done regarding wage returns to non-cognitive skills in developing countries, which are likely to be different. Employers in developing countries might, for example, reward skills that deal with the precise execution of tasks (such as being conscientious and emotionally stable) more than skills that deal with intellectual curiosity and independent working (such as openness to experience or extraversion).

We use a novel matched employer-employee data set from Bangladesh to estimate wage returns to cognitive and non-cognitive skills while taking into account one important feature of the labor market in developing countries: the choice of hiring channel. We thus consider both the direct and indirect effects of skills on wages. To our knowledge, no paper has so far considered both the direct and indirect effect of cognitive and non-cognitive skills on wages, especially in a developing country context. This paper does both, first looking at the effect of cognitive and non-cognitive skills on hiring channels chosen and then estimating outcomes of the job search and returns to wages for those skills, given the initial selection into hiring channels.²

The cognitive skills included are measured by numeracy and literacy tests. These cognitive skills are not ‘pure’ cognitive skills in the sense of capturing only intelligence. Instead, they are learning outcomes, capturing levels of intelligence needed to accomplish a task as well as the effort on that task and motivation to complete it. As such, they are an imperfect measure of pure cognitive skills, but they are the only measure available in the data set and have been used as proxies in the literature (Hanushek and Woessmann, 2008). The non-cognitive skills considered are measured by the Big Five personality test (openness to experience, conscientiousness, extraversion, agreeableness, and emotional stability); a set of

²The idea of the paper is to endogenize the hiring channel. We certainly admit that other choices could be endogenous as well, e.g. job choice or task choice; however, this paper deliberately decided to only look at the hiring channel.

socio-emotional skills (hostile attribution bias and grit) is also included. All are based on a self-assessed questionnaire. Given the nature of the survey (a matched employer-employee data set) we are able to include firm fixed effects, which allows us to account for firm-specific heterogeneity in rewarding certain skills. Additionally, we explore demand-side preferences for specific skills by decomposing within-firm wage gaps between workers who have been hired through formal channels and those hired through networks to better understand whether those gaps are driven by firm characteristics or preferences for certain skills.

The survey covers the formal sector, which is the sector that will contribute most strongly to the shift from agriculture into higher skilled sectors in Bangladesh. While most Bangladeshi employment still occurs in the informal sector, three quarters of new jobs over the last ten years were added in the more formal, non-agricultural sector (World Bank, 2015). Still, formal sector employment in Bangladesh represented only about 11.5 percent of total employment in 2010 (Asian Development Bank and Bangladesh Bureau of Statistics, 2012). While our results are thus only applicable to this small share of the Bangladeshi labor market, our paper still contributes to a better understanding of which skills are required in the part of the labor market likely to be the primary driver of faster GDP growth and poverty reduction.

We find that non-cognitive skills do not affect the selection of the hiring channel, but they do have a direct effect on wages, after correcting for initial selection. These significant correlations are not visible in simple OLS regressions and differ by hiring channel; they therefore illustrate the benefit of first taking into account selection into different hiring channels. For those hired formally, the wage returns to an additional standard deviation of openness to experience is 2.5 percent. However, not all non-cognitive skills benefit from positive wage returns: those hired formally are punished for having higher levels of conscientiousness and hostile attribution bias. Looking at differential results by occupation shows that the result for conscientiousness is driven by white collar workers, among whom other traits, such as creativity or flexibility are arguably more important than diligently completing tasks. Among those hired through networks, hostile attribution bias also shows a negative wage return. Network hires can benefit from a positive return to emotional stability though with a one standard deviation increase in the trait associated with a 1.9 percent wage increase. These are considerable returns, especially when compared to average returns to education in developing countries. Returns to skills thus seem to align with the task content of work.

We then explore the demand side to understand whether firm characteristics and preferences for certain skills can help explain differential returns to skills. We find that employers who value communication skills more are associated with a larger within-firm wage gap between formal and network hires, while those who value teamwork more are associated with a smaller wage gap, though this result

only holds for professional workers, such as managers or technicians. We explain this result through the firm’s decision to hire through different channels among highly skilled workers. If the employer values communication skills, they will choose a hiring channel that allows them to observe these skills. As communication skills can arguably be well observed during formal interviews, this increases the wage gap between formal and informal hiring. Teamwork skills, on the other hand, might be more difficult to observe reliably in a simple job interview, which is why firms valuing these skills might rely more on networks to provide otherwise unobservable information about a worker. This mechanism only seems to hold for professional workers; among non-professional workers, other mechanisms, such as firms hiring through networks to overcome moral hazard problems of workers shirking, as observed in Heath (2018), could dominate.

The nature of this paper is purely descriptive, as the data do not allow us to draw any causal inferences. Still, we believe that it is valuable to engage in a descriptive analysis given the lack of information on returns especially to non-cognitive skills in a developing country setting. Further, to our best knowledge, no paper has previously examined the potential interplay between hiring channels and non-cognitive skills, a factor that could be of particular importance in a developing country, in which hiring through networks is very prevalent.

The remainder of this paper is organized as follows: section 2.2 provides an overview of the literature and lays out our conceptual framework, section 2.3 introduces the data, section 2.4 describes the methodology used, section 2.5 presents our results, and section 2.6 concludes.

2.2 Literature and conceptual framework

2.2.1 Literature review

The impact of education and experience on wages has been widely discussed since the seminal work of Becker (1964) and Mincer (1974).³ Still, focusing solely on returns to education ignores the innate multidimensionality of human capital, which combines cognitive ability and non-cognitive skills. Lack of data has meant that cognitive and non-cognitive skills have been oftentimes part of the unobservable.

Cognitive skills are often approximated via standardized test scores in developed countries or literacy and numeracy tests in developing countries. Hanushek and Woessmann (2008) provide evidence that it is the possession of cognitive skills, rather than mere school attainment, that is most powerfully related to individual earnings, and that average years of education becomes insignificant once test scores are included as an additional control variable. Still, it is important to note that cognitive skills measured by numeracy and literacy are less of a measure of true

³Psacharopoulos and Patrinos (2004) provide a global overview and estimate that the average rate of return to another year of schooling is about 10 percent.

cognitive ability than other instruments, such as Raven’s Progressive Matrices. While most of the literature has focused on the United States, similar results have been found for other developed countries. Hanushek et al. (2015) use a cross-country data set for OECD countries and find that returns to cognitive skills (numeracy and literacy) are considerable but smaller on average than in the US. In developing countries, Lee and Newhouse (2012) find that higher cognitive ability is associated with measures of better job quality.⁴

Non-cognitive skills such as personality traits and behaviors have recently been included in the analysis of labor market outcomes. Being able to learn, to lead, to communicate, to work in a team, or to deliver results in a timely manner might in the end be as important as cognitive ability. Indeed, non-cognitive skills have been found to be strongly associated with higher earnings, to have a positive effect on wages beyond the effect of pure cognitive ability (Heckman et al., 2006; Heineck and Anger, 2010), and to be a predictor of labor market outcomes (Borghans et al., 2008a). Mueller and Plug (2006) show that the effect of personality traits on earnings is of a similar magnitude to that of cognitive skills. In the framework of the Big Five, conscientiousness (Nyhus and Pons, 2005) and emotional stability (Drago, 2011) have been linked to better job performance and higher wages. Recently, “grit”, perseverance in the pursuit of long-term goals, has taken a prominent place in predicting success in several settings, such as educational attainment and grade point average (Duckworth et al., 2007), suggesting that such perseverance might be more important than, or at least as important as, actual talent.

Research on returns to non-cognitive skills has largely focused on developed countries, with a few notable exceptions. Blom and Saeki (2011) find evidence that employers of engineers in India stress interpersonal skills such as reliability and willingness to learn above cognitive skills such as literacy and numeracy. In Peru, Díaz et al. (2013) find that returns to perseverance are as high as returns to average cognitive ability. Other papers have found rather mixed evidence: Glewwe et al. (2017) show that, in China, both cognitive and non-cognitive skills are important for the school-to-work transition, but they do not predict wages; similarly, Acosta et al. (2015) find a larger impact of cognitive than non-cognitive skills on labor market outcomes; and OECD (2015) suggests that raising cognitive skills might ultimately be more important than raising socio-emotional skills in determining incomes. However, recent intervention aimed at improving non-cognitive skills shown some success, such as among adolescents in India (Krishnan and Krutikova, 2013).

This discussion shows that personality traits have both a direct and an indirect

⁴Assessing the differential impacts of years of schooling and cognitive skills on labor market outcomes is not without problems, since including intelligence tests as control variables for ability could worsen rather than improve their inherent bias, as these tests are themselves subject to measurement error (Griliches, 1977). Cawley et al. (2001) illustrates this difficulty: using the US National Longitudinal Survey of Youth (NLSY), they find that cognitive ability and schooling are so highly correlated that their effects on wages cannot be estimated without imposing strong parametric structures on any estimation.

impact on wages by determining, for instance, labor market participation, the school-to-work transition, or formality status. In this paper, we are focusing on one indirect factor: the use of social networks in the hiring process. If individuals self-select into different channels of job search depending on their human capital, the hiring channel could then mitigate wage returns to skills. This is particularly important in developing countries, where formal institutions are traditionally weak and social networks are employed frequently in job searches (Fafchamps, 2006). Social networks have been found to both increase and decrease wages as outlined below.

On the supply side, informal networks are understood to be preferred by workers because they are less expensive and characterized by a higher probability of finding a job (Holzer, 1988). On the demand side, the use of social networks has traditionally been justified by mitigating selection problems through reduced asymmetric information between employers and employees and improved matching (Montgomery, 1991; Simon and Warner, 1992). In these models, current employees possess information about unobserved characteristics of applicants or the match quality. Jobs obtained via social networks should then result in higher wages (e.g. Kugler, 2003; Simon and Warner, 1992). Burks et al. (2015) look at employee referrals in three industries (call centers, trucking, and high-tech) and find that workers who obtained their jobs through employee referral earn slightly higher wages and have lower turnover and recruiting costs. Heath (2018) examines the garment sector in Bangladesh, in which jobs are rarely formally published. She finds that firms use referrals to mitigate the moral hazard problem by being able to punish referrers if they refer an unproductive worker.

Having to use social networks to find a job can also be perceived as a negative signal to employers, however, as those relying on networks might simply be workers in greater need of a job of any sort (Granovetter, 1995). If the latter channel is at work, workers searching for jobs via social networks exhibit a lower reservation wage, suggesting that finding a job through a social network will thus be negatively related to wages. Empirical evidence can also be found in support of a negative relationship between social network job search and wages (e.g. Bentolila et al., 2010; Berardi, 2013).

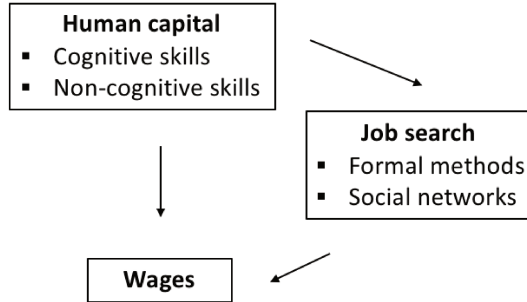
Most of the models in the network literature assume exogenous network formation and homogeneity of workers and firms. Taking into account endogenous network formation (particularly concerning the size of a network) and heterogeneity of workers and firms can lead to ambiguous prediction of network hiring on wages (Beaman, 2016). Human capital has been found to affect network formation. Lee et al. (2014) suggest that individuals with higher cognitive skills have access to broader social networks and are better able to signal their productivity through their social network. Individuals who possess more of certain non-cognitive skills (openness to experience, extraversion, and emotional stability) similarly have access

to larger and more diverse social networks (Pollet et al., 2011; Wu et al., 2008). Caliendo et al. (2015) further show that non-cognitive skills also influence job search as individuals with an internal locus of control exert higher search effort.

2.2.2 Conceptual framework

The literature demonstrates that cognitive and non-cognitive skills are likely to have a direct and an indirect effect on wages. The direct effect of non-cognitive skills stems from personality being considered part of a worker’s endowment, which is rewarded directly by the employer if this endowment matches the requirements for the job. The indirect effect stems from cognitive skills and personality affecting channels that lead to a job such as occupational choice (Cobb-Clark and Tan, 2011), educational attainment (Heckman et al., 2006), or intensity of job search (Caliendo et al., 2015). This paper combines the direct and indirect effects by first looking at the effect of cognitive and non-cognitive skills on hiring channels chosen and then estimating outcomes of the job search and returns to wages for those skills, given the initial selection into hiring channels (Figure 2.1).

Figure 2.1 – Conceptual framework illustrated



Non-cognitive skills in our paper are captured by the Big Five framework of personality traits. The Big Five are a widely used list of key traits, which are understood to capture the broadest level of personality traits.⁵ The Big Five framework includes: (1) openness to experience, which captures one’s tendency to be open to new experiences (aesthetic, cultural or intellectual); (2) conscientiousness, one’s tendency to be organized, hardworking, and responsible; (3) extraversion, directing one’s interest towards the outer world of people and things; (4) agreeableness, the tendency to act cooperatively and in an unselfish manner; and (5) emotional stability, predictability and consistency in emotional reactions with absence of rapid mood changes. The survey further includes the socio-emotional skills grit and

⁵The Big Five factor model is usually attributed to Allport and Odbert (1936), who theorized that important human individual differences are encoded in language. Allport and Odbert used personality-describing words from English dictionaries, which they condensed into five broad factors using factor analysis. The Big Five taxonomy has since been replicated across cultures (John and Srivastava, 1999) and developmental stages of the life course (Soto et al., 2008).

hostile attribution bias. Grit is the tendency to sustain interest in long term goals and persistence; hostile attribution bias refers to a systematic bias which leads to individuals interpreting the ambiguous behavior of others as hostile towards them.

To illustrate the conceptual framework, assume that a worker is very agreeable. Individuals who score highly on this personality trait are usually considerate, generous, willing to compromise their interests with those of others, and likely to have a larger and closer group of friends. Thus, we assume that scoring more highly on agreeableness increases a worker's probability of finding a job through friends or other social connections. On the job, agreeableness is perceived as positive, especially in teamwork settings. However, agreeable individuals are also likely to compromise their interests with those of others – not necessarily a valued tendency on the workplace. In terms of wage returns, the trait agreeableness could then first have an indirect effect on wages purely through enabling network hiring (positive or negative). In addition, agreeableness could have a direct effect on wages (positive or negative) depending on how well the trait enables the individual to perform his job.

Considering both direct and indirect effects, we expect education to have a positive effect on formal hiring, as the network channel has been found to be used more often by workers with a lower socio-economic status (Topa, 2010). Similarly, we expect cognitive skills to have a positive effect on the probability of having found the job through formal channels. People who are more extroverted, more outgoing, and take on more leadership roles have access to larger social networks and might thereby receive more information and offers through the network, increasing their likelihood of finding a job through their network. As mentioned in the example above, we assume agreeableness to have a positive effect on network hiring. The effect of conscientiousness, which has been linked to hard work and strong self-control, and emotional stability, which has been associated with better task performance, is unclear. On the one hand, friends or family might be more willing to provide a recommendation for a hard worker, increasing hiring through networks. On the other hand, the worker might prefer to opt for formal, official hiring channels instead of back-doors. Grit refers to perseverance in pursuit of long-term goals. As social networks often provide the quicker option to find a job, we assume that people who score higher on the grit scale continue their formal job search until they find a job. Grit would then have a negative effect on the probability of finding a job through networks.

Regarding direct effects, we expect our estimates for wage returns to education to be positive and potentially non-linear and our estimates for the returns to cognitive skills to be positive and significant. Literature on wage returns to the traits of the Big Five framework in developing countries is scarce. In developed countries, the literature has found positive effects for conscientiousness, extraversion, emotional stability, and grit on career success (Duckworth et al., 2007; Judge et al.,

1999). Conscientiousness and grit quite naturally coincide with professional success, extraversion can be especially useful in a business environment and emotional stability enables performance on specific tasks. Hostile attribution bias is assumed to have a negative effect on wages, but the effect of the remaining two Big Five personality traits on wages is less clear. Openness to experience is related to flexibility, creativeness, and intellectual orientation, which could have a positive effect on wages; at the same time, openness to experience has also been linked to more autonomy and non-conformity, which might not necessarily be rewarded in jobs for which strict obedience and subordination are, on the contrary, highly expected from employees. Similarly, agreeableness coincides with being more likable and cooperative, which is beneficial in a teamwork setting, but might be a hindrance if the worker places his co-workers' interests above his own.

2.3 Data

This paper is based on the 2012 Bangladesh Enterprise-Based Skills Survey (ESS), a matched employer-employee survey commissioned by the World Bank with the aim of assessing whether the educational system in Bangladesh is producing graduates with skills relevant to and demanded by firms (Nomura et al., 2013).⁶ The survey covers formal sector firms in the industrial and manufacturing sectors. Most of Bangladeshi employment occurs in the informal sector; it is therefore important to note that results from this data only hold for a rather selected sample of workers. In particular, the 2010 Bangladeshi Labor Force Survey (the survey closest to this data set) shows that only 11.5 percent of employment in Bangladesh occurs within the formal sector (Asian Development Bank and Bangladesh Bureau of Statistics, 2012). This share is slightly higher among men (13.2 percent), but it remains low. While the formal sector employment grew to 14 percent of total employment in 2016, it remains small on a global scale. Additionally, workers in formal firms are likely to be selected, due to better working conditions and wages in formal firms. In Bangladesh, the incidence of informal employment decreases with educational attainment (World Bank, 2018). For example, only 3.7 percent of employed Bangladeshi men have a high school degree and another 4.2 percent obtained a Bachelors degree, but their shares among formal workers are 10.6 and 18.7 percent, respectively (Asian Development Bank and Bangladesh Bureau of Statistics, 2012).

The survey covers firms in five industries: manufacturing, commerce, finance, education, and public administration. In total, the sample contains 500 firms and 6,981 individuals, stratified by economic sector and firm size: small (less than 20 employees), medium (21-70), and large (71+). Despite its limitation to five sectors, the survey is quite representative, as these sectors cover 87 percent of

⁶Two authors of this paper were part of the team designing and implementing the survey.

formal sector enterprises and 91 percent of total formal sector employment in Bangladesh (Nomura et al., 2013). Thus, the data used in this study only allow us to draw conclusions about a small part of the Bangladeshi labor market; still, conclusions drawn for the formal sector are representative.

The survey unit of the ESS is the firm; the survey consists of two modules, one each for employees and employers. The employee part of the survey is conducted for a sub-sample of employees in the sampled firms.⁷ It contains detailed information of each individual’s background, educational attainment, and numeracy and literacy skills, as well as their personality traits. Measures for numeracy and literacy stem from questions of the National Student Assessment conducted by the Bangladeshi Department of Primary Education (Nomura et al., 2013), designed to assess skills that workers who have completed primary school should possess. Again, the sample of workers in formal firms is on average much more educated than the population in general. According to the labor force survey, only 14.1 percent of all formal sector workers in 2010 did not complete primary education (Asian Development Bank and Bangladesh Bureau of Statistics, 2012). The test level thus seems highly appropriate and the test is administered to all workers, regardless of whether or not they have completed primary education. Personality measures are based on the Big Five typology of personality tests, with the short Big Five Inventory (BFI-S) included in the survey. It was originally developed by John and Srivastava (1999) and has been validated in large panel surveys such as the German Socio-Economic Panel Survey (McCrae and Costa, 2008).

The employee survey also asks workers to elicit information about their job search process. The question was asked in a nested way. First, workers are asked how they found their current job. Answer possibilities include media advertisement/posting, informal networks (including a reference from somebody), a school, public employment services, private employment services, job fairs, internet posting, and others (specify). Second, those who replied “informal networks” were then subjected to a follow-up question that inquired which informal network was used. Answer possibilities include family/relatives, friends, political affiliation, school alumni, and same village/town. For the remainder of this paper, all workers who said they used a job search method other than “informal networks” will be classified as having found their job through “formal channels”. Those who replied informal networks and answered the follow-up question on what type of network they used will be classified as having found their job through “networks”.

Responses for the employer module come from business owners and high-level managers; its questions deal with information on recruitment and training of employees, as well as an assessment of the workplace, and firm performance. It contains information on senior management itself (such as gender and education),

⁷Every 3rd person in a small firm; every 5th and 7th person in a medium and large firm; and if employment exceeds 200, every 30th person is interviewed (Nomura et al., 2013).

the firm’s preferred hiring channels, the importance of certain selection criteria for hiring potential employees (such as academic performance, skills, or affiliation with an informal network), whether the company has formal performance reviews for its workers, and the importance that it places on types of skills in its workforce (such as problem solving, motivation, or their ability to work as part of a team).

2.4 Methodology

Our methodological approach consists of estimating different models to assess the impact of cognitive and non-cognitive skills on wages and the choice of the hiring channel. We restrict our sample to male workers with non-missing skills variables, since men and women might differ in their endowment of non-cognitive skills (Fortin, 2008) or the same personality trait might be valued differently by employers according to gender (Heineck and Anger, 2010). Nordman et al. (2015) find that personality traits reduce the male-female wage gap in the upper part of the wage distribution using the same data set as this paper.⁸

We start from Mincer-type wage regressions and take into account the indirect effects of the use of networks through bimodal and multinomial switching models. We then move to the demand side to look at firm-level determinants of the wage gap between formal and network hires within the same firm. We decompose this observed within-firm wage gap to understand whether it is driven by firm characteristics or skill preferences.

2.4.1 Returns to skills

Our basic model is a simple Mincer specification:

$$\ln w_{ij} = \beta_0 + \beta_1 A_i + \beta_2 Cog_i + \beta_3 NonCog_i + \delta_j + \epsilon_{ij} \quad (2.1)$$

where $\ln w_{ij}$ is the natural logarithm of the hourly current wage for individual i in firm j . A is a vector of worker i ’s demographic characteristics, including years of formal education, total work experience, and a dummy for being married. Returns to education are likely to be non-linear, especially in developing countries (Kuépié et al., 2009; Söderbom et al., 2006), which we capture with a simplified model using a low-order polynomial (Card, 1999). We further introduce controls for occupation, as Heckman et al. (2006) have shown that individuals sort into occupations and education based on their personality traits, meaning our estimate of returns to personality could be overestimated if it simply captures occupational effects. Cog includes individual i ’s cognitive skills, the standardized score of the worker on the numeracy and literacy tests, and $NonCog$ his non-cognitive skills, the standardized scores of each dimension of the Big Five personality assessment

⁸Nakata and Nomura (2015) use the same data set and analyze the decomposition of the wage gap between workers hired formally and through networks.

and the standardized scores of the socio-emotional skills grit and hostile attribution bias. δ_j captures firm fixed effects to capture the effect that firm-level determinants of wages might have on wage setting (Abowd et al., 1999; Groshen, 1991).

The non-cognitive skills data in the survey suffer from a large amount of missing values. We therefore only keep individuals who answered at least two out of the three questions per dimension.⁹ We then correct the items for acquiescence,¹⁰ before assessing the internal validity of each trait by computing Cronbach's α . Correcting for acquiescence improves the internal validity for all constructs except for grit, for which it decreases validity substantially (from 0.22 to 0.09). We therefore keep the raw grit dimension and use acquiescence corrected version for the other skills. The threshold for good validity is $\alpha \geq 0.7$. None of our dimensions reach that threshold, but several come close. In ascending order, the α s are: 0.22 (grit), 0.33 (hostile attribution bias), 0.34 (agreeableness), 0.42 (openness to experience), 0.45 (extraversion), 0.61 (emotional stability), and 0.63 (conscientiousness).¹¹ The internal validity of our constructs is therefore not ideal and implies that results could suffer from measurement error, which would bias our results towards zero. We also consider creating scales based on the data and not our pre-defined dimensions, using exploratory factor analysis.¹² The three factors that can be retained from the data are a combination of emotional stability and extraversion (*Factor 1*), a combination of conscientiousness and grit (*Factor 2*) and a combination of agreeableness and openness to experience (*Factor 3*). This gives evidence to the observation that while one can theoretically distinguish between the different Big Five traits, in practice, they tend to be somewhat correlated. Due to the nature of the exploratory factor analysis, however, this approach reduces our sample size to those with non-missing skills variables only (N=3,102), which is why we stick to the Big Five distinction for most of the paper.

All our variables are measured at the same point in time (year 2012), after schooling has been completed and the worker has entered the labor market. If some of our control variables in the past influenced the degree to which others developed, our OLS estimates for returns to education or skills will be biased downwards, potentially underestimating the true effect. To illustrate, if cognitive ability is increased through education, by including controls for cognitive skills (literacy and numeracy) as well as educational attainment, our estimates for returns to cognitive skills would be the true partial effect. Cognitive skills and non-cognitive skills have been shown to be malleable by the educational system but also to be predictors of

⁹This reduces our sample from 6,092 men to 4,678 men. We thus lose about 23 percent of the raw sample.

¹⁰Acquiescence refers to a response bias in the form of being more likely to agree or disagree with questions in general.

¹¹We also checked α only for those who have a complete non-missing non-cognitive questionnaire. This decreases our sample size to 3,102 but does not significantly improve α .

¹²We use Kaiser's criterion, Catell's scree plots, and Horn's parallel analysis (Cattell, 1966; Horn, 1965; Kaiser, 1958) to decide how many factors to retain, settling on 3 factors. We then rotate factor loadings using an oblique, olimin rotation to allow for latent factors to be correlated. Final factors are scored using regression scoring and standardized.

educational attainment (Heckman et al., 2006).¹³ Additionally, measurement error in both cognitive and non-cognitive skills is quite likely. Our estimates should therefore be interpreted as lower bounds.

In a second specification, we use starting wages at firm entry instead of current wages as our dependent variable. Starting wages have the advantage of having been set before an employer is intimately familiar with a given worker. It is especially interesting to see the effect of cognitive and non-cognitive skills on starting wages, since at this point in time the employer’s information about workers consists only of observable information, such as experience and educational attainment, and knowledge received through the hiring channel (formal interviews, or information transmitted through a social network). If we find significant returns to some variable for current but not for starting wages, this suggests that effects are not spurious but stem from the employer learning about the worker’s productivity and match to their job.

Pairwise correlations between skills and personality traits show small correlations between the different personality traits and cognitive skills, but significant correlations between the personality traits themselves (see Table A2.1 in the Appendix). Our correlations are smaller than those found by Cunha and Heckman (2008), who show correlations of 0.3 between cognitive and non-cognitive factors, whereas our highest correlation coefficient is of the order of 0.16. Still, they carry the expected sign (e.g. grit and conscientiousness correlate positively and significantly with numeracy). The correlation between cognitive factors and years of education attained is very large for literacy (0.75) and still quite substantial for numeracy (0.49); correlations between years of education and personality traits are small.

2.4.2 Skills and the type of hiring channel

OLS regressions provide our starting point, but they confound the direct and indirect effects that cognitive and non-cognitive skills have on wages, as they are unable to account for unobserved characteristics that influence the choice of hiring channel and could also influence wages that the worker receives once employed. More extroverted workers might have a higher probability of finding their job through networks such as friends and family, for example. Once employed, the worker’s wage is influenced by having been hired through friends and networks (the indirect effect of extraversion). His wage will also be determined by the value the employer places on the worker’s extraversion on the job (the direct effect).

An endogenous switching model is able to correct for both the endogenous sample selection and the switching impact of wage determinants and has been

¹³ Almlund et al. (2011) discuss how using previously measured traits as predictors of later outcomes is problematic if the traits evolve over time. However, psychological research has demonstrated the stability of personality traits beginning in young adulthood (Mischel and Shoda, 2008), which has also been shown in surveys (Cobb-Clark and Tan, 2011).

used to investigate hiring channels in developed (Delattre and Sabatier, 2007) and developing countries (Berardi, 2013). We therefore estimate bimodal and multinomial endogenous switching models.

Workers can be hired through formal channels or through networks. Each worker is only observed in one regime at a time, leading to the following two equations:

$$\ln w_{ij}^F = \beta_0^F + \beta_1^F A_i + \beta_2^F Cog_i + \beta_3^F NonCog_i + \delta_j^F + \epsilon_{ij}^F \quad (2.2)$$

$$\ln w_{ij}^N = \beta_0^N + \beta_1^N A_i + \beta_2^N Cog_i + \beta_3^N NonCog_i + \delta_j^N + \epsilon_{ij}^N \quad (2.3)$$

where $\ln w_{ij}^F$ is the natural logarithm of the hourly wage rate of worker i hired through formal channels in firm j ; $\ln w_{ij}^N$ the natural logarithm of the hourly wage rate of worker i hired through networks in firm j ; vectors A , Cog , and $NonCog$ are vectors of worker i 's demographic characteristics, his cognitive skills, and his non-cognitive skills, respectively; δ_j are firm fixed effects, and ϵ_{ij}^F and ϵ_{ij}^N are the error terms for formal and network hires. We use the same worker characteristics as in the OLS regressions in section 2.4.1.

The switching regression then sorts individuals to one of the two regimes. Choosing which channel to use when engaging in job search is not exogenous, but depends on the expected gains or losses associated with finding a job formally or through networks, given one's level of skills. Worker i therefore engages in job search through his social networks (NET) if

$$NET_{ij}^* = \gamma Z_{ij} + u_{ij} > 0 \quad (2.4)$$

where Z_{ij} is a vector of explanatory variables for job search through social networks and u_{ij} is the error term. NET_{ij}^* is unobserved. We do, however, observe whether or not the individual was actually hired through his networks and thereby whether or not he used networks as at least one of his potentially multiple channels of job search:

$$NET_{ij} = \begin{cases} 1 & \text{if } NET_{ij}^* > 0 \\ 0 & \text{if not} \end{cases} \quad (2.5)$$

We are able to control for a wide range of characteristics at the worker and firm level. Still, workers searching through networks might differ from those searching through formal channels in their unobserved characteristics. Networks have long been thought to form among similar people (the ‘‘homophily’’ principle) (Lazarsfeld and Merton, 1954; McPherson et al., 2001); at the same time, the size and diversity of the network (‘‘strength of weak ties’’) is highly relevant for job search by linking individuals to information that is unavailable in their own circles (Granovetter, 1973). The relative importance of diversity versus similarity seems to depend on the current position of the individual in the social hierarchy (Lin, 1999). If

a worker is in a lower social position, relying on weaker ties and reaching out vertically will put the individual in contact with people in a higher social position. If the individual is already in a high position, reaching out vertically provides no such benefit. The individual would then benefit more from searching among his strongest ties instead, reaching out horizontally.

We introduce *mother has no formal education* as an identifying variable to approximate the aspect of network quality which could influence its usage. The data allow us to control for a worker’s education, cognitive skills, and personality. Since these are able to control for almost all of the genetic factors that could correlate between a mother and her child (except possibly health), we believe that the only way a mother’s education could have an effect on her child’s wages is through providing access to differently shaped social networks, thus satisfying the exclusion restriction.¹⁴ We further introduce *monthly household income* as an additional exclusion restriction to control for the fact that those from wealthier households might be able to bribe their way into jobs.

We estimate the switching model using, first a bimodal selection correction, distinguishing between workers hired formally or through social networks; we further attempt to use a multinomial selection correction, allowing for five different hiring channels (formal, family, friends, village, and political or school alumni organizations). Hausman and Small-Hsiao tests of Independence of Irrelevant Alternatives (IIA) assumption confirm that they are independent. Still, as we can only include firm characteristics and not firm fixed effects, we will place minimal emphasis on this additional distinction.¹⁵

¹⁴We also estimated the model using the formal education of the father and identifying the model by its functional form (without exclusion restriction). We opted for the formal education of the mother instead of the father because the latter was not a good predictor of hiring channel choice and an exclusion restriction is generally the preferred option. We further checked the robustness of the instrument for different subgroups (by educational attainment, levels of cognitive and non-cognitive skills), and occupational classification (blue/white collar). The instrument is robust for all subgroups except levels of educational attainment. Testing the instrument by levels of educational attainment shows that it only holds for those with medium levels of education (with the highest degree being either junior secondary school, secondary school, or high school). To us, this suggests the following within an instrumental variable framework: those with low levels of education search predominantly through social networks. The quality of the networks does not matter for them. Those with high levels of education predominantly search through formal channels, thus using mother’s education as a proxy for the quality of the network is irrelevant for them. Those affected by the instrument in our case then are those with intermediate levels of education who might still search among family networks if those are of good quality (proxied by the level of formal education of the mother).

¹⁵The estimates that we provide based on a multinomial selection model rely on the selection correction developed in Bourguignon et al. (2007). Since a multinomial model with firm fixed effects was not converging, these models will use firm characteristics instead, which are industry, firm size, the sex of the top manager, and whether the firm exports.

2.5 Results

2.5.1 Descriptive statistics

For reasons outlined previously, we restrict our sample to male workers with non-missing non-cognitive skills variables. In total, we keep 4,678 male workers and 487 firms in our sample.¹⁶ Table A2.2 presents descriptive statistics of the firms in our sample.

The data include firms from five economic sectors: commerce (wholesale, retail), education, finance, manufacturing, and public administration, with manufacturing representing a larger share of the sample due to the importance of this sector in the Bangladeshi economy. Firms in commerce, education, and finance tend to be stand-alone companies, while firms in the manufacturing and public administration sectors typically belong to a larger (parent) company. Firms in the public administration sectors are almost all part of publicly owned organizations, though the picture is more mixed in the education sector, which is comprised of a mixture of government-owned, autonomously owned, and NGO-owned companies.¹⁷ Public sector companies include a variety of economic activities, ranging from agriculture to administrative support and social work. Companies in the remaining sectors (commerce, finance, and manufacturing) are private enterprises or individually owned entities. Table A2.2 shows that the use of networks is extremely prevalent among firms: 33 percent of firms use social networks as their main channel of advertising job openings. For about half the sample (54 percent), social networks are at least one of their hiring channels.

Table A2.3 presents the characteristics of a random sample of employees within the sampled firms. Column (1) depicts means for the entire male sample. Workers surveyed are on average 32 years old. They were on average 26 years old at hiring at their current job and currently have about 6 years of work experience. The majority of workers are located in Dhaka, the capital of Bangladesh. 31 percent of workers have at most primary education, 46 percent have completed secondary education, and the remaining 23 percent have obtained tertiary education. Workers score worse on the literacy test than on the numeracy test, and they score lowest on the acquiescence-corrected personality trait hostile attribution bias and highest on conscientiousness. Most workers are skilled white or blue collar workers, which can be explained by the rather specific sample of formal firms. A slight majority of workers obtained their current job via social networks (54 percent).

¹⁶By limiting our sample to male workers only, we lose about 12 percent of the original sample of workers and by limiting it to those with at least two answer per personality trait dimension, we lose an additional 23 percent of male workers. We further exclude 13 firms who did not have any male workers with non-missing skills variables. Our resulting sample is slightly better educated, slightly more likely to work in professional occupations, and slightly less likely to work in construction or elementary occupations.

¹⁷The school system in Bangladesh consists of a heterogeneous mix of education providers, including the government, secular private sector initiatives, religious authorities, and NGOs (World Bank, 2013).

Comparing workers who found their job through formal job search methods (column (3)) and those who found their job through networks (column (5)) further reveals differences between the two groups. Workers who found their job through formal hiring channels earn higher wages on average. Figure A2.1 illustrates that this wage gap persists over different economic sectors. Formally hired workers are slightly older, are more likely to work in skilled white-collar occupations (76 percent vs. 38 percent), and are less likely to live in Dhaka. Formally hired workers are more likely to be working on permanent contracts (96 percent), though even workers hired through networks predominantly have permanent contracts (86 percent).¹⁸ They score higher on both numeracy and literacy tests, are more educated (7 percent of the formal hires have obtained at most primary education compared to 50 percent of network hires), and are less conscientious, less extroverted, and less emotionally stable.¹⁹

2.5.2 Wage returns to different types of skills

Hourly wages

We estimate OLS regressions with the natural logarithm of the current hourly wage as the dependent variable and a basic set of covariates that is gradually expanded. The first set of covariates consists of those commonly used in Mincer type regressions, explaining wages as a function of human capital. The covariates included are years of education, total work experience²⁰, and quadratic effects for these two variables, as well as a dummy variable for being married. We add measures of cognitive ability (or, rather learning outcomes, based on standardized scores on numeracy and literacy tests) and non-cognitive traits (standardized values of the Big Five personality test and the socio-emotional traits hostile attribution bias and grit). Lastly, we include firm fixed effects and dummies for occupations.²¹ By including cognitive and non-cognitive skills as well as firm fixed effects we can control for a large amount of otherwise unobserved worker and firm heterogeneity.

Results are displayed in Table 2.1. One caveat of including many controls is that we could run into a “bad controls” problem: we control for firm fixed effects and occupation, but these variables could themselves be part of an individual’s choice, given his skills. If this were the case, we would be over-controlling, especially if

¹⁸We have some missing values for the contract type variable, hence the two different sample sizes in the table.

¹⁹The survey does not contain information on whether workers within the firm are formal or informal workers. However, more than 90 percent of employment in formal firms in Bangladesh is formal employment (Asian Development Bank and Bangladesh Bureau of Statistics, 2012). We can assume that any significant results are caused by differences in hiring channels and not in formality status.

²⁰The data allow us to split total work experience into work experience prior joining the current firm and tenure at one’s current job. However, we decided against using the more disaggregated measures due to the inherent endogeneity of tenure.

²¹We include dummies for 10 different occupations: managers, professionals, technicians and associate professionals, clerical support workers, service workers, sales workers, skilled agricultural workers, construction and related workers, machine operators/drivers, and elementary occupations.

Table 2.1 – Log hourly wages regressed on cognitive and non-cognitive skills, OLS

Log hourly wages	(1)	(2)	(3)	(4)	(5)
Years of education	-0.007 (0.008)	-0.007 (0.008)	-0.014* (0.008)	-0.024*** (0.008)	-0.028*** (0.008)
Years of education (sqrd)	0.004*** (0.000)	0.004*** (0.000)	0.004*** (0.000)	0.005*** (0.000)	0.003*** (0.000)
OP (std)		0.001 (0.011)	0.003 (0.009)	0.004 (0.009)	0.001 (0.009)
CO (std)		-0.011 (0.012)	-0.019* (0.011)	-0.021* (0.012)	-0.020* (0.011)
EX (std)		0.015 (0.012)	0.001 (0.009)	0.002 (0.009)	0.005 (0.009)
AG (std)		0.003 (0.009)	0.013* (0.007)	0.014** (0.007)	0.011* (0.007)
ES (std)		0.004 (0.013)	0.007 (0.011)	0.007 (0.011)	0.008 (0.010)
HAB (std)		0.003 (0.009)	-0.010 (0.008)	-0.010 (0.008)	-0.013* (0.007)
GR (std)		0.005 (0.011)	0.005 (0.009)	0.004 (0.009)	0.001 (0.009)
Reading score				0.040*** (0.013)	0.034** (0.013)
Numeracy score				0.005 (0.013)	0.005 (0.012)
Indvl controls	YES	YES	YES	YES	YES
Firm fixed effects			YES	YES	YES
Occupation dummies					YES
N	4,678	4,678	4,678	4,678	4,678
R ²	0.459	0.459	0.433	0.461	0.485

Notes: Standard errors clustered at the firm-level in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Individual controls are a dummy for being married, total work experience and total work experienced squared. OP = openness to experience; CO = conscientiousness; EX = extraversion; AG = agreeableness; ES = emotional stability; HAB = hostile attribution bias; GR = grit; Personality traits are acquiescence corrected. Numeracy and literacy scores are standardized.

particular firms attract particular workers and pay differential wages. We include covariates successively, and comparing results with and without firm fixed effects and occupation dummies shows that including firm fixed effects increases the size and significance of some of our covariates. This leads us to conclude that, indeed, particular firms might attract particular workers and means that column (5) of Table 2.1 reports our preferred estimates. This simple model is able to explain 49 percent of the variation in hourly current wages.

In accordance with literature from developing countries (e.g. Kuépié et al., 2009; Söderbom et al., 2006), we find convex returns to education; this means that returns to education increase with the level of education attained.²² To illustrate the non-linearity of returns to education, we tested a specification with education dummies instead of a continuous education variable. Returns to education are insignificant at low levels of education (i.e. there are no significant

²²We also estimated a basic model with a linear education term. We find a wage return of 6.9 percent to an additional year of education, which is in line with other studies for Bangladesh, such as Asadullah (2006), who uses national household survey data from 1999-2000 and finds returns of 6.2 percent for men. However, as our squared years of education term is highly significant, we will continue with a non-linear specification of education.

returns to having completed primary education compared to having no schooling or incomplete primary education), but returns are significant and increasing with the level of education obtained, with returns highest for those who have obtained a post-graduate degree.²³ This pattern suggests low-quality primary education in Bangladesh but could also be due to a different reward structure to educational attainment in the formal sector (Kuépié et al., 2009).

Cognitive skills have a positive effect on wages, even when controlling for educational attainment. Returns are rather low, however: column (5) shows that a one standard deviation increase in the literacy score increases wages by only 3.4 percent. Estimates for developed countries range between 3.8 percent in Sweden to 20 percent in the United States (Hanushek and Zhang, 2009), but these estimates look at annual earnings and not wages; estimates for developing countries might be even larger, though results seem to depend on the estimation strategy and sample (Hanushek and Woessmann, 2008).

In terms of non-cognitive skills, we find several significant covariates: a one standard deviation increase in conscientiousness is associated with a 2 percent decreases in log hourly wages; a similar increase in hostile attribution bias is associated with a 1 percent decrease in wages. An increase in agreeableness is associated with a positive wage premium, however. These findings only appear after we include firm fixed effects in column (3), and the significant coefficient for hostile attribution bias only shows once we also include occupation dummies. This suggests that firm specific wage setting, rewarding particular skills differently (and doing so by occupation, in the case of hostile attribution bias), is driving our results. This is intuitive, as one could imagine a firm in the manufacturing sector to not care much about a workers' agreeableness if that person works in construction, while this could be particularly important for somebody working in customer care in a firm in commerce.

As a robustness check, we also constructed non-cognitive skills variables using exploratory factor analysis instead of relying on pre-defined definitions. Three factors can be identified in the data. The first is made up of mostly items from the extraversion and emotional stability scale; the second relates to work ethic (conscientiousness and grit); the third captures openness to experience and agreeableness. Building non-cognitive skills from exploratory factor analysis instead of pre-defined dimensions reduces our sample size to 3,102 observations, as not every worker completed the full non-cognitive skills questionnaire. Using these factors in the OLS regressions (shown in columns (1) to (3) of Table A2.4) illustrates that none of the factors are significant in the OLS wage regressions, though the coefficients carry the same signs as in Table 2.1 (i.e. positive sign for factor 1 and a negative sign for factor 2). Due to the drop in sample size, we will rely on the separate Big Five dimensions for the remainder of the paper and refer to the

²³Table not shown but available upon request.

constructs from factor analysis as robustness checks.

An additional potential caveat of including education, learning outcomes (literacy, numeracy), and non-cognitive skills is that we potentially introduce multicollinearity into our estimations. This is rather obvious for the case of education and literacy and numeracy, given that the latter are usually developed in the educational system. It is, however, also possibly the case for the non-cognitive skills variables, since they have been shown to influence schooling outcomes (educational attainment) and learning (through effort applied, for example) (e.g. Cunha and Heckman, 2008; Heckman and Rubinstein, 2001). We therefore test for multicollinearity in the OLS regressions by looking at the variance inflation factor (VIF). The VIF is the ratio of the variance of a multiple covariate model to the variance of a single covariate model. As a rule of thumb, a VIF larger than 10 is seen as a sign of multicollinearity (see e.g. Hair et al., 1998). Table A2.5 displays the variance inflation factors per covariate included for different OLS specifications. According to this rule, the VIFs for education and education squared are too high in the very basic Mincer model displayed in column (1). Comparing a basic Mincer model (column (1)) to a slightly augmented one including non-cognitive skills (column (2)) shows that including additional covariates does not change the VIF of those in the basic model. Further, the VIF for the included non-cognitive skills are all rather low (ranging from 1.1 to 2.1). Including learning outcomes in column (3) substantially increases the VIF for years of schooling from 13.3 to 17.1 but leaves the non-cognitive skills unaffected. Thus, the problematic variable in terms of multicollinearity is education. Indeed, column (4) without the education variables shows that the VIF for every covariate is below 10, and the value of literacy's VIF drops. This illustrates that multicollinearity does not seem to be a problem when it comes to education and non-cognitive skills, or non-cognitive skills and learning outcomes²⁴, but multicollinearity is present when including both years of schooling and literacy. We thus estimate our preferred OLS regression from Tables 2.1 without years of education (and its square) as covariates. OLS results are displayed in column (1) of Table A2.6 and show that the coefficients for the non-cognitive skills remain unchanged. The coefficient for literacy more than doubles in size, though, from 0.034 with education covariates to 0.088 without. The collinearity between education and literacy could occur for a number of reasons. For example, both literacy and years of schooling are outcomes of cognitive (intelligence) and non-cognitive skills. If this were the driving force behind the collinearity though, one would also expect numeracy to be highly collinear (which does not react to whether or not education is included) and for both literacy and education to be sensitive to the inclusion of the non-cognitive skills. Since this is not the case, it seems more likely that the ability to read and write could act as a signal for the

²⁴Indeed, additional models, not shown, demonstrate that starting the models with cognitive skills and subsequently adding non-cognitive skills leaves the VIF of cognitive skills basically unchanged.

quality of education received, particularly in an environment where educational quality is diverse, which is why we decide to keep both literacy and education in our main regressions.

Earnings

While most of the paper is based on hourly wages, this section provides a descriptions of returns to total individual earnings. This is useful as one component of the returns to skills might be differences in labor supply. Table A2.7 depicts the returns to total current earnings. Results are broadly comparable to those in Table 2.1, with the only difference being that hostile attribution bias is no longer significant when looking at current earnings. Indeed, looking at the determinants of hours worked (column (6)) shows that hostile attribution bias is the only significant determinant of hours worked that we can find, with those being more hostile towards others working more hours. This suggests sorting of people with high hostile attribution bias (who thus think that others around them might be hostile towards them) into jobs with a high hourly requirement, but not necessarily high pay. Indeed, those with higher hostile attribution bias can mostly be found in unskilled occupations. This supplemental analysis suggests some potential labor supply responses; for comparability reasons, we will be focusing on hourly wages for the remainder of this paper.

2.5.3 Choice of hiring channel

The OLS wage estimates are likely to be biased if the skills variables we are looking at (cognitive and non-cognitive) predict the hiring channel a worker chooses. Table 2.2 presents a probit model in which the dependent variable is the probability of having been hired through social networks. This model is the first equation of the selection model (equation 2.5) that will be developed in subsequent sections. We start from a basic model with only the identifying variables (mother's education and monthly household income), education, and non-cognitive skills, and subsequently add cognitive skills, occupational dummies, and firm fixed effects. Moving from a model without fixed effects (column (2)) to a model with fixed effects (column (4)) reduces our sample size, as the fixed effect model relies on at least one formal and one network-hired worker per firm for identification. This is not the case for 229 out of 487 firms in our sample and shows that not all firms use both hiring channels. The single-hiring-channel firms are mostly from education and public administration for those who only hire through the formal channel and manufacturing for those who only hire through networks. Keeping only firms that have more than one hiring channel is essential as we are interested in why firms would hire workers through one channel or the other. However, this means, on aggregate, that in our preferred specifications with firm fixed effects our sample has slightly fewer manufacturing firms compared to the original sample.

Results in Table 2.2 show that after including firm fixed effects, non-cognitive skills do not have any significant relation with hiring channels. Contrary to our hypothesis, being more agreeable, for example, does not significantly increase one's probability of having been hired through networks. In addition to our selection identification variables, the only significant predictors of having been hired through networks are years of schooling and the reading score. For both covariates, the effect goes in the same direction: an additional year of schooling and a better reading score both reduce the probability of having been hired through networks. On the one hand, this result could be due to individuals without literacy skills being unable to access formal means of job search (such as newspaper ads). On the other hand, literacy could act as a signal for having received a quality education. If it is difficult for employers to assess the real quality of a degree received, requiring workers (in formal job ads) to be literate could work as a proxy for quality.

So far, we have combined all types of social networks into one category. However, the type of social network activated might matter for wages. We can distinguish between one formal hiring channel and four different types of social networks (family, friends, village connections, and political and school alumni associations). The number of observations within each channel is balanced, except for the last category.²⁵ Due to the small number of observations in the politics and school category and based on the suggestions of a Wald test for combining categories, we will combine it with the friends category. Table A2.8 depicts the result and demonstrates the relevance of both of our identifying variables. Crucially, the estimates are based on firm characteristics (industry, size, whether the firm exports, and gender of the CEO) instead of firm fixed effects, as models with the latter would not converge. As we are therefore not able to fully control for firm heterogeneity in the hiring decision, the results presented in this section should be handled with caution.

Results are similar compared to the binary distinction between formal hiring channels and networks only in the sense that we find that the literacy score increases a worker's probability of having been hired through formal channels and decreases his probability of having been hired through family or through friends. It is insignificant for the village hiring channel. Looking at the non-cognitive skills now also reveals significant results, though, they differ by channel: emotional stability is positive and significant among those hired through family connections, as is grit, but the latter decreases one's chances of being hired through friends. Further, hostile attribution bias is negatively related to being hired through village connections. This suggests that there could be a role for non-cognitive skills in affected hiring channels; still, given that we only control for firm characteristics and not firm fixed effects, we place less emphasis on this result.

²⁵960 workers found their job through family members, 1,012 through friends, and 489 through village connections, but only 76 workers found their job through political and school alumni associations.

Table 2.2 – Probability of being hired through social networks: marginal effects after probit

Hired through networks	(1)	(2)	(3)	(4)
Years of education	-0.078*** (0.012)	-0.050*** (0.012)	-0.059*** (0.016)	-0.040** (0.017)
Years of education (sqrd)	0.000 (0.001)	-0.001 (0.001)	-0.000 (0.001)	-0.001 (0.001)
OP (std)	-0.005 (0.013)	-0.003 (0.012)	0.003 (0.019)	0.003 (0.019)
CO (std)	0.030* (0.016)	0.028* (0.016)	0.007 (0.021)	0.010 (0.022)
EX (std)	-0.005 (0.015)	-0.006 (0.014)	-0.015 (0.021)	-0.017 (0.020)
AG (std)	0.007 (0.011)	0.004 (0.011)	0.015 (0.015)	0.013 (0.015)
ES (std)	0.014 (0.015)	0.016 (0.014)	0.028 (0.019)	0.026 (0.019)
HAB (std)	-0.011 (0.011)	-0.009 (0.011)	-0.013 (0.015)	-0.013 (0.015)
GR (std)	-0.011 (0.013)	-0.011 (0.012)	-0.013 (0.017)	-0.011 (0.018)
Reading score		-0.071*** (0.019)		-0.057** (0.026)
Numeracy score		0.011 (0.017)		0.004 (0.025)
Mother no formal educ	0.127*** (0.026)	0.120*** (0.025)	0.134*** (0.034)	0.122*** (0.034)
HH inc: 5,000-7,500tk	-0.063 (0.066)	-0.065 (0.065)	-0.161* (0.092)	-0.141 (0.094)
HH inc: 7,501-10,000tk	-0.072 (0.066)	-0.095 (0.067)	-0.184** (0.093)	-0.163* (0.094)
HH inc: 10,001-15,000tk	-0.042 (0.067)	-0.080 (0.068)	-0.234** (0.095)	-0.200** (0.096)
HH inc: 15,001-20,000tk	0.006 (0.070)	-0.044 (0.072)	-0.203** (0.100)	-0.165 (0.102)
HH inc: 20,001-30,000tk	-0.079 (0.076)	-0.175** (0.077)	-0.302*** (0.108)	-0.275** (0.111)
HH inc: 30,001-50,000tk	0.088 (0.092)	-0.012 (0.098)	-0.205 (0.127)	-0.182 (0.131)
HH inc: 50,001tk and above	0.091 (0.164)	-0.024 (0.170)	-0.085 (0.229)	-0.070 (0.237)
Indvl controls	YES	YES	YES	YES
Occupation dummies	YES	YES	YES	YES
Firm fixed effects			YES	YES
N	4,655	4,655	3,110	3,110
Pseudo R^2	0.287	0.328	0.357	0.361

Notes: Standard errors clustered at the firm-level in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Individual controls are a dummy for being married, work experience and work experienced squared. OP = openness to experience; CO = conscientiousness; EX = extraversion; AG = agreeableness; ES = emotional stability; HAB = hostile attribution bias; GR = grit; Personality traits are acquiescence corrected. Literacy and numeracy scores are standardized. Marginal effects reported are at the means of covariates. Our sample size is reduced to 4,655 individuals instead of 4,678 due to missing values in the household income variable.

2.5.4 Skills and endogenous hiring channel

The results from the previous section illustrate that cognitive skills influence the selection of hiring channel, while less evidence is found for non-cognitive skills, as when we distinguish between the different Big Five traits, none of them is

statistically significant. Still, to account for the selection into hiring channel based on literacy skills, the following section presents the switching model. Table 2.3 depicts the results, modeling selection as a bimodal choice between formal channels and networks. Columns (1) and (2) of Table 2.3 depict the wage regression for the formal hiring channel, equation (2.2), and columns (4) and (5) the wage regression for the network hiring channel, equation (2.3), both while first taking into account selection into hiring channel. Columns (3) and (6) present OLS estimations for comparison.

Table 2.3 – Returns to skills using an endogenous switching model

Log hourly wages	Formal hiring			Network hiring		
	Selec Corr (1)	Selec Corr (2)	OLS (3)	Selec Corr (4)	Selec Corr (5)	OLS (6)
Years of education	-0.041*** (0.015)	-0.051*** (0.016)	-0.028 (0.019)	-0.021*** (0.005)	-0.028*** (0.006)	-0.030*** (0.009)
Years of education (sqrd)	0.005*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.004*** (0.000)	0.003*** (0.000)	0.004*** (0.001)
OP (std)	0.027** (0.011)	0.025** (0.010)	0.024* (0.013)	-0.010 (0.009)	-0.013 (0.008)	-0.013 (0.012)
CO (std)	-0.025* (0.014)	-0.023* (0.014)	-0.028 (0.018)	-0.017 (0.011)	-0.017 (0.011)	-0.019 (0.015)
EX (std)	0.005 (0.014)	0.008 (0.013)	0.013 (0.015)	0.003 (0.010)	0.001 (0.010)	0.003 (0.012)
AG (std)	0.017 (0.011)	0.011 (0.010)	0.009 (0.011)	0.012 (0.008)	0.011 (0.008)	0.010 (0.010)
ES (std)	0.002 (0.014)	0.004 (0.013)	0.000 (0.016)	0.017 (0.011)	0.019* (0.010)	0.017 (0.013)
HAB (std)	-0.022** (0.010)	-0.023** (0.010)	-0.020 (0.013)	-0.009 (0.007)	-0.013* (0.007)	-0.012 (0.010)
GR (std)	-0.005 (0.011)	-0.011 (0.011)	-0.007 (0.015)	0.006 (0.009)	0.003 (0.008)	0.003 (0.011)
Reading score		0.026 (0.019)	0.049* (0.026)		-0.005 (0.012)	0.001 (0.015)
Numeracy score		0.001 (0.015)	-0.004 (0.024)		0.012 (0.009)	0.012 (0.014)
Indvl controls	YES	YES	YES	YES	YES	YES
Firm fixed effects	YES	YES	YES	YES	YES	YES
Occupation dummies		YES	YES		YES	YES
Inverse Mills ratio	-0.267*** (0.055)	-0.250*** (0.052)		0.114** (0.049)	0.150*** (0.048)	
N	4,655	4,655	2,127	4,655	4,655	2,528
R ²			0.687			0.672

Notes: Standard errors clustered at the firm-level in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Individual controls are a dummy for being married, work experience and work experienced squared. OP = openness to experience; CO = conscientiousness; EX = extraversion; AG = agreeableness; ES = emotional stability; HAB = hostile attribution bias; GR = grit; Personality traits are acquiescence corrected. Instruments used for identification: mother has no formal education and monthly household income. Estimated using a two-step model. Our sample size is reduced to 4,655 individuals instead of 4,678 due to missing values in the household income variable.

Results show that selection is indeed present; both inverse mills ratios (IMR) are significant. The IMR for the formal hiring channel is negative, meaning that without the selection correction, the estimates would have been biased downwards. Returns estimated from the OLS are thus likely to be lower than the true returns once we take selection into account. The opposite holds true for the IMR for

the network hiring channel. Looking at selection corrected estimates, we find that years of education are convex, increasing, and significant in both channels (formal and networks), though at the average number of years of education (12.8 in the formal channel, 7.2 in the network channel) returns are negative – but effectively equal to zero in the network channel. Literacy skills, which were highly significant in the OLS regressions and in the selection equation, do not show any significant association with wages in either the formal hiring channel or the network hiring channel once the initial selection into type of hiring channel has been controlled for. Numeracy is not significant either. Cognitive skills thereby act only by enabling workers to access the formal hiring channel; once a worker has been hired through the formal channel, the role of literacy vanishes. Thus, the positive literacy coefficient in the OLS regressions (Table 2.1) simply captures a wage premium for having been hired through formal channels.

Non-cognitive skills, none of which were significant in the selection equation when looking at binary selection (formal vs. networks), are significant in the wage regressions. For those hired through formal channels as well as those hired through networks, hostile attribution bias decreases wages. Hostile attribution bias refers to a cognitive bias according to which individuals interpret others' action towards them as hostile. In a work setting, such an individual is likely to be less effective in any setting that involves other workers, such as teamwork or communication with customers or suppliers. A negative effect of hostile attribution bias is thus expected (except for maybe management positions, where adverse behavior and a suspicion towards others could feasibly lead to more success). The negative effect of conscientiousness on wages that we found in the combined OLS regressions in Table 2.1 is in fact driven by those in the formal hiring channel. As mentioned previously, hiring channel does coincide with occupational selection. As such, those hired through formal channels are more likely to work in white-collar occupations, in which conscientiousness and a high level of attention to detail (doing work very thoroughly) might not be strictly required. Instead, traits that could be more important might be those related to creativity and coming up with new ideas. Indeed, the trait correlating with that sort of behavior, openness to experience, is highly significant for those hired through formal channels (and even shows up in the formal-channel only OLS regression in column 3). Looking at those hired through networks, emotional stability is the only trait that exhibits a positive and significant correlation with wages. While the trait is only significant at the 10 percent level, it could relate to workers being better able to remain balanced throughout the work day, to be tolerant of others' behaviors, to perform tasks well, and to not be stressed easily. This then also captures dimensions of agreeableness, which is no longer significant in either hiring channel.

Columns (3) and (6) provide results from separate OLS regressions by hiring channel for direct comparison with the selection-corrected returns. For formal

channels, the coefficients for most non-cognitive skills are similar in size but increase in significance when taking into account selection. This is in line with expectations, given the negative IMR. The coefficient for years of education almost doubles in size and becomes highly statistically significant when taking into account selection, while the coefficient for reading score is halved and loses its significance. For network hiring, a similar effect is not visible and coefficients for the selection-corrected estimations in column (5) and basic OLS in column (6) are very similar in terms of coefficient size. The selection corrected standard errors are slightly smaller, meaning that some of the coefficients gain significance. Still, it seems that selection into formal channel is more prevalent than selection into network hiring. This suggests that network hiring could be a secondary option when looking for work, or that it is the only option available for some low-skilled workers. In sum, the returns to different personality traits seem to mirror the types of tasks that individuals hired through each channel perform, which are more likely (though not exclusively so) higher skilled, more white-collar occupations for those hired through formal channels, and low- or unskilled (possibly manual or menial tasks) for those hired through networks.

The different rewards to skills by hiring channel give rise to the notion that firms might hire workers through different channels depending on their occupation (i.e. blue-collar workers being hired predominantly through social networks, with white-collar workers hired largely through formal channels). The data do not fully discount this claim: about 40 percent of workers in the sample work in blue-collar jobs and 78 percent of them were hired through networks (compared to only 39 percent of workers in white-collar occupations). We illustrate differential returns by broad type of occupation (blue- and white-collar), by introducing interactions between non-cognitive skills and being a white collar worker (Table A2.9).

For network hires, all interaction terms except for grit are insignificant. Being gritty and a white-collar worker actually has a surprising negative return for those hired through networks. This is unexpected and does not only affect a small number of people, as about 40 percent of those hired through networks are white-collar workers. It could suggest that what is wanted from those workers is creativity (relating back to the positive effect of openness in formal channels in Table 2.3, which is also positive here, but insignificant), or that other traits such as obedience and listening to one's superiors are more important than persistence and task completion. Among formal channels, being a white-collar worker is only significantly associated with emotional stability. Conscientiousness is still negative but no longer significant. Table A2.9 thus suggests that there might be differential returns depending on broad occupational type, though the evidence is not particularly conclusive. One has to keep in mind that among those hired through formal channels, only about 20 percent of the sample are not in white-collar jobs and are thus a potentially non-representative comparison group.

In addition to the bimodal selection model, we estimate multinomial models with four different states: one formal hiring channel and three informal hiring channels (family, friends, and village connections). Table A2.10 depicts wage returns to skills, taking into account this multinomial selection. Few of the ρ , the selection coefficients, are significant, which means that our model attempts to correct for selection which might not even be present. This could be driven by us being unable to fully account for firm heterogeneity in hiring decisions, as we only include firm characteristics. Given that, and a different selection correction using Bourguignon et al. (2007), results in Table A2.10 are not necessarily comparable to those obtained earlier. Differences in returns to skills by more disaggregated hiring channel, distinguishing between the different networks used, are visible, but only one trait (emotional stability) is significant in only one channel (friends). These results show that there could be differences depending on which network was activated to find the job, but, they are at best suggestive and certainly not conclusive.

2.5.5 Starting wages and wage growth

In addition to looking at current wages, our data allow for an analysis of starting wages and thus also wage growth (the difference between current and starting wages). This is especially interesting when it comes to wage returns to skills, as employers cannot fully observe a worker’s skillset at the time of hiring; instead, skills and productivity are only fully revealed over time (“employer learning”). Employers are assumed to use school attainment and other observable signals to predict productivity and set starting wages accordingly. Over time, wages are then adjusted to match observed productivity. Similarly, the different hiring channels offer different opportunities for employers to obtain information about a potential employee prior to hiring – either through information obtained in the formal hiring process (such as personal interviews or assessment centers) or by means of information received through networks (such as who recommended the applicant or a personal assessment of the applicant by the recommending person). All of our variables are measured at the same point in time, the present. However, some of them can be assumed to be constant over time and thus be important for wage setting. This holds especially true for levels of formal education (assuming that most workers do not engage in further formal education once they are employed and would only benefit from on-the-job-training), but it also most likely holds for non-cognitive skills. Indeed, non-cognitive skills are assumed to be rather stable throughout adult life and have been shown to only react slightly to major life events (Cobb-Clark and Tan, 2011). As the mean age at hiring in our sample is 26 years, we assume that while the survey (and its skills assessment) was conducted after the worker had started his current job, skills have not changed significantly between hiring and the time of assessment. For literacy and numeracy skills, the

assumption seems to be a little more dubious: somebody who uses these skills a lot as part of his workday might improve on these measures, while somebody who does not use them might lose proficiency.

Table 2.4 presents the switching model for starting wages and wage growth, as well as separate OLS regressions for comparison. We identify the selection equation only via the educational attainment of the mother, as household income is only available at present time and could have been different when workers began their jobs.

Table 2.4 – Returns to skills using an endogenous switching model - Starting and current wage

	Starting wage				Wage growth			
	Formal		Network		Formal		Network	
	OLS (1)	SelCor (2)	OLS (3)	SelCor (4)	OLS (5)	SelCor (6)	OLS (7)	SelCor (8)
Years of education	-0.014 (0.028)	-0.027 (0.024)	-0.020 (0.012)	-0.019** (0.008)	0.008 (0.008)	0.011* (0.005)	0.001 (0.002)	0.001 (0.001)
Years of education (sqrd)	0.003** (0.001)	0.003*** (0.001)	0.004*** (0.001)	0.003*** (0.001)	-0.000 (0.000)	-0.000* (0.000)	-0.000 (0.000)	-0.000 (0.000)
Reading score	0.007 (0.031)	-0.004 (0.028)	0.014 (0.024)	0.013 (0.016)	-0.002 (0.006)	-0.001 (0.006)	0.002 (0.004)	0.003 (0.003)
Numeracy score	0.010 (0.030)	0.012 (0.023)	0.012 (0.018)	0.011 (0.013)	-0.006 (0.005)	-0.007 (0.005)	-0.001 (0.003)	-0.001 (0.002)
OP (std)	0.024 (0.019)	0.025* (0.015)	-0.022* (0.013)	-0.021* (0.011)	0.006 (0.004)	0.005 (0.003)	0.002 (0.002)	0.001 (0.002)
CO (std)	-0.040* (0.024)	-0.039* (0.020)	-0.032 (0.021)	-0.031** (0.015)	-0.009* (0.005)	-0.009* (0.004)	0.005 (0.003)	0.005* (0.003)
EX (std)	-0.001 (0.028)	-0.001 (0.019)	0.016 (0.016)	0.015 (0.013)	-0.005 (0.006)	-0.005 (0.004)	-0.004 (0.003)	-0.004 (0.002)
AG (std)	0.025 (0.019)	0.028* (0.015)	0.012 (0.014)	0.013 (0.011)	-0.002 (0.006)	-0.002 (0.003)	-0.001 (0.002)	-0.001 (0.002)
ES (std)	0.026 (0.029)	0.024 (0.020)	0.027 (0.018)	0.030** (0.014)	0.005 (0.004)	0.005 (0.004)	-0.002 (0.003)	-0.002 (0.002)
HAB (std)	-0.010 (0.019)	-0.012 (0.014)	-0.008 (0.011)	-0.011 (0.010)	-0.003 (0.003)	-0.002 (0.003)	0.000 (0.002)	0.001 (0.002)
GR (std)	0.019 (0.020)	0.015 (0.016)	-0.006 (0.016)	-0.007 (0.012)	-0.001 (0.004)	-0.001 (0.004)	0.004* (0.002)	0.004* (0.002)
Occup dummies	YES	YES	YES	YES	YES	YES	YES	YES
Firm fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Indvl controls	YES	YES	YES	YES	YES	YES	YES	YES
Inverse Mills ratio		-0.113 (0.083)		0.055 (0.068)		0.020 (0.019)		-0.014 (0.012)
N	2,127	4,678	2,528	4,678	2,127	4,678	2,528	4,678
R ²	0.541		0.550		0.205		0.313	

Notes: Standard errors clustered at the firm-level in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Individual controls are a dummy for being married, prior work experience, prior work experienced squared, and age at entering the firm. OP = openness to experience; CO = conscientiousness; EX = extraversion; AG = agreeableness; ES = emotional stability; HAB = hostile attribution bias; GR = grit; Personality traits are acquiescence corrected. Numeracy and literacy scores are standardized. Instrument used for identification: mother has no formal education. Estimated using a two-step model, showing the second stage of that model. SelCor stands for selection corrected estimates.

Looking at the starting wages regressions reveals similar results as present wages: openness to experience has a positive and significant return for those hired through formal channels, conscientiousness is once again negative, and agreeableness is positive (and significant this time). Hostile attribution bias is no longer significant,

though it remains negative. All of these are traits that can arguably be identified during a formal interview; the employer could observe if the worker is open to new ideas and thinks outside the box (openness to experience), if he is organized or not (conscientiousness), and whether or not he is friendly (agreeableness). Hostile attribution bias might not show during an interview (or the worker might attempt to hide it), which likely is why it is not punished at initial wage setting. However, once a worker is observed on the job, this trait could more readily show and therefore influence future wage development. It thus seems that employers do reward individuals for traits they deem desirable and can observe during worker selection.

Among network hires, openness to experience, and conscientiousness are punished during initial wage setting and emotional stability is rewarded. When workers are hired through networks, the only information that the employer has is through information they can gather from the network. Employers could hire through networks because it allows them to select unobservably good workers, as networks might have information that employers could not observe. They could also hire through networks because it allows them to mitigate a moral hazard problem, punishing referral providers if the referred worker does not perform well. Heath (2018) finds evidence of the latter mechanism at play in Bangladeshi garment factories. Our data do not allow us to distinguish between both mechanisms (in fact, we do not even know if the job was obtained as a referral, or if networks simply provided information about a job opening). Still, given a comparable setting, if the moral hazard mechanism was at play, workers that would be referred could arguably be those who are more emotionally stable – and therefore less likely to suffer from stress and anxiety and to be more productive as a result (Drago, 2011). This could also explain why workers do not seem to benefit from higher openness to experience, which refers to workers who are curious and prefer new experiences to routines.

Taking wage growth instead of current wages or starting wages as the dependent variable reveals a small but significantly positive effects of grit and conscientiousness for hired through networks and a small but significantly negative effect for those hired through formal channels. Among the latter group, this again reflects that employers probably value creativity and coming up with new ideas more than simply executing tasks on time. In the wage growth equations, individual effects that are constant over time should have zero effect on wage growth, due to differentiation between current and starting wage equations which they enter in similar ways. As we observe a positive and significant effect of conscientiousness and grit among network hires, one possible explanation is that these traits evolve over time as workers improve their performance and adapt closer to work requirements. Lastly, the inverse mills ratios are insignificant for both, the network and the formal hiring channel (though with the same sign), meaning that estimates from OLS and after

correcting for selection are very similar. This could be driven in part by our ability to include only one instrument to identify selection instead of two and thus our consequent inability to fully account for a worker’s decision to engage in formal or network job search. At the same time, starting wages could also simply be measured with more error, as workers do not remember them accurately.

2.5.6 Determinants of the within-firm wage gap between formal and network hires

Finally, we aim to understand whether employer biases and preferences can help explain part of the wage gap between formal and network hires, exploring the employer side of this matched survey. Results for this section are based on a restricted sample of firms that have at least two formal and two informal hires, leaving us with 171 firms – about two thirds of the sample used in previous sections. Table A2.11 presents characteristics of the firms included in this restricted sample in comparison to the sample of firms used in previous sections. Compared to the switching model section, our firms in this section are slightly more likely to be in commerce and manufacturing and less likely to be in public administration. Firms in the reduced sample are, on average, bigger than firms in the original sample, which is to be expected when we impose the restriction of having at least two formal and two network hires per firm. Few differences between workers can be observed apart from a slightly better educational attainment among workers in the reduced sample.

The previous sections have shown that returns to skills are affected by the hiring channel through which workers found their jobs. The differential returns to the same skill type varying by hiring sector leads to wage gaps between similar workers within the same firms. This could in fact reflect preferences of firms for certain skill sets among certain types of hires. To the degree to which employers value some skills more and have underlying assumptions or information about the availability of these skills in different hiring channels (formal/networks), this could affect the wage premium paid. We employ a hierarchical modeling approach (Bryk and Raudenbush, 1992; Meng, 2004; Nordman et al., 2015; Nordman and Wolff, 2009) to capture the determinants of this within-firm formal-network hiring channel wage gap. Decomposing this wage gap allows us to better understand which employer biases or preferences predict a larger or smaller gap.

As before, we include firm fixed effects in our worker-level wage equations for formal and network hires and estimate those as follows:

$$\ln w_{ij}^F = \beta_0^F + \beta_1^F A_i + \beta_2^F Cog_i + \beta_3^F NonCog_i + \delta_j^F + \epsilon_{ij}^F \quad (2.6)$$

$$\ln w_{ij}^N = \beta_0^N + \beta_1^N A_i + \beta_2^N Cog_i + \beta_3^N NonCog_i + \delta_j^N + \epsilon_{ij}^N \quad (2.7)$$

In the regression analysis, we control for a worker’s demographic characteristics,

A , his cognitive skills Cog and non-cognitive skills $NonCog$. Due to these controls, the effect of the firm fixed effect, δ , then simply reflects a premium paid by the firm to its employees beyond their observable characteristics. The difference between δ_j^F and δ_j^N can thus be interpreted as an estimate of the within-firm wage premium or penalty for having been hired through formal channels. We then use OLS regressions with the difference in firm fixed effects as the dependent variable to estimate the effect of firm level characteristics on the size of the within-firm wage gap.

$$\widehat{\delta}_j^F - \widehat{\delta}_j^N = \beta_0 + \beta_1 C_j + \beta_2 SkillsEmployees_j + \beta_3 SkillsHiring_j + \epsilon_j \quad (2.8)$$

where C_j is a vector of firm characteristics at the firm level, including industry, firm size, whether the firm exports, whether it provides on-the-job training, and the sex and education of the top manager. Respondents of the employer survey (a high ranking manager of the firm) were further asked to judge the importance of a battery of skills and values and the importance of a number of skills in the hiring decision among two types of workers in their firm, namely professional workers (managers, professionals, technicians) and non-professionals (such as clerical support workers, construction, and elementary occupations). The employers were thus asked not to judge an individual employee but a broad group of occupations. Employers were asked (on a scale of 1-10) how important they think it is for their employees to have the following skills: communication, teamwork, problem solving, literacy, numeracy, costumer care, responsibility, motivation, and creativity, as well as general and advanced vocational job-specific skills. They were also asked (on a scale of 1-10) how important the following criteria are for their hiring decision: academic performance, work experience, skill set, interview, informal network/recommendation, and political affiliation. Standardized values of these skills variables are included in the regression (vectors $SkillsEmployees_j$ and $SkillsHiring_j$, respectively). As the questions were asked separately for both broad occupational types, we also run separate regressions and, in line with the distinction between white- and blue-collar workers earlier in this paper, do not expect to obtain the same results for each professional group.

We normalize the within-firm wage gap to be bound in the (0,1] interval: $wg_j = e^{-(\widehat{gap}_{max} - \widehat{gap}_j)}$ where \widehat{gap}_{max} is the sample maximum of the estimated within-firm wage gap and \widehat{gap}_j is the estimated wage gap for formal and network hires within-firm j . The sample mean is low at 0.11, as the distribution is quite skewed. We drop the five most unequal firms, which leaves us with a sample mean of 0.48. Figure A2.2 shows that the within-firm wage gap is approximately normally distributed (panel a) and that it is not concentrated in a single industry (panel b).

Table 2.5 displays the results of equation 2.8, separately for professional and

non-professional workers. A positive coefficient refers to an increased within-firm wage gap between formal and network hires (i.e. formal hires enjoying a wage premium beyond their observable characteristics); a negative coefficient symbolizes the opposite. Looking at the regression for professional workers, an employer placing more value on communication skills, correlates with a larger wage gap between formal and network hires. An employer placing more value on teamwork skills, is associated with a smaller wage gap, meaning that workers hired through formal channels and networks earn wages that are comparatively equal. No significant correlation is visible for non-professional workers in column (2).

Table 2.5 – Determinants of the within-firm hiring channel wage gap

Normalized within-firm wage gap	Professional worker (1)	Non-professional worker (2)
<i>Importance of skills among employees</i>		
Communication skills	0.032* (0.018)	-0.003 (0.020)
Team work skills	-0.052*** (0.020)	0.002 (0.024)
Problem solving skills	0.008 (0.021)	-0.034 (0.023)
Literacy skills	-0.007 (0.022)	-0.000 (0.019)
Numeracy skills	0.009 (0.021)	0.005 (0.018)
Customer care skills	-0.007 (0.017)	-0.016 (0.018)
Responsibility	-0.023 (0.020)	-0.018 (0.026)
Motivation	0.006 (0.022)	0.026 (0.028)
Creativity	0.007 (0.025)	0.025 (0.023)
Vocational job-specific skills	-0.012 (0.019)	-0.034 (0.023)
<i>Importance in hiring decision</i>		
Academic performance	0.008 (0.018)	0.001 (0.016)
Work experience	0.024 (0.016)	0.019 (0.019)
Skill set	-0.008 (0.019)	0.017 (0.020)
Interview	-0.004 (0.021)	-0.002 (0.020)
Network/Recommendation	-0.019 (0.016)	-0.013 (0.014)
Political affiliation	0.022 (0.018)	0.033 (0.059)
Firm characteristics	YES	YES
N	171	171
R^2	0.239	0.240

Notes: Standard errors are bootstrapped (500reps); *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Firm characteristics include industry, firm size, a dummy for exporting firms, whether the firm has performance reviews, whether it has bonuses based on firm performance, or individual performance, whether it requires interviews for vacancies, a dummy for the main job advertising channel being networks, whether the firm provides on the job training, the proportion of the top management being women, the gender of the manager answering the questionnaire and their educational attainment.

From a theoretical point of view, employers could choose to hire through networks instead of formal channels for two main reasons: networks could provide information about potential workers that would otherwise be unobservable, and hiring based on referrals could reduce moral hazard, as referees can be punished as well if referred workers do not perform according to expectations. Heath (2018) has shown that the latter mechanism is at play among Bangladeshi garment workers, and section 2.5.5 has demonstrated that a similar mechanism could be at play here, since hiring channels do correlate with occupational attainment, with those being hired through networks more likely to work in lower skilled occupations, more similar to those observed in Heath (2018). Table 2.5 illustrates, that in fact, both explanations could be valuable in the data set at hand. The category ‘professional workers’, displayed in column (1), captures very high-skilled workers, from managers to technicians and associates. Among those, a significant share is still hired through networks (about 38 percent). Among these highly skilled workers, moral hazard is arguably less of a problem as they might be more motivated; instead, the employer could use networks to capture otherwise unobservable characteristics. Applying this to Table 2.5, employers might reasonably observe a potential employees’ communication skill during a formal job interview. If communication skills are important to the employer (or specific position), the employer would then predominantly search through formal channels. Indeed, if the employer thinks communication skills among their professional workers is important, the within-firm wage gap between formal and network hires is larger. Teamwork, on the other hand, might be more difficult to observe in a job interview (unless the employer is willing to include a group exercise). If the employer thinks that teamwork is an essential skill, they might predominantly search through social networks, which could provide information about how a potential worker would engage in a teamwork setting. Actually, if the employer deems teamwork important, the within-firm wage gap between formal and network hires decreases. This suggests that employers could be strategic in terms of which hiring channel to use, depending on what type of skills they require and the ability of the hiring channel to provide information about this skill.

We do not observe any significant correlation between skill importance as judged by the employer and the within-firm wage gap between formal and network hires for non-professional workers. This could be due to the fact that the skills required for non-professional workers were not part of the battery of skills included in the questionnaire (such as, for example, working diligently, not shirking, or handling stress well). If the main reason for hiring non-professional workers through social networks is to circumvent moral hazard, under the assumption that these types of occupations might not be as dependent on individual skills to ascertain a certain productivity, had we been able to, for example, observe the importance that an employer places on a worker not shirking in the questionnaire, we would have

expected a negative coefficient (i.e. a smaller within-firm wage gap between formal and network hires).

2.6 Conclusion

This paper provides a descriptive illustration of estimates of the wage returns to educational attainment, cognitive skills (or rather learning outcomes), and non-cognitive skills while taking into account selection into hiring channel using a novel matched employer-employee data set from Bangladesh.

We aim to take into account both the potential direct and indirect effects of cognitive and non-cognitive skills on wages, focusing on one indirect effect – the choice of hiring channel. The interplay between skills and hiring channel in determining labor market outcomes is a particularly important issue in developing countries, where the role of social networks is large and research on the effects of non-cognitive skills scarce. We incorporate the fact that the same unobserved characteristics could drive both the selection of the hiring channel and the wage through endogenous switching models. The data further allow for an innovative exploration of both demand and supply side factors in explaining returns to skills and selection into hiring channels.

We find that a higher score on the literacy test increases the probability of choosing formal hiring channels and decreases the probability of having found a job through social networks. Non-cognitive skills do not seem to correlate with hiring channel in the bimodal model. A brief multinomial model which further distinguishes between different network based hiring channels shows a potential, but weak, role for non-cognitive skills.

Cognitive skills (literacy) affect wages only through increasing the probability of choosing the formal over other hiring channels. Once this initial selection into hiring channel has been corrected for, there are no further significant wage returns to literacy. However, it seems that in our case literacy could simply act as a signal for good quality education. Literacy and years of education are highly collinear, while this is not the case for education and non-cognitive skills. Looking at wage returns to non-cognitive skills, we see that these differ by hiring channel, which relates to different jobs being predominantly filled through one channel or the other. These significant correlations are not visible in simple OLS regressions, illustrating the benefit of first taking into account selection into different hiring channels. Those hired through formal channels benefit from higher returns to openness to experience but lower returns to conscientiousness and hostile attribution bias. Those hired through networks enjoy higher wages for higher levels of emotional stability, which has been linked to task performance, but are also punished for higher hostile attribution bias. Exploring the time dimensions of wages, we investigate whether non-cognitive skills could also contribute to wage growth. We find that for those

hired through networks, grit and conscientiousness are both associated with higher wage growth over time. We should not expect any effects of individual effects that are constant over time on wage growth. The positive correlations of grit and conscientiousness with wage growth among network hires thus seem to suggest that these traits evolve over time as workers improve their performance and adapt closer to work requirements.

We then explore the demand side to understand whether firm characteristics and preferences for certain skills can help explain differential returns to skills. We find that employers who value communication skills more are associated with a larger within-firm wage gap between formal and network hires, while those who value teamwork more are associated with a smaller wage gap. These results only hold for professional workers, such as managers or technicians. We explain this by a firm's decision to hire through different channels among highly skilled workers. If an employer values communication skills, it will choose a hiring channel that allows them to observe these skills, even if this channel is potentially more costly for the employer. As communication skills are observable during formal interviews, this increases the wage gap between formal and informal hiring. Teamwork skills, on the other hand, might be more difficult to observe reliably in a simple job interview, which is why firms valuing these skills might rely more on networks to provide otherwise unobservable information about a worker. This mechanism only seems to hold for professional workers, though; among non-professional workers, other mechanisms, such as firms hiring through networks to overcome moral hazard problems such as workers' shirking as observed in Heath (2018), could dominate.

This paper illustrates that literacy remains important in developing countries by enabling workers to access their full potential – in this case by being hired through formal channels – even though it might not have an additional wage return apart from facilitating this first selection. It could also act as a signal for the quality of education received in a country where the average quality is low, with the literacy test capturing primary school-level reading skills. Despite the low level of knowledge tested, the average score even among those with a high school degree was only 6 correct answers out of 8 questions. Returns to non-cognitive skills are comparable to those found in developed countries, though our sample is also based on the formal sector only.

This paper provides a descriptive analysis of hiring channels and non-cognitive skills in developing countries and illustrates some channels that could be at work to explain why firms would choose to hire through one channel or another. To our knowledge, it is the first paper to do so in a developing country context, in which hiring through networks is highly prevalent. Future research attempting to provide more causal estimates for these mechanisms would certainly be beneficial. Our estimates show that non-cognitive skills play a potentially important role in developing countries' labor markets, but more research is necessary to understand

whether and to what extent returns to non-cognitive skills differ in other labor market segments, such as in informal firms, among the self-employed, or more broadly in rural areas. Further, the analysis in this paper highlights the need for more reliable measurement instruments to capture non-cognitive skills. The internal validity of most non-cognitive constructs used here is less than ideal. Developing reliable instruments for low-literacy developing country settings would help expand this quickly developing strand of research.

Appendix for Chapter 2: Tables and Figures

Table A2.1 – Correlations between the skills measures

	(1) Educ	(2) Lit	(3) Num	(4) OP	(5) CO	(6) EX	(7) AG	(8) ES	(9) HAB	(10) GR
Years of education	1.00									
Literacy score	0.75*	1.00								
Numeracy score	0.49*	0.60*	1.00							
Openness to Experience	0.01	0.01	0.00	1.00						
Conscientiousness	0.01	0.00	0.08*	0.39*	1.00					
Extraversion	-0.06*	-0.09*	-0.12*	0.27*	0.48*	1.00				
Agreeableness	0.00	-0.01	0.04*	0.32*	0.40*	0.28*	1.00			
Emotional Stability	-0.04	-0.05*	-0.01	0.30*	0.65*	0.53*	0.32*	1.00		
Hostile Attribution Bias	-0.05*	-0.06*	-0.03	0.18*	0.11*	0.23*	0.23*	0.05*	1.00	
Grit	0.09*	0.12*	0.16*	-0.06*	-0.13*	-0.26*	-0.04*	-0.32*	-0.07*	1.00

Source: 2012 Bangladesh Enterprise Based Skills Survey (ESS).

Notes: The value in each cell is the pairwise correlation; *** = $p < 0.01$; Male workers only. N = 4,678.

Table A2.2 – Characteristics of firms in samples

Share of firms	Original	Restricted
Commerce	0.150 (0.36)	0.150 (0.36)
Education	0.150 (0.36)	0.154 (0.36)
Finance	0.150 (0.36)	0.148 (0.36)
Manufacturing	0.400 (0.49)	0.396 (0.49)
Public administration	0.150 (0.36)	0.152 (0.36)
At most 20 employees	0.382 (0.49)	0.376 (0.48)
21 - 70 employees	0.232 (0.42)	0.236 (0.43)
More than 70 employees	0.386 (0.49)	0.388 (0.49)
Firm located in Dhaka	0.528 (0.50)	0.528 (0.50)
Main channel of job advert: networks	0.336 (0.47)	0.331 (0.47)
One channel of job advert: networks	0.540 (0.50)	0.536 (0.50)
N	500	487

Source: 2012 Bangladesh Enterprise Based Skills Survey (ESS). The restricted sample consists of male workers with non-missing skills variables only. Standard deviation in brackets.

Table A2.3 – Characteristics of employees

	Mean (1)	SD (2)	Formal (3)	SD (4)	Networks (5)	SD (6)	Difference (7)	P-value Diff (8)
Age	31.68	8.38	33.30	8.10	30.31	8.37	3.00	0.00
Married	0.77	0.42	0.83	0.37	0.72	0.45	0.11	0.00
Lives in Dhaka	0.59	0.49	0.51	0.50	0.65	0.48	-0.14	0.00
Age at hiring	26.04	6.47	26.88	5.86	25.33	6.86	1.55	0.00
Work experience	6.28	6.13	6.96	6.63	5.71	5.61	1.25	0.00
Primary education	0.31	0.46	0.07	0.26	0.50	0.50	-0.43	0.00
Secondary education	0.46	0.50	0.49	0.50	0.43	0.50	0.06	0.00
Tertiary education	0.23	0.42	0.44	0.50	0.06	0.24	0.37	0.00
Literacy score	4.62	2.60	5.99	2.02	3.45	2.47	2.54	0.00
Numeracy score	5.71	2.01	6.36	1.66	5.16	2.11	1.20	0.00
Openness to experience	2.13	0.60	2.13	0.61	2.14	0.60	-0.01	0.62
Conscientiousness	2.44	0.80	2.39	0.82	2.48	0.78	-0.09	0.00
Extraversion	1.88	0.72	1.82	0.75	1.92	0.69	-0.10	0.00
Agreeableness	2.08	0.60	2.06	0.59	2.09	0.61	-0.03	0.10
Emotional stability	2.32	0.82	2.26	0.85	2.37	0.79	-0.12	0.00
Hostile attribution bias	1.86	0.73	1.84	0.71	1.87	0.74	-0.03	0.19
Grit	2.10	0.55	2.11	0.56	2.10	0.54	0.01	0.62
Log Hourly Wage	3.62	0.61	3.89	0.59	3.39	0.51	0.50	0.00
Manager	0.05	0.22	0.05	0.22	0.06	0.23	-0.01	0.27
Skilled white collar	0.55	0.50	0.76	0.43	0.38	0.48	0.38	0.00
Skilled blue collar	0.27	0.44	0.10	0.30	0.41	0.49	-0.31	0.00
Unskilled	0.13	0.34	0.09	0.29	0.16	0.37	-0.07	0.00
Permanent contract	0.91	0.29	0.96	0.19	0.86	0.35	0.10	0.00
Fixed-term contract	0.05	0.23	0.01	0.11	0.09	0.29	-0.08	0.00
Part time contract	0.03	0.16	0.02	0.15	0.03	0.17	0.00	0.35
Seasonal contract	0.01	0.10	0.00	0.00	0.02	0.14	-0.02	0.00
Found job through network	0.54	0.50	0.00	0.00	1.00	0.00	-1.00	
N	4,678		2,141		2,537			
N contract type	4,650		2,140		2,510			

Source: 2012 Bangladesh Enterprise Based Skills Survey (ESS).

Notes: Sample excludes employees for whom personality questions are missing (about 16 percent). Male workers only. Maximum score for literacy and numeracy is 8. Maximum score for the personality trait variables is 4.

Figure A2.1 – Current wages by hiring channel and industry

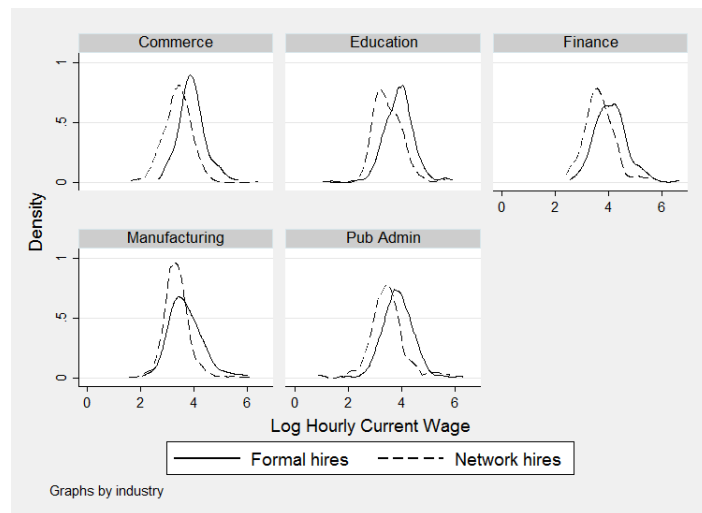


Table A2.4 – Estimates for log hourly wages regressed on cognitive and non-cognitive skills with non-cognitive skills derived from exploratory factor analysis

Model	OLS:			probit:	selection correction:	
Dependent variable	log hourly wages			network hiring	log hourly wages	
	(1)	(2)	(3)	(4)	Formal (5)	Network (6)
Years of education	-0.002 (0.009)	-0.004 (0.010)	-0.022** (0.010)	-0.051** (0.023)	-0.044** (0.020)	-0.018** (0.007)
Years of education (sqrd)	0.004*** (0.000)	0.004*** (0.000)	0.003*** (0.000)	0.000 (0.001)	0.003*** (0.001)	0.003*** (0.000)
Factor 1 (ES/EX) (std)	0.010 (0.022)	0.019 (0.019)	0.025 (0.017)	-0.042 (0.044)	-0.015 (0.023)	0.057*** (0.018)
Factor 2 (CO/Grit) (std)	-0.008 (0.018)	-0.013 (0.016)	-0.018 (0.015)	0.074** (0.034)	0.028 (0.021)	-0.038*** (0.015)
Factor 3 (OP/AG) (std)	0.011 (0.012)	0.003 (0.010)	-0.003 (0.009)	0.016 (0.024)	0.010 (0.013)	-0.020* (0.010)
Reading score			0.035** (0.016)	-0.116*** (0.039)	0.024 (0.024)	-0.007 (0.015)
Numeracy score			0.006 (0.016)	0.010 (0.033)	0.008	0.001
Mother no formal educ				0.130*** (0.045)		
HH inc: 5,000-7,500tk				-0.119 (0.124)		
HH inc: 7,501-10,000tk				-0.138 (0.126)		
HH inc: 10,001-15,000tk				-0.239* (0.132)		
HH inc: 15,001-20,000tk				-0.185 (0.141)		
HH inc: 20,001-30,000tk				-0.323** (0.147)		
HH inc: 30,001-50,000tk				-0.211		
Indvl controls	YES	YES	YES	YES	YES	YES
Firm fixed effects		YES	YES	YES	YES	YES
Occup dummies			YES	YES	YES	YES
Inverse mills					-0.198*** (0.054)	0.144*** (0.048)
N	3,102	3,102	3,102	1,974	3,096	3,096
R ²	0.459	0.412	0.474	0.396		

Notes: Standard errors clustered at the firm-level in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Individual controls are a dummy for being married, total work experience and total work experienced squared. The decrease in the sample size compared to the regressions using the Big Five is due to some non-cognitive skills variables having missing values, dropping those individuals from the factor analysis. OP = openness to experience; CO = conscientiousness; EX = extraversion; AG = agreeableness; ES = emotional stability; GR = grit; Exploratory factor analysis based on acquiescence corrected items.

Table A2.5 – Variance inflation factor for different OLS models

	Model 1	Model 2	Model 3	Model 4
	(1)	(2)	(3)	(4)
Years of schooling	13.3	13.3	17.1	
Years of schooling (sqrd)	13.4	13.5	14.2	
Experience	6.2	6.3	6.3	6.2
Experience (squared)	6.0	6.0	6.0	6.0
Married	1.1	1.1	1.1	1.1
OP (std)		1.3	1.3	1.3
CO (std)		2.0	2.1	2.1
EX (std)		1.6	1.6	1.6
AG (std)		1.3	1.3	1.3
ES (std)		2.1	2.1	2.1
HAB (std)		1.1	1.1	1.1
GR (std)		1.2	1.2	1.2
Reading score			2.8	1.6
Numeracy score			1.7	1.7

Notes: Standard errors clustered at the firm-level in parentheses; OP = openness to experience; CO = conscientiousness; EX = extraversion; AG = agreeableness; ES = emotional stability; HAB = hostile attribution bias; GR = grit; Personality traits are acquiescence corrected. Variance inflation factor calculated after OLS regressions. All covariates included are shown in the table. Literacy and numeracy scores are standardized.

Table A2.6 – OLS and endogenous switching models without education as a covariate

Model	OLS:	selection correction:	
	(1)	Formal	Networks
	(1)	(2)	(3)
Reading score	0.088*** (0.013)	0.025 (0.021)	0.003 (0.015)
Numeracy score	0.003 (0.013)	0.000 (0.016)	0.015 (0.012)
OP (std)	0.004 (0.009)	0.024** (0.011)	-0.007 (0.011)
CO (std)	-0.022* (0.012)	-0.014 (0.015)	-0.021 (0.014)
EX (std)	0.002 (0.010)	0.001 (0.014)	-0.004 (0.013)
AG (std)	0.014* (0.008)	0.010 (0.011)	0.017* (0.010)
ES (std)	0.007 (0.010)	0.009 (0.015)	0.024* (0.014)
HAB (std)	-0.013* (0.008)	-0.028*** (0.010)	-0.016* (0.009)
GR (std)	0.008 (0.009)	-0.010 (0.012)	0.008 (0.011)
Inverse mills ratio		-0.409*** (0.054)	0.420*** (0.053)
N	4,678	4,655	4,655
R ²	0.434		

Notes: Standard errors clustered at the firm-level in parentheses; *** p<0.01, ** p<0.05, * p<0.1. All specifications include individual controls (dummy for being married, work experience and work experienced squared), occupation dummies, and firm fixed effects. OP = openness to experience; CO = conscientiousness; EX = extraversion; AG = agreeableness; ES = emotional stability; HAB = hostile attribution bias; GR = grit; Personality traits are acquiescence corrected. Literacy and numeracy scores are standardized. Instruments used for identification of switching model: mother has no formal education and monthly household income. Switching model estimated using a two-step model. The difference in sample size between column 1 and columns 2 and 3 is due to some missing values in the identifying variable 'household income'.

Table A2.7 – Log earnings regressed on cognitive and non-cognitive skills and determinants of hours worked per week – OLS

	Log current earnings					Hours worked
	(1)	(2)	(3)	(4)	(5)	(6)
Years of education	-0.016** (0.007)	-0.010* (0.005)	-0.023*** (0.006)	-0.028*** (0.005)	-0.028*** (0.005)	-0.056 (0.133)
Years of education (sqrd)	0.004*** (0.000)	0.004*** (0.000)	0.004*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	-0.008 (0.006)
OP (std)	0.007 (0.010)	0.005 (0.007)	0.006 (0.007)	0.003 (0.007)	0.003 (0.007)	0.061 (0.153)
CO (std)	-0.005 (0.011)	-0.015* (0.009)	-0.018** (0.009)	-0.017** (0.008)	-0.017** (0.008)	0.273 (0.234)
EX (std)	-0.004 (0.011)	-0.006 (0.007)	-0.003 (0.007)	-0.000 (0.007)	-0.000 (0.007)	-0.220 (0.214)
AG (std)	0.003 (0.008)	0.012** (0.006)	0.012** (0.006)	0.009* (0.005)	0.009* (0.005)	-0.057 (0.195)
ES (std)	0.015 (0.012)	0.010 (0.009)	0.010 (0.009)	0.011 (0.008)	0.011 (0.008)	0.070 (0.203)
HAB (std)	0.001 (0.007)	-0.004 (0.006)	-0.004 (0.006)	-0.006 (0.006)	-0.006 (0.006)	0.282* (0.147)
GR (std)	0.017* (0.010)	0.009 (0.007)	0.008 (0.007)	0.005 (0.007)	0.005 (0.007)	0.278 (0.175)
Reading score			0.040*** (0.011)	0.033*** (0.011)	0.033*** (0.011)	-0.147 (0.251)
Numeracy score			0.018* (0.010)	0.018* (0.010)	0.018* (0.010)	0.597** (0.260)
Indvl controls	YES	YES	YES	YES	YES	YES
Firm fixed effects		YES	YES	YES	YES	YES
Occupation dummies				YES	YES	YES
N	4,678	4,678	4,678	4,678	4,678	4,678
R ²	0.446	0.492	0.497	0.548	0.548	0.036

Notes: Standard errors clustered at the firm-level in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Individual controls are a dummy for being married, work experience and work experienced squared. OP = openness to experience; CO = conscientiousness; EX = extraversion; AG = agreeableness; ES = emotional stability; HAB = hostile attribution bias; GR = grit; Personality traits are acquiescence corrected. Literacy and numeracy scores are standardized.

Table A2.8 – Probability of being hired through different hiring channels: marginal effects after multinomial logit

Being hired through	Formal (1)	Family (2)	Friends (3)	Village (4)
Years of education	0.042*** (0.014)	-0.013* (0.008)	-0.011 (0.008)	-0.018*** (0.006)
Years of education (sqrd)	0.000 (0.001)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)
OP (std)	-0.000 (0.013)	-0.009 (0.011)	-0.002 (0.010)	0.012* (0.007)
CO (std)	-0.008 (0.018)	-0.000 (0.012)	0.009 (0.013)	-0.001 (0.009)
EX (std)	0.009 (0.016)	-0.003 (0.012)	-0.015 (0.011)	0.009 (0.009)
AG (std)	-0.004 (0.012)	0.015 (0.009)	-0.010 (0.008)	-0.002 (0.006)
ES (std)	-0.023 (0.016)	0.023* (0.013)	-0.008 (0.012)	0.008 (0.009)
HAB (std)	0.010 (0.012)	0.000 (0.008)	0.010 (0.008)	-0.020*** (0.006)
GR (std)	0.010 (0.013)	0.033*** (0.010)	-0.031*** (0.009)	-0.012 (0.008)
Reading score	0.079*** (0.022)	-0.054*** (0.014)	-0.029** (0.014)	0.004 (0.009)
Numeracy score	-0.022 (0.020)	0.022* (0.012)	0.002 (0.011)	-0.002 (0.008)
Mother no formal educ	-0.132*** (0.027)	0.030* (0.017)	0.052*** (0.018)	0.050*** (0.015)
HH inc: 5,000-7,500tk	0.135 (0.087)	-0.062 (0.044)	-0.044 (0.046)	-0.029 (0.033)
HH inc: 7,501-10,000tk	0.162* (0.088)	-0.039 (0.045)	-0.082* (0.047)	-0.040 (0.036)
HH inc: 10,001-15,000tk	0.151* (0.090)	-0.086* (0.045)	-0.029 (0.046)	-0.036 (0.037)
HH inc: 15,001-20,000tk	0.118 (0.093)	-0.095* (0.049)	0.003 (0.050)	-0.026 (0.037)
HH inc: 20,001-30,000tk	0.254*** (0.098)	-0.061 (0.057)	-0.106* (0.060)	-0.087* (0.051)
HH inc: 30,001-50,000tk	0.086 (0.122)	-0.033 (0.068)	-0.032 (0.072)	-0.021 (0.064)
HH inc: 50,001tk and above	1.334*** (0.190)	-2.371*** (0.128)	0.685*** (0.148)	0.352*** (0.087)
Indvl controls	YES	YES	YES	YES
Occupation dummies	YES	YES	YES	YES
Firm characteristics	YES	YES	YES	YES
Observations	4,655	4,655	4,655	4,655
Pseudo R^2	0.219	0.219	0.219	0.219

Notes: Standard errors are bootstrapped (200reps) and clustered at the firm-level; *** p<0.01, ** p<0.05, * p<0.1; Marginal effects reported are at the means of covariates. The hiring channel 'politics and school' has been combined with the hiring channel 'family' due to the small number of observations in the former and this being suggested by a Wald test of whether two outcomes can be combined. Individual controls experience, experience squared, and a dummy for being married. Firm characteristics include industry, firm size, a dummy for exporting firms, and the gender of the CEO. Personality traits are acquiescence corrected. Literacy and numeracy scores are standardized. The table reports marginal effects at means of covariates.

Table A2.9 – Endogenous switching model. Second stage of log hourly wages regressed on cognitive and non-cognitive skills and interactions with being a white collar worker

Log hourly current wage	Formal (1)	Networks (2)
Reading score	0.066** (0.028)	-0.033** (0.014)
Numeracy score	-0.014 (0.026)	0.022** (0.011)
Reading score # white collar	-0.061* (0.032)	0.064*** (0.019)
Numeracy score # white collar	0.018 (0.029)	-0.024 (0.017)
OP (std)	0.032 (0.021)	-0.021* (0.011)
CO (std)	0.009 (0.028)	-0.003 (0.014)
EX (std)	0.010 (0.024)	-0.004 (0.012)
AG (std)	-0.012 (0.022)	0.011 (0.010)
ES (std)	-0.046* (0.026)	0.014 (0.014)
HAB (std)	-0.000 (0.022)	-0.022** (0.009)
GR (std)	-0.011 (0.022)	0.016 (0.011)
OP (std) # white collar	-0.007 (0.023)	0.016 (0.015)
CO (std) # white collar	-0.042 (0.030)	-0.031 (0.020)
EX (std) # white collar	-0.005 (0.027)	0.012 (0.017)
AG (std) # white collar	0.029 (0.024)	-0.001 (0.015)
ES (std) # white collar	0.063** (0.030)	0.010 (0.020)
HAB (std) # white collar	-0.028 (0.024)	0.019 (0.014)
GR (std) # white collar	-0.002 (0.024)	-0.028* (0.015)
Indvl controls	YES	YES
Occupation dummies	YES	YES
Firm fixed effects	YES	YES
Inverse Mills ratio	-0.234*** (0.051)	0.133*** (0.047)
N	4,655	4,655

Notes: Standard errors clustered at the firm-level in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Individual controls are education, education squared, a dummy for being married, work experience and work experience squared. OP = openness to experience; CO = conscientiousness; EX = extraversion; AG = agreeableness; ES = emotional stability; HAB = hostile attribution bias; GR = grit; Personality traits are acquiescence corrected. Literacy and numeracy scores are standardized. Instruments used for identification: mother has no formal education and monthly household income. White collar occupations: managers, professionals, clerks, service and sales workers. Estimated using a two-step model.

Table A2.10 – Returns to skills using an endogenous switching model - Multiple hiring channels

Log hourly wages	Formal (1)	Family (2)	Friends (3)	Village (4)
Years of education	-0.049 (0.030)	-0.035 (0.029)	-0.028 (0.028)	-0.061* (0.035)
Years of education (sqrd)	0.003*** (0.001)	0.002 (0.002)	0.002 (0.002)	0.005** (0.002)
OP (std)	0.019 (0.018)	-0.016 (0.039)	-0.007 (0.034)	-0.005 (0.050)
CO (std)	-0.003 (0.019)	0.002 (0.036)	-0.011 (0.034)	-0.042 (0.046)
EX (std)	0.028 (0.022)	0.036 (0.040)	0.027 (0.037)	0.029 (0.064)
AG (std)	0.003 (0.015)	0.014 (0.034)	0.037 (0.026)	0.032 (0.038)
ES (std)	0.008 (0.025)	0.071 (0.046)	0.071* (0.041)	0.060 (0.054)
HAB (std)	-0.030 (0.020)	-0.033 (0.043)	-0.022 (0.038)	-0.014 (0.077)
GR (std)	-0.014 (0.026)	0.037 (0.059)	0.079 (0.048)	0.035 (0.067)
Reading score	0.031 (0.051)	-0.024 (0.076)	-0.050 (0.056)	-0.094 (0.091)
Numeracy score	-0.027 (0.031)	0.010 (0.041)	0.013 (0.035)	0.113* (0.058)
Indvl controls	YES	YES	YES	YES
Occupation dummies	YES	YES	YES	YES
Selection variables	YES	YES	YES	YES
Firm characteristics	YES	YES	YES	YES
ρ_1	-0.403* (0.220)	-0.807 (0.640)	0.077 (0.365)	-0.393 (0.571)
ρ_2	0.365 (0.924)	0.469 (0.378)	1.412** (0.630)	1.344 (0.992)
ρ_3	-0.868 (0.927)	-0.793 (0.802)	-0.070 (0.280)	-0.491 (1.078)
ρ_4	1.475 (1.028)	1.313 (1.067)	1.329 (0.861)	0.109 (0.407)

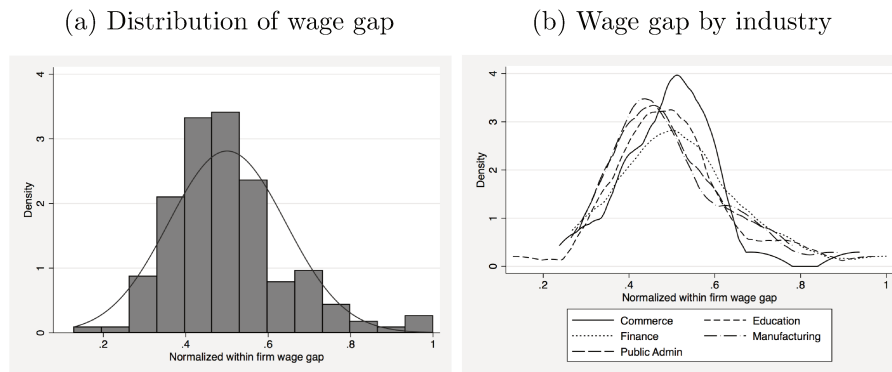
Notes: Standard errors are bootstrapped (200reps) and clustered at the firm-level; *** p<0.01, ** p<0.05, * p<0.1; Individual controls include experience, experience squared, and a dummy for being married. Firm characteristics include industry, firm size, a dummy for exporting firms, and the gender of the CEO. Personality traits are acquiescence corrected. Numeracy and literacy scores are standardized. Instruments used for selection identification are education of the mother and monthly household income. Selection corrected using Bourguignon et al. (2007).

Table A2.11 – Characteristics of firms used in wage gap analysis sample

Sample	Sample 1	Sample 2	Sample 3
Used for	OLS regressions	Selection correction with firm fixed effects	Decomposition of within firm wage gap
Restrictions	Male & non-missing NC	Same as (1)	Same as (1)
Hiring channels	all	≥ 1 formal & network	≥ 2 formal & network
Percentage of firms	(1)	(2)	(3)
Commerce	0.150 (0.36)	0.147 (0.36)	0.157 (0.37)
Education	0.154 (0.36)	0.194 (0.39)	0.192 (0.39)
Finance	0.148 (0.36)	0.159 (0.37)	0.157 (0.37)
Manufacturing	0.396 (0.49)	0.368 (0.48)	0.401 (0.49)
Public administration	0.152 (0.36)	0.132 (0.34)	0.093 (0.29)
At most 20 employees	0.376 (0.48)	0.260 (0.44)	0.169 (0.38)
21 - 70 employees	0.236 (0.43)	0.271 (0.45)	0.285 (0.45)
More than 70 employees	0.3888 (0.49)	0.469 (0.50)	0.547 (0.49)
Firm located in Dhaka	0.528 (0.50)	0.609 (0.49)	0.657 (0.48)
Main channel of job advert: networks	0.331 (0.47)	0.217 (0.41)	0.192 (0.39)
One channel of job advert: networks	0.536 (0.05)	0.481 (0.50)	0.477 (0.50)
Total number of firms	487	258	171

Source: 2012 Bangladesh Enterprise Based Skills Survey (ESS). Sample 1 represents the sample used for the initial OLS regressions. Sample 2 represents the sample of firms used for the endogenous switching model including firm fixed effects. Sample 3 is the sample used for the decomposition of the within-firm wage gap.

Figure A2.2 – Within firm wage gap



Chapter 3

The determinants of trust: Evidence from rural South India

Abstract¹

Trust and participation in social networks are inherently interrelated. We make use of the demonetization policy in India, an unexpected and unforeseeable exogenous variation that had direct effects on networks but not on interpersonal trust, to causally identify the effect of social networks in determining trust. We use first-hand quantitative and qualitative data from rural South India and control for a variety of individual characteristics that could influence network formation and trust, such as personality traits and cognitive ability. We find that social interactions only had a significant effect on levels of trust among men. Further, we find important differences along the lines of caste membership. Among lower castes, who live in homogeneous neighborhoods and relied on their neighbors and employers to cope with the shock, making use of one's network more intensely increases levels of trust placed in neighbors. Among middle castes, who live in more heterogeneous neighborhoods and relied predominantly on other caste members to cope, a larger network size leads to higher levels of trust placed in kin among employees but lower levels of trust in neighbors (who tend to be more dissimilar). This paper thus shows that social interactions can foster trust, though this is dependent on the type of interaction occurring. The paper also demonstrates the importance of having clearly defined in- and out-groups in trust measures, given the highly segregated nature of social interactions in rural South India.

¹This chapter is co-authored with Christophe Jalil Nordman.

3.1 Introduction

Many transactions in developing countries, from obtaining personal credit to workplace interactions and business transactions, involve personal, informal relationships, relying on so-called social capital instead of formal institutions. Social capital refers to the “actual and potential resources which are linked to the possession of a durable network of more or less institutionalized relationships of mutual acquaintance and recognition” (Bourdieu, 1980, p.2). Trust in those around one is thus an essential ingredient of social capital, enforcing transactions in the absence of formal markets (Fukuyama, 1995; Putnam, 2001). Trust can be understood as an “optimistic expectation or belief regarding other agents’ behavior” (Fafchamps, 2006, p.1183) arising from a variety of sources, such as repeated social interactions or a general knowledge about the share of trustworthy and cheating agents in a given population and their incentives. It has been shown to play an important role in economic performance (Knack and Keefer, 1997; Slemrod and Katuščák, 2005).

Despite the importance of trust for facilitating informal interactions, our understanding of the origins and determinants of trust remains limited. Fehr (2009) notes that informal institutions, such as social networks, are likely to shape trust. However, a causal relationship is difficult to establish due to the inherent endogeneity; individuals’ beliefs about the trustworthiness of others are influenced by experiences of others’ trustworthiness, which in turn feeds back into interpersonal interactions and beliefs. Nooteboom (2007, p. 33) phrases the dilemma as follows: “Trust is both an outcome and an antecedent of relationships. It forms a basis for relationships, and thus generates social capital. It may be based on institutions, and it may be built from relationships, and then it arises from social capital.”

In attempting to understand the causal relationship between social interactions and trust, the economic literature has resorted to the use of unexpected shocks (conflict, violence) and economic games. Rohner et al. (2013) take the example of civil conflict in Uganda and, exploiting variations in the spatial and ethnic nature of fighting, find that more intensive exposure to fighting decreases generalized trust and increases ethnic identity, attributing their findings to a breakdown of civic and economic cooperation within society. Similarly, Fearon et al. (2009) look at the effect of a positive shock - the arrival of a donor-aided, community-driven reconstruction program - on social cohesion in Liberia, measured by the amount of funding a community raised through public good games, finding that a simple participatory politics program, designed to increase community committee structures and support those structures in meeting community needs, increases the amount of money that a community raised. The role of social interactions thus seems crucial; Rohner et al. (2013) attribute the reduction in trust following conflict to a breakdown of cooperation, whereas Fearon et al. (2009) find that increased social interaction through a community based program leads to increased

group cohesion. Indeed, social connections have been found to be an important determinant of trusting others (Glanville et al., 2013).

The aim of this paper is to causally identify the determinants of trust in a setting marked by high levels of informal transactions and strict social hierarchies: rural South India. In this setting, traditional agrarian structures, based on a strict segmentation and hierarchy of occupations according to caste and gender, are increasingly contested and reconfigured, with social networks playing a growing role. Rapid development in the South Indian state of Tamil Nadu, one of India's most developed, urbanized, and industrialized states, has resulted in a complex society wherein old structures coexist with new forms of relationships in both the labor market and social hierarchies (Guérin et al., 2015).

This paper contributes to the literature by causally identifying the determinants of trust in this dynamic rural setting using the demonetization policy in India as a source of exogenous variation and relying on first-hand quantitative and qualitative data sources. Demonetization, the ban of the two highest value banknotes in circulation on November 8th, 2016, was unexpected, unforeseeable, and took place overnight and led to severe cash shortages. Households could not have prepared for it and were hit by an exogenous variation in money supply, which is especially relevant in a cash-based economy such as India. Resulting cash shortages led to an increased demand for informal credit, and people were forced to rely on their social networks more than ever to cope with the shock. This external variation thus shifted individuals' reliance on their networks, revealing information about who they could rely on in times of crisis (based on how they judge others' trustworthiness). The shock did not have any direct effect on levels of trust placed into neighbors or kin: we will assume and show that any effect should purely operate through changed patterns of social interaction.

The setting is quite similar to the cash-deprived economy of early modern England, described by Muldrew (1998), where cash shortages led to an increased demand for informal credit and a multiplicity of informal transactions. As formal credit guarantees provided by the state were weak and demand for informal credit high, households found ways to provide informal credit or material exchanges to each other based on trust. In such an economic system, neighbors were encouraged to judge one another's credit and thrift. This mechanism of coping with shocks had already been well-established in South India before the demonetization shock and intensified as a result. The South Indian setting is special, however, due to its dependence on caste as social capital (Munshi and Rosenzweig, 2016), as illustrated by the following example:²

Gomathi (female, 26 years old) is an agricultural coolie (laborer), living with her husband, who migrates part of the year to another state for

²The life stories included in this paper stem from semi-structured interviews that were conducted by the authors in December 2016, approximately one month after demonetization.

work. When asked who she would ask for help while her husband was away, she mentioned her family and the ‘people around her’. Asked what she meant by the latter, she described a reciprocal system, in which she could always ask her female neighbors, members of the same caste, for small urgent amounts of money. To quote: “This kind of help, they [other women] never ask any interest. But at the same time, they also demand 100 rupees from me whenever they need it. So you are in a position to give 100 rupees whenever they demand as well.”

Trust is likely to be determined not only by social interactions but by a variety of individual characteristics, such as gender, age, height (Dohmen et al., 2008) or cognitive ability and personality traits. Jones (2008) surveys the literature on cooperation games (prisoner’s dilemma games) and finds that students from schools with higher SAT scores, a standardized test widely used for college admissions in the United States, cooperate more than those from schools with lower scores. In the game’s setting, trust arises as one player (the investor) has to decide whether to send her endowment to the other person or keep it. The decision to send money (and how much to send) depends, then, on the investor’s beliefs about the other person’s trustworthiness (willingness and probability to cooperate) and the investor’s willingness to make herself vulnerable to the actions of another person (Hong and Bohnet, 2007). Dohmen et al. (2008) use data from the German Socio-Economic Panel and present evidence that psychometric measures (measured by the Big Five, a personality test thought to capture the broadest level of personality traits) have predictive power for trust and reciprocity. Our paper is special in its ability to include measures of individual cognitive and non-cognitive ability in a rural developing country setting. Assuming consistent measurement of these traits, we can thereby include variables into the regression that are usually part of the unexplained individual heterogeneity captured by the error term. Further, the determinants of trust could well vary between countries and cultural areas. India, for example, has above average values on positive reciprocity on a global level (Falk et al., 2018), justifying the focus on this particular region.

We find that social interactions have a significant effect on levels of trust among men, only. Further, we find important differences along the lines of caste membership. Among lower castes, who live in homogeneous neighborhoods and relied on their neighbors and employers to cope with the shock, making use of one’s network more intensely increases levels of trust placed in neighbors. Among middle castes, who live in more heterogeneous neighborhoods and relied predominantly on other caste members to cope, a larger network size leads to higher levels of trust placed in kin among employees but lower levels of trust in neighbors (who tend to be more dissimilar). Higher network density among this group also leads to lower trust in kin, as those kin tend to be weaker links. This paper thus shows that social interactions can foster trust, though this is dependent on the type of interaction

occurring. The paper also demonstrates the importance of having clearly defined in- and out-groups in trust measures given the highly segregated nature of social interactions in rural South India

The remainder of this paper proceeds as follows: section 3.2 provides context for the study region and the demonetization policy; section 3.3 offers a brief theoretical framework that is useful for understanding the presupposed mechanisms; section 3.4 introduces our data set and the construction of the main variables; section 3.5 describes the empirical strategy; section 3.6 depicts our results and robustness checks and section 3.7 concludes.

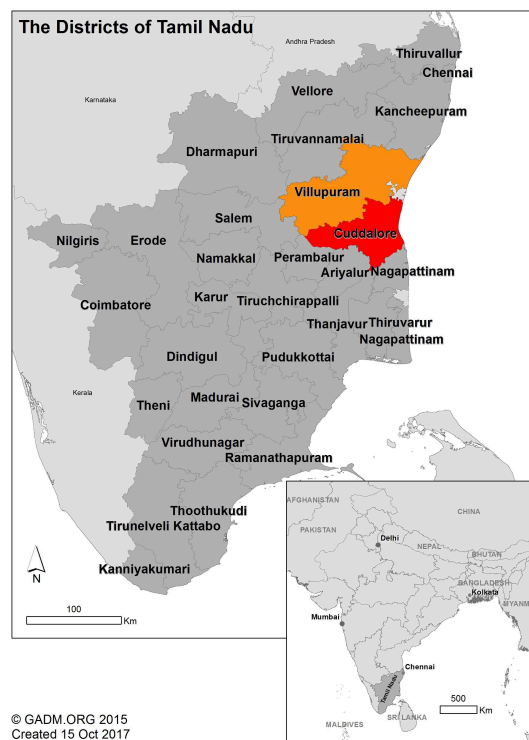
3.2 Background

3.2.1 Tamil Nadu

The data collected for this paper stem from Tamil Nadu, a state in the extreme South of the Indian Subcontinent (marked in color in Figure 3.1). Like India as a whole, it has seen impressive economic growth over the last several decades, but it is also one of India's most developed, urbanized, and industrialized states. The changes in recent years have been accompanied by strong inequalities between urban and rural areas, however. This two-tier development has resulted in a complex society wherein old structures coexist with new forms of relationships in the labor market and in social hierarchies.

Over the last three decades, in the region studied, members of upper castes (oftentimes landholders) have moved away from local villages to nearby towns, selling their land to members of middle castes, thereby initiating a restructuring of land and labor (Guérin et al., 2015). As a result, the protection traditionally provided by landholders has gradually been replaced by a contractualization of labor. Such a land transfer from the traditionally dominant caste to the intermediate and lower castes has reshaped local power structures, and therefore network structures. The fragmentation of land and the associated changed organization of labor supply have then led to the development of non-agricultural employment, while simultaneously increasing the importance of networks and en-

Figure 3.1 – Map of the study region



couraging intra-caste solidarity (for instance, with regard to accessing urban jobs). New occupations in rural non-farm employment have also increased connections between urban and rural areas and promoted social and geographical mobilities (Bremar, 1996; Guérin et al., 2013). Local inequalities remain strong: while the situation of the Dalits has been improving due to a combination of temporary migration and government schemes (Guérin et al., 2015), they continue to be disadvantaged on the labor market as the vast majority of employers are from middle and upper castes.

Social policies targeting the poor and lower castes have led to an increased participation of lower castes through new forms of activism (trade unions, farmers' associations, autonomous caste associations), serving to reinforce local community networks (Vijayabaskar and Kalaiyarasan, 2014). It is in this changing economic and social landscape that we aim to investigate the determinants of trust.

3.2.2 Demonetization

On November 8th, 2016 at 8pm local time, Indian Prime Minister Narendra Modi announced the ban of the 500 and 1,000 rupee notes, the two highest value banknotes in circulation. From midnight onward, these two notes were no longer legal tender and had to be exchanged in banks for new notes, affecting about 86 percent of the entire money supply. The policy was supposed to contribute to the formalization of the economy by fighting corruption, the illegal economy, counterfeit money, and terrorism, in addition to fostering the digitization of banking.

The implementation process faced many technical challenges, leading to severe cash shortages. Due to the importance of cash in the Indian economy (98 percent of transactions are estimated to be in cash), this measure had strong impacts on employment, daily financial practices, and network use for more than three months, as people relied more strongly on their networks to sustain their economic and social activities. During the first two months, cash withdrawals were limited (first to 2,000 and later 4,000 rupees per day per bank card) and lines at ATMs long, making obtaining cash a time consuming experience. The policy shock hit rural households particularly strongly, as 80 percent of ATMs are located in urban and peri-urban areas, making it more difficult for rural households to travel to them. Further, new notes were unequally distributed. In the state of Tamil Nadu, 44 percent of newly delivered notes were distributed to three private banks with only 900 branches, while public banks, with over 9,000 branches (many in rural areas), received the remainder (Ghosh et al., 2017). Further, few rural households had access to a bank account prior to demonetization, and most who did only used it to receive transfers from government schemes. Most rural households were thus hardly ever in touch with the formal banking system.

Informal social networks have been successful in mitigating the impact of this shock in multiple ways (Guérin et al., 2017). Rich individuals in our study region

were able to get rid of their old notes through social relations and business tactics such as prematurely paid advances, while poorer ones could rely on their networks for informal loans; all of these relationships rely on the necessary condition of trust. Demonetization led to new markets to exchange old notes at discount rates ranging from 18-40 percent (Ghosh et al., 2017). Still, this mitigation mechanism only holds for those who are integrated into social networks (Guérin et al., 2017), illustrated by the following example from our qualitative fieldwork:

Sabeema is a female tailor who manufactures clothes for the women in her community, mostly from the same street. She experienced a reduction in customers and is increasingly working for credit. Her husband is employed as a TV-mechanic in a nearby city. He is usually paid weekly, but he had not been paid for a few weeks as his employer did not have any cash available. As a result of the double shortfall of wages, the family had to reduce their food consumption. They were not able to ask their network for help since everybody in their network was in a similar situation.

This reliance on informal credit channels such as friends and family, moneylenders, and black markets to exchange bills hit the poor and marginalized especially hard, as they saw their oftentimes meagre cash holdings losing value or being worthless. As such, social networks can have inequitable consequences when dealing with shocks, potentially widening the gap between those with and without connections (Fafchamps, 2006; Guérin et al., 2017).

3.3 Conceptual framework

Trust can be defined as one individual voluntarily placing resources (of whatever kind) at the disposal of someone else. In economic terms, the individual expects to be better off after making her resources available, with better off defined according to whatever goal the investor has in mind (Fehr, 2009). Given this definition, we provide a brief conceptual framework to illustrate the hypothesized relationship between social networks and trust at the core of this paper. The conceptual framework relies heavily on Guiso et al. (2008), who formalized Berg et al.'s (1995) trust game.

Assume that an individual lives in an economy consisting of two types of agents: trustworthy agents and non-trustworthy agents. The individual is then embedded in one of two potential social networks: an honest network, in which the share of trustworthy agents predominates, and a cheating network, in which the share of non-trustworthy agents is in the majority. The individual knows that there are both cheating and trustworthy agents in the economy and knows that either one could be present in her network. However, in this stylized framework, the

individual is not *a priori* aware of the type of network that she's embedded in, leading to the following distribution (where $q_1 > q_2$):

Table 3.1 – Distribution of trustworthy agents and cheaters

	Social Network	
	Honest	Cheater
Share of trustworthy agents	q_1	q_2
Share of cheaters	$1-q_1$	$1-q_2$

In the first period, the individual is endowed with her initial endowment x . In the second period, the individual can invest her endowment within her network, without knowing which of the two types of receivers (trustworthy or cheater) will receive her investment. This simplistic framework disregards the role of reputation as an information-sharing mechanism among individuals, which could influence an individual's propensity to engage (or not) with a specific recipient. All the individual knows prior to investing is that there might be both cheaters and honest recipients, without being able to tell who is who.

With a trustworthy receiver, the individual's investment accrues a positive return, $r > 0$, whereas with a cheater receiver, the individual accrues a loss, $l < 0$. In this framework, the individual needs to make the decision of whether or not to invest her endowment under a condition of uncertainty, as the type of receiver (trustworthy or cheater) is only revealed afterwards. It is only through investing, i.e. through interacting with the receiver, that the individual gains knowledge about the type of receiver and the type of network she's embedded in. Thus, reputation building only occurs at the individual level, and an individual will only keep interacting with those who are honest (reputable). The set-up thus implies Bayesian updating of the network in which an individual lives and interacts. Specifically, an individual might think that her network is a trustworthy one. As a result of a shock, for example, the individual might then grant another individual a loan, thereby investing her endowment. Only after this interaction does the individual realize that her network has a high number of cheaters (i.e. she lives in a cheater environment) and that her endowment is lost. The individual then updates her beliefs about the type of network she lives in.

Let $A = q_1r + (1 - q_1)l$ denote the expected return if the receiver is part of an honest network and $B = q_2r + (1 - q_2)l$ the expected return if she is drawn from a cheater network. Given an initial endowment of $x = 1$, we assume that $A > 1$ and $B < 1$, so that the expected return is positive if the population is honest and negative if it is not.

Let h and nh represent the true distribution of honest people (h) and cheaters (nh) in one's network. In line with this set-up, we assume that individuals who do not invest (do not interact) do not learn about the true distribution of h and nh people in their network. People choose not to interact in the first period if their

prior is that they live in a cheater network. In a two period game, this then means that only people who have interacted in the first period will interact in the second, since they are the only ones who are able to update their beliefs.

Given these assumptions, an individual who interacts in the first period and finds out she lives in an honest network will thus always interact in a potential second period, since the expected value of $A > x$. However, if she finds out in the first period that she lives in an cheater network, she will not interact in a potential second period since $B < x$.

This framework is useful to keep in mind when looking at the present case of social network interaction and demonetization. In line with this framework, the first period in which the individual decides whether or not to interact with her network based on her priors in our case aligns with the time period just after demonetization. Following this shock, individuals decide to interact more only if they believe that they live in an honest network.³ We thus expect to see that after the demonetization shock, people who have the prior that they live in an honest environment will interact more with their network and thereby in a next step learn about the true distribution of honest people h and cheaters nh in their network. As a result of this interaction, the individual then updates her priors if necessary. If, contrary to her initial beliefs, it turns out that she lives in a cheater network, this will lead to a reduction in measured levels of trust; if her initial belief of living in an honest network is confirmed, her measured levels of trust should increase.

In line with the model, we expect this increase in interaction to stem from those who already used networks before. Further, in our context, these interactions could be heavily defined by the social hierarchy in place. As such, we do not necessarily expect a similar outcome for different castes.

3.4 Data and descriptive statistics

3.4.1 Description of the survey

This paper is based on a novel data set from rural Tamil Nadu, entitled Networks, Employment, Debt, Mobilities, and Skills in India Survey (NEEMSIS), which was conducted in 2016/2017.⁴ The survey was collected over two periods, first from August 2016 to early November 2016 and then from January to March 2017.⁵

The survey was collected in 10 villages in the Cuddalore and Villupuram districts

³Alternatively, the shock could have lowered the trustworthiness threshold at which an individual is willing to engage. As will be shown later in this paper, this should have led to individuals listing a larger amount of potential ties, which is not the case.

⁴The survey was collected by a team of IRD and IFP researchers, including the authors of this paper. More information can be found on <https://neemsis.hypotheses.org>

The 2016/2017 survey is based on the structure of the 2010 Rural Employment and Microfinance (RUME) program, creating a household panel (2010-2016/2017). For this paper, we will only make use of the second wave (2016/2017).

⁵The break in the survey was unrelated to environmental factors (demonetization).

of Tamil Nadu in an economy dominated by agriculture⁶ but benefiting from the proximity of two large industrial towns (Neyveli and Cuddalore) and a regional business center (Panruti) (see Figure 3.1 for a map of the geographical area of data collection). The survey uses a stratified sample framework according to first agro-ecological considerations (dry/irrigated agriculture in villages), then urban proximity, and lastly social groups (caste representation). The caste representation was based on self-classification of individuals into castes using local terminologies, which were then categorized into three main categories (Dalits, middle castes, upper castes). The two largest caste groups in the region are Vanniyars and Paraiyars, the former classified as a middle caste⁷ and the latter one of the major Dalit communities⁸ in Tamil Nadu. Despite the Vanniyars' traditionally rather low rank, they are land-owners in the region studied, dominating politically. The upper caste group in the studied zone consists of Mudaliyars, Chettiars, Naidus, Reddiyars, Settus, and Yathavars, who make up only a small proportion of the village populations. In each village, the sample was then determined to stem half from the Ur part of the village, in which mostly upper and middle castes live, and half from the Colony part of the village, which contains mostly Dalits.

The NEEMSIS consists of comprehensive household and individual level modules, completed by the household head, and a randomly chosen younger member of the household (older than 18 and younger than 35). The total sample size of the individual survey is 952 individuals. This individual-level survey provides more detailed information on labor force participation, labor outcomes, and social networks, alongside a cognitive and a non-cognitive skills assessment. The cognitive skills assessment includes Raven's Colored Matrices⁹, which have been previously used in cognitive skills assessment in low-literacy populations in developing countries (e.g. Serneels, 2008); a literacy test (four questions); and a numeracy test (four questions). The non-cognitive skills self-assessment consists of a Big Five questionnaire, including questions about the Big Five personality traits as well as about grit.¹⁰ The language in the question set was adjusted to accommodate a

⁶The sowing and transplanting season takes place from September to December and the harvest season is from January to March. This means that our first sample (pre-demonetization) was interviewed during sowing seasons and the second sample (post-demonetization) during harvest season. This is potentially problematic as more work is available during the sowing season since harvesting is done mechanically. However, we do not actually observe any significant differences in employment shares between the pre- and post-demonetization samples (see section 3.6.4).

⁷Additional middle caste groups present in the region of our survey are Padayachis, Gramanis, Navithars, Nattars, Kulalars and Asarai.

⁸A few Arunthathiyars who are part of the Dalit community are also present in the region.

⁹Raven's Colored Progressive Matrices (CPM) are a cognitive, visual, non-verbal test that does not require any level of formal education. It captures the ability to think and make sense of complex data and logical reasoning. The CPM consists of 36 questions of increasing difficulty.

¹⁰The Big Five are a list of key traits capturing the broadest level of personality traits. These are: (1) openness to experience, which captures one's tendency to be open to new experiences; (2) conscientiousness describes one's tendency to be organized, hardworking, responsible; (3) extraversion, which encompasses directing one's interest towards the outer world of people and things; (4) agreeableness, the tendency to act cooperatively and in an unselfish manner; and (5) emotional stability, the predictability and consistency in emotional reactions with absence of rapid mood changes. The survey further includes

low-literacy population and a careful translation to local Tamil was developed after numerous discussions and tests among the survey team including local enumerators. The social networks module includes information about membership in associations (e.g. self-help finance groups, village councils, sports groups) and detailed information on actual and potential interactions with others, as is explained in more detail in the next section.

3.4.2 Construction of the social network variables

Using the detailed social networks module of our survey, we construct two different social network variables, capturing interactions of individuals in our data set with a variety of actors. Interactions in our data cover formal and informal social capital as well as actual and potential interactions. Formal interactions include membership in associations (such as a farmers' association). Informal interactions include all sorts of social connections that an individual may have made. The data on interactions was collected using a name generator which was included as part of the individual survey. The name generator follows sociological research approaches (McCallister and Fischer, 1978) and invites the respondent to recall and elicit people ('alters') with whom they maintain certain types of direct relationships in order to delineate the core members of the network (Marsden, 2005). These include borrowing from and lending to each other, helping others or seeking help in finding work, relying on connections for help with a business or supply of tools, and so on. As part of this name generating process, we also collected background information on these alters (such as caste, age, gender, education) and on the relationship between the survey respondent and alters. As we only have a single measure of formal social capital (number and types of associations of which an individual is a member), we will not consider differential effects of formal and informal social networks; instead, we combine both into composite measures.

The actual ties refers then to links an individual has explicitly made. This includes having borrowed or lent money to others, recommended somebody for a loan (or received a recommendation from somebody), recommended somebody for a job (or received a recommendation), or received help with a loan. The potential ties consist of all connections that an individual could use if the need occurred. This includes questions regarding whom the individual would ask for help if in need of information, help with the business, help with finding a job, or recruiting workers. It also includes household size, counting family members 15 and older only, as the survey does not allow for family members to be included in the borrowing/lending links mentioned before.¹¹

grit, which is the tendency to sustain interest in long term goals and persistence.

¹¹While we already capture quite a number of interactions, this social network data set is by no means a complete representation of reality. We are relying to a large degree on interactions of an economic nature (financial practices and labor interactions), without being able to capture an individual's full network. For example, while we are including loans taken out for marriage as part of the actual ties, the pure growth in one's potential interactions due to the merging of two families cannot be taken

Lin (2001) defines the structural foundation of social capital as “resources embedded in a social structure that are accessed and/or mobilized in purposive actions” [ibid, p.40]. In this framework, our potential ties variable would refer to something akin to the resources component of social capital, the part that is potentially accessible to an individual through her social network. Our actual ties variable then relates more to concepts of activation, accessibility, and mobilization. These are the resources that an individual can access not only in theory but in practice. We use the network data to construct two different measures of social networks, which will be our main independent variables of interest for this study.

The first measure of social networks that we will look at is *total network size*, which is the sum of the ties that we observe.

$$size_i = a_i + p_i,$$

where a_i are the actual and p_i the potential ties of individual i .

The rationale of this network variable follows from Johny et al. (2017) who consider intra-village social networks in poor rural areas in Kerala, a state in South India, and find that the number of connections a household has is more important than alternative measures of network centrality such as degree or eigenvector.¹²

The second measure of social networks that we consider relates to *network usage* rather than of pure size. We define network density, the share of connections used as follows:

$$density_i = \frac{a_i}{a_i + p_i},$$

where a_i are the actual and p_i the potential ties of individual i .

Thus, either definition captures a different idea within the broader concept of social networks. Given their different definitions, we do not necessarily expect them to influence trust measures in the same way. Based on our conceptual framework described in Section 3.3, we would expect $size_i$ to influence trust positively: if, after the demonetization shock, an individual still elicits a large number of ties, this means that the individual, who increases interactions as a result of demonetization, was confirmed in their belief to be living in an honest environment, or at least confirmed in their belief of who can be relied on. We would expect $density_i$ to potentially have a negative effect on measured levels of trust. A higher value on the $density_i$ measure means that individuals have to use their networks more intensely; this includes making resort to ties that would not usually be activated. If the individual has to rely on weaker ties, this could suggest that their belief of living in an honest environment was rejected, as closer ties were unwilling (or unable) to

into account given the structure of the data. We only capture the size of the survey unit, the nuclear household.

¹²Eigenvector centrality is a measure of the influence of a node in the network. It takes into account the number of neighbors, but also the importance of those neighbors, i.e. whether those neighbors are themselves central to the network.

help. As a result of higher network usage, we might then expect lower levels of trust.

3.4.3 Measuring trust

We use three different measures of trust, which are all related to interpersonal trust, i.e. trust in other people:

1. People in my neighborhood can be trusted.
2. Among employees, kin members are more trustworthy than non-kin members.
3. Are you generally trusting of other people?

Trust in neighborhood

Villages in rural South India are highly segregated by caste: middle and upper castes tend to live in a part of the village called ‘Ur’, while lower castes, Dalits, tend to live in the ‘Colony’. Upper castes tend to live alongside middle castes in ‘Ur’. These parts are oftentimes separated physically. In several survey villages, for example, Ur is located on one side of a cross-country road, while Colony is located on the other. Neighborhoods in the study region can thus be highly homogeneous in terms of caste membership (especially ‘Colony’) and the socio-economic status of their inhabitants. This is a common finding in India, where spatial segregation leads to a high level of local social connectedness within caste networks (Munshi, 2016a), thus leading to closed intra-group and weaker inter-group relations. Caste groups within villages are usually big enough to support a local community, which would then foster ties with other villages through intra-caste marriages. Munshi (2016a) find that there are on average about 30 different castes per village; in our survey, we can distinguish between 3-8 different castes. As a result of the social segregation, one would thus expect a country like India to score highly on questions about trust in neighbors. Using the World Values Survey, Munshi (2016a) show that almost 90 percent of people in India say that they trust their neighbors. In measuring trust with the question ‘People in my neighborhood can be trusted’ it is thus important to keep in mind the different reference points for the different castes.

Trust in kin vs. non-kin among employees

The second measure of trust refers to an even closer in-group (kin) versus the out-group (non-kin). South India has tight kinship structures, which has been negatively correlated with out-group cooperation in other countries (Herrmann et al., 2008). The question ‘Among employees, kin members are more trustworthy than other non-kin members’ relates to the distinction of kin vs. non-kin in a very specific environment: the workplace. In South India, labor contractors are an important intermediary (“maistries”); they are the middlemen between, for example, the

contractor responsible for painting a house and the oftentimes unskilled laborers who carry out the work. These middlemen are primarily responsible for finding the laborers to work on the project, including making sure the laborers show up to work and assuring that they have done quality work. Because castes are traditionally occupational units, choosing the right laborers is important for the labor contractor, who therefore oftentimes resorts to hiring his own kin due to informal mechanisms of ensuring quality work (Munshi, 2016a).

Generalized trust

The last question with which we measure trust is most closely related to measures typically used in surveys, such as the World Values Survey. The question ‘Are you generally trusting of other people?’ is not specifically related to the context of rural South India, but it is thought to capture the concept of generalized trust. It is also the most difficult question to answer and use for the analysis, though, as trust is in and of itself dependent on circumstances (Nooteboom, 2007); for instance, one might trust someone in one condition but not in another seen as beyond that person’s level of competence. Generalized trust is oftentimes understood as a broader definition of trust, placing more weight on trusting people beyond the local community.

Answers for all three questions were recorded on a Likert answer scale ranging from completely disagree to fully agree. A Likert scale was chosen to elicit answers to prevent problems with ambiguous wording (Miller and Mitamura, 2003).¹³ In general, as they refer to survey questions and not results from a trust game, all three measures capture a combination of people’s beliefs about others’ trustworthiness, betrayal aversion, etc. (Fehr, 2009; Glaeser et al., 2000). They will be standardized for ease of interpretation in the regression analysis.

The three different questions were chosen as they all represent different aspects of trust in others that are important in the context of rural South India.¹⁴ We decided to keep all three measures of trust separately instead of combining them in an index, as the literature notes the importance of distinguishing between trust in different actors (e.g. Haddad and Maluccio, 2003). This is particularly important in India, as it relates to both the specific cultural context structured by high levels

¹³Miller and Mitamura (2003) examine trust questions included in the World Values Survey and find that reducing answer possibilities to a simple agree or disagree can lead to conflicting answers and misinterpretations regarding the concept actually measured, which related closer to a measure of caution than levels of trust.

¹⁴All three measures are related to prosociality. We looked at other correlates of prosociality in our data, i.e. facets of the Big Five traits agreeableness, openness to experience, and extraversion. Conducting factor analysis over the entirety of the Big Five questionnaire and our trust measures to see whether other questions are in fact very closely related to the three questions chosen, we find that all of the three trust measures used in this paper load on the same factor. The only other question loading on that factor is “Do you enjoy being with people?” which reflects strongly on positive behavioral dispositions to others. While this is certainly related to trust, trust refers more clearly to a belief rather than a social preference towards social interaction. We therefore decided to proceed with the three questions described in this section.

of social segregation and the context of the shock that is used for identification. Indeed, one would expect the demonetization shock to primarily foster interactions locally, which might not translate to any effects if measured by a broad question regarding generalized trust in people. Further, all three measures are purely related to trust in people. This is important to keep in mind, as measures of trust that are related to more formal institutions, such as trust in banks or trust in associations, could potentially be directly influenced by the demonetization policy. Indeed, as will be explained in more detail in section 3.5.2, it seems unlikely demonetization had a direct effect on trust in people directly.

3.4.4 Descriptive statistics

Table A3.1 provides descriptive statistics of the main variables used in the analysis. The sample is restricted to those with non-missing cognitive and non-cognitive skills values. The individuals in our sample are on average 42 years old. A little more than half of the sample is male and most people (about 82 percent) are married. The survey uses a stratified sample based on caste, with about half belonging to the lowest castes (Dalits). The other half belongs predominantly to middle castes, with only a few people (10 percent) identifying as belonging to upper castes. Education in the villages covered is still low: 38 percent of villagers did not complete primary education and another 21 percent stopped after completing primary. The education variables hide important generational differences, though, as younger generations have shown fast improvements in educational attainment. Indeed, the vast majority of people with at most primary education is concentrated among those aged 40 and older, while less than 4 percent of the 19-29 years old have completed less than primary education. Similarly, almost all of the educational attainment above a secondary school degree (“10 Standard”) is concentrated among the young, of whom 21 percent have obtained a Bachelors degree.

In addition to educational attainment, we included more objective measures of cognitive ability in the survey (the Raven test) as well as measures of functional learned ability such as literacy and numeracy. On average, individuals answered only slightly more than 13 out of 36 questions of the Raven’s test correctly. Interestingly, the vast disparities by age group in terms of educational attainment cannot be observed in the Raven test; while younger people (18-29 years old) perform better than older ones, the differences are small and not statistically significant (i.e. the younger cohort answered on average 14 questions correctly versus 11.5 questions among those 45-60 years old). The numeracy and literacy questions were set up to test basic, primary school-level knowledge. The low means (less than 2 correct answers for each) reflect the oftentimes poor quality of education in rural India. Measures of non-cognitive skills were included to capture the multidimensionality of skills. Individuals in our sample score highest on the trait conscientiousness

and lowest on openness to experience.¹⁵ Our measure of social network density illustrates that individuals use on average only 15 percent of their social network and that, on average, they have a total of 8.22 ties in their social networks.

Trust in the sample is high: on average, individuals score about 3.9 out of 5 on the question asking whether neighbors can be trusted and 3.45 out of 5 on the question on whether kin can be trusted more than non-kin. Generalized trust is slightly lower at 3.2. This is in line with other surveys that also find very high levels of trust in neighbors in India (as cited in Munshi, 2016a).

3.5 Empirical strategy

3.5.1 OLS

In order to estimate the effect of social networks on trust, we start from a basic OLS regression:

$$Trust_i = \beta_1 SN_i + \beta_2 X_i + \varepsilon_i \quad (3.1)$$

where $Trust_i$ represents our outcome of interest, different measures of trust, for individual i . SN_i captures the social network of individual i (total network size or network density), X_i is a vector of individual and household control variables that is thought to affect the level of trust, and ε_i is the error term, capturing any remaining individual heterogeneity.

X_i includes among other things information on individual i 's cognitive ability and personality traits, as personality traits have been shown to affect levels of trust (Dohmen et al., 2008; Freitag and Bauer, 2016). This paper is special in its ability to include measures of individual cognitive and non-cognitive ability in a rural developing country setting. Assuming consistent measurement of these traits, we can thereby include variables into the regression that are usually part of the unexplained individual heterogeneity captured by the error term.

3.5.2 Instrumental variables

The correlation captured in the previous section is likely to suffer from endogeneity bias. For example, if we happen to find a positive relationship between social networks and levels of trust, this could be consistent with our hypothesis that social interactions foster trust, but the correlation could also be explained by people who

¹⁵In the analysis, we first correct non-cognitive items for acquiescence bias, i.e. the tendency to answer more in one direction (agree or disagree) over the other and then aggregate and standardize the traits. The acquiescence score for the sample is 2.84, meaning that given the 5 option Likert scale, slight acquiescence is present in the sample, with individuals more likely to disagree with a statement than to agree. Cronbach's α , a measure of internal consistency of a construct, are mostly at or above the desirable value of 0.7. The value of α per trait in ascending order are: 0.60 (agreeableness), 0.61 (extraversion), 0.68 (grit), 0.77 (emotional stability), 0.78 (openness to experience), and 0.85 (conscientiousness).

are by nature more trusting forming larger and more extended social networks. Social network could thus be an endogenous variable. In order to estimate the causal effect of networks on trust, we make use of the demonetization shock as a source of exogenous variation that affects social networks but does not affect trust in other people directly.

Using demonetization as a source of exogenous variation for our study is possible because about two-thirds of our sample was interviewed before (November 2016) and the other third about two months after (January - April 2017) demonetization had occurred. The chronological sequence of household data collection was almost random, or at least had no obvious and systematic collection plan across the 10 villages. As such, around two thirds of the first subsample had not experienced the sudden demonetization shock when we interviewed them; the other third experienced the shock and may have used their networks to cope.

The framework is as follows:

$$SN_i = \alpha X_i + \gamma D + \mu_i \quad (3.2)$$

$$Trust_i = \alpha X_i + \beta \widehat{SN}_i + \varepsilon_i \quad (3.3)$$

where D is a dummy variable, taking the value of 1 for individuals who have been interviewed after demonetization and 0 otherwise; X_i is a vector of individual and household control variables that is thought to affect the level of trust. This includes information on individual i 's cognitive ability and personality traits, \widehat{SN}_i is the predicted value of SN_i , our measures of social networks, recovering an exogenous measure of SN_i .

For any IV strategy to identify the local average treatment effect (LATE) consistently, the instrument must satisfy two conditions: (1) it must be correlated with our measures of social networks, and (2) it must not be correlated with μ_i , thus it must not be correlated with factors directly affecting levels of trust. Failure to satisfy these conditions can lead to inconsistent estimates, asymptotic bias, and large standard errors (Bound et al., 1995; Wooldridge, 2010).

Relevance

Our data show that interaction increased as a result of demonetization. This is in line with our conceptual framework described in Section 3.3, as we expect to see that after the demonetization shock, people who have the prior that they live in an honest environment will interact more with their network and thereby in a next step learn about the true distribution of honest people h and cheaters nh in their network. Data on lending behavior, for example, shows that while only 5 percent of individuals in our sample claim to have lent money to anybody before demonetization, this figure jumps up to 11 percent among those interviewed after

demonetization.

Guérin et al. (2017) provide a first overview of how individuals in our study region coped with the sudden shortage in cash, suggesting that individuals had to rely on their networks more than they usually do. This holds for both richer and poorer individuals: the better off made use of their networks to dispense of old and now invalid notes, enabling them to prevent having to endure the long lines at banks and to cash in potentially illegal notes, while poorer individuals relied on their networks for informal loans to cover shortages in wages. Qualitative evidence supports the view that those who are part of a supportive network made use of it to cope with the shock and were able to mitigate its risks, while those who did not have the ‘right’ networks suffered. The first example below describes the small business of a woman who belongs to a well-connected family and the effect of demonetization on that business.

Bargath (female, 32 years old) sells chicken from her home in the Ur part of one of the villages. Bargath is part of a dynasty of chicken vendors: both her father and grandfather were involved in the same business. Her brothers are still involved in chicken farming and selling, though neither of them lives in the same village. Bargath sells mostly to customers from the Colony part of the village. Bargath’s father taught her never to sell on credit - a guiding principle she has employed in her small business. When asked whether she had experienced any change in her business dealings as a result of demonetization, Bargath replied that for her, demonetization did not have any effect whatsoever, as she could rely on her extended family members. She went on to explain that her supply of chicken had not changed as it came directly from her brothers. Further, she was able to accept “old” 500 rupee notes, as her brothers would then take care of exchanging the money for her. She continued not to sell for credit; however, as she was able to accept notes that were officially no longer legal tender, her customers remained able to pay her.

The second example describes a shop owner who was unable to deal with the demonetization shock through networks and instead had to resort to a loan from a moneylender.

Saleem Basha (male, 41 years old) runs a small local grocery shop. Following demonetization, he had to start selling goods on credit since customers did not have any cash at hand. He further had to take out a loan from a moneylender in order to buy supplies for his shop. In his opinion, if he did not take out a loan in order to continue offering goods, customers would take their business elsewhere and not return.

The examples above illustrate the role that social interactions had in coping with the unexpected demonetization shock. While they describe the mechanisms at

hand, they are also not fully representative of the sample population: in general, women in the study area are less able to access resources (for example, in our data set, the majority of loans have been taken out by men).

Exclusion restriction

The exclusion restriction requires that the instrument (demonetization) does not correlate with factors directly affecting the outcome (trust in people) other than through its impact on social network variables and that the instrument should be close to random assignment. The instrument only affects trust through its effect on social networks.

First, conceptually, the component of trust that we think we measure and that could be changed in a rather short period of time (about 2-5 months had passed between the demonetization shocks and the interview) is not necessarily people's preferences, but rather their beliefs about others' trustworthiness (Fehr, 2009). It seems likely that demonetization changed these beliefs only through the fact that demonetization increased the likelihood of *interacting* with others. This is exactly the framework that we have in mind and described in Section 3.3: only those who invest (interact) learn about the true distribution of h and nh people in their network. Therefore, only those who interact will update their priors about others' trustworthiness, leading to an increase in measured trust if the individual's belief of living in an honest environment is confirmed and a decrease in trust if her initial belief is rejected.

Second, individuals themselves did not think that demonetization as such had a direct impact on their levels of trust. A short additional demonetization module was administered to those who were interviewed from January - April 2017 (after demonetization). This module contained questions about whether or not individuals think demonetization influenced their answers. The question was asked after the answers to the other questions had been elicited and should therefore not frame the answers to the trust questions themselves, meaning they remain comparable between the pre- and post-demonetization samples. Table 3.2 presents answers to these questions. It becomes clear that most individuals did not think demonetization changed their levels of trust. Among those who did experience a change, the share of individuals experiencing a positive or negative change is almost the same for neighborhood and employee trust.

Looking closer into changing trust due to demonetization lets us draw out two interesting observations: first, 78 percent of those who claim not to have experienced a change in the trust questions also claim not to have had to ask anybody else for help because they did not need to. No change in trust levels thus correlates with no additional social interactions. Second, those who claim that demonetization affected their levels of trust (either positively or negatively) also reported having interacted more, whether through asking others for help

Table 3.2 – Change of trust in:

	Neighborhood	Employees
Increase	10.6	11.8
Decrease	15.0	14.3
No change	74.4	73.9
N	273	272

Source: NEEMSIS (2016-2017); authors' computations.
Note: Question asked to post-demonetization sample only.

(12 percent), through realizing there was nobody there to help (34.5 percent), or through asking but being refused help (8.6 percent). Again, this suggests that the effect of demonetization on trust only acts through the channel of social networks. In the IV framework, what we are estimating is the LATE: the average effect of X on Y for those whose treatment status has been changed by our instrument. We are thus identifying the effect of a social network on trust – the underlying research question of this paper – among those who interacted as a result of the treatment (demonetization).

Third, the proposed instrument should be as good as randomly assigned across the 10 surveyed villages. The chronological sequence of household data collection did not follow any systematic collection plan in the sense that we did not start our data collection in the poorest or richest villages, nor in the ones closest to, or furthest away from, the regional hub (Panruti), which could arguably have significantly altered the composition of the pre- and post-demonetization sample. Table A3.2 depicts descriptive statistics of the individuals interviewed by timing of interview (before or after the demonetization shock). Despite demonetization being *a priori* as good as randomly assigned, Table A3.2 shows that this does not hold in practice. Indeed, a Hotelling's T-squared generalized means test rejects the hypothesis that both samples are equal. We will therefore use matching based on covariates to balance the pre- and post-demonetization samples.

Balancing the pre- and post-demonetization samples

Despite the demonetization shock falling randomly into our survey collection time schedule, the previous section has shown that there are significant differences in the pre- and post-demonetization sample. We therefore use matching techniques to balance the samples. Given the rich nature of the data collection, we can match based on a number of covariates that could influence the outcome, including individual characteristics such as personality traits or cognitive skills, age, marital status, education, gender, and caste. We further match based on household characteristics that could affect the outcome, most notably consumption (food expenses, health expenses, ceremonial expenses), household income, and characteristics of the

household's dwelling (access to electricity, water, sanitation, and type of house).¹⁶ In total, we are matching based on 12 individual characteristics and 8 household characteristics. We use nearest neighbor matching and restrict our sample to those for whom we have common support (see Figure A3.1). We use full covariate matching instead of matching based on the propensity score for several reasons: matching on covariates is usually better in terms of asymptotic efficiency (Angrist and Pischke, 2009); our data set includes a large set of covariates for matching, including some individual characteristics such as personality traits and cognitive skills that are oftentimes considered to be part of the unobservables; and the process of matching on observables requires the researcher to focus on the covariates determining outcomes (trust in our case) and choosing the appropriate covariates to match on. While this latter reason could be prone to error, we still have a better idea of what could determine trust than what could determine treatment assignment (being interviewed before or after the demonetization shock), since in our case treatment was not based on certain individual characteristics such as age or gender. Instead, and while this is not fully reflected in the unmatched sample, from the point of view of the data collection, treatment was essentially distributed randomly. Matching reduces our analysis sample from 885 to 663 individuals for whom we have common support. Importantly, we manage to match almost all of the individuals from the treatment group (238 out of 255 individuals in the unmatched sample).¹⁷ Table A3.3 displays the matched sample and shows that differences between the post- and pre-demonetization samples are no longer statistically significant for the covariates that we matched on. Conditional on matching, we thus conclude that the demonetization shock is as good as randomly distributed.

LATE framework

The previous sections have shown that our instrument is relevant and as good as randomly assigned. This assumption holds conceptually, as the demonetization shock was both unexpected and implemented uniformly across the country instantaneously, as well as empirically, conditional on matching on covariates. Given these necessary conditions, it is important to note that the effect that we are estimating is likely to be a Local Average Treatment Effect (LATE), as we are

¹⁶Matching on personality traits is based on the assumption that differences in personality traits between the pre- and post-demonetization samples are due to us interviewing fundamentally different people and not due to any direct effect of demonetization on personality traits. As individuals interviewed before and after do significantly vary in their non-changeable characteristics such as gender, educational attainment and (to an extent) age (see Table A3.2) and as personality traits seem rather stable among adults and only slightly related to adverse life events (Cobb-Clark and Schurer, 2012), this assumption is not unreasonable. We also match without the cognitive and non-cognitive skills variables. This reduces our ability to control for individual heterogeneity, which might be important, especially with regards to trust formation. Results hold in coefficient sign and mostly in significance.

¹⁷17 treated individuals fall outside the area of common support and are therefore excluded from the final sample when we restrict it to those with common support.

covering the effect on the compliers, i.e. those who adjusted their interactions as a result of the treatment (demonetization). While everybody interviewed after January 2017 did by definition live through demonetization, not everybody reacted to it in the same way. In the results section, we therefore consider heterogeneous effects by characteristics that we think could influence somebody’s chances of being a complier, such as caste category and gender. Given the specificity of the setting, the results obtained in this study are internally valid, but they are unlikely to be applicable to other settings that do not have the same strict social hierarchies, which fundamentally determine the type of social interactions that are possible.

3.6 Results

3.6.1 OLS estimates of the determinants of trust

Table A3.4 depicts our first results, separate OLS regressions with the different measures of trust as the dependent variables (trust in neighbors, trust in kin among employees, and generalized trust). All three trust measures and both measures of social networks (size and density) have been standardized for easier interpretation. The OLS regressions are based on the balanced samples obtained from matching on covariates, as described in the previous section, and are weighted by the matching weights retained. In the regressions, we control for a variety of potential individual and household determinants of trust: age, gender, being married, caste membership, educational attainment, the standardized score on the Raven’s test (a cognitive test), the standardized scores on the numeracy and literacy tests, and standardized and acquiescence corrected personality traits (Big Five and Grit) for individuals; and household expenses on food, health, and ceremonies (to control for a consumption effect as a result of demonetization) as well as household income for households. We further include village-area fixed effects to capture village-specific heterogeneity and cluster standard errors at the village level (the highest level and the level at which treatment occurs as suggested by Cameron and Miller (2015)).

Table A3.4 shows that total network size is positively correlated with agreeing that, among employees, kin can be trusted more than non-kin (column (4), significant at the 1 percent level); it is also positively correlated with more trust in people in general and less trust placed in neighbors, though the last two are not statistically significant. Network density shows the opposite picture: those who have a higher network density (that is, those who use more connections as a share of their total connections) are more likely to trust neighbors and less likely to prefer kin over non-kin, but are also less likely to trust people in general. None of these effects are statistically significant. While these analyses are probably subject to bias due to the endogeneity of the social network measures being both the cause and consequence of trust, it already illustrates that network size and

network density are distinct concepts that do not necessarily have the same effects.

Looking at other correlates of trust, middle and upper castes are significantly more likely than lower castes to agree that neighbors, and people in general, can be trusted, and that kin employees cannot be trusted more than non-kin employees. Compared to the largest educational category (no completed primary education), people from the second largest educational category, those with some high school education, are significantly more likely to agree that those in their neighborhoods can be trusted and that kin can be trusted more than non-kin.

Lastly, non-cognitive skills are significant determinants of levels of social trust, but the effects vary between the different measures of trust, illustrating again that while the three different measures of trust seem similar, they capture different underlying concepts. We can compare our results for the personality traits to the literature to see if our determinants are similar. This is possible for the generalized trust measure (columns 5 and 6), which is most similar to the trust measure used in the literature. Indeed, the signs for the personality traits in the OLS that we obtain are similar to the results obtained from OLS by Dohmen et al. (2008) for a German sample: we find that openness to experience, extraversion, and agreeableness have a positive correlation with trust. Our coefficient for emotional stability is negative, which is the opposite of Dohmen et al. (2008), who find that more emotionally stable people are in fact more trusting. While many trust measures are not significant in the OLS, this does by no means mean that the hypothesized relationship between network measures and trust does not hold. Instead, the OLS results are likely to be biased precisely due to the endogeneity of our network measures. We therefore make use of the demonetization shock in the next section to overcome issues related to the endogeneity of the network measures.

3.6.2 First stage results: determinants of network size and density

Columns (1) of Tables 3.3 and 3.4 depict the first stage regressions for our network measures network size and density, respectively. Both show that the demonetization shock is a strong predictor of our network measures. The F-statistic for the first stage of network size is 14.04 and 41.2 for network density. Both values are above the F-statistic of 10 recommended by Stock et al. (2002). The first stage coefficients make sense intuitively (as described below) and reduced form regressions¹⁸ of the outcome variables (trust measures) on the instrument (demonetization) show the same relationship. Further, our IV estimation is identified with only one potential instrument, which makes it median-unbiased and less subject to the weak instrument critique (Angrist and Pischke, 2009). Therefore, we believe that demonetization is a strong instrument with which to proceed.

The experience of demonetization decreases network size by more than half a standard deviation (0.64) on average for an otherwise equal sample (thanks to

¹⁸Results not reported, but available upon request.

the matching), and it increases network density by about a third of a standard deviation (0.36). As mentioned earlier, network size is the sum of all actual and potential connections an individual has either made (mostly related to financial and labor market practices) or claims that she can make if necessary (the potential connections). Table A3.3 shows that both the amount of actual ties that an individual has made and the amount of potential ties changes with demonetization, with the number of actual ties increasing (marginally) and the amount of potential ties decreasing (more strongly). The mechanism at hand thus seems driven by potential ties, which relates back to the conceptual framework described in section 3.3; the demonetization shock forces individuals to interact with their networks to cope, which reveals information about their network. In a first step, individuals then update their beliefs about their networks. There is some suggestive evidence of this in the data: among both Dalits and middle castes, those who lived through demonetization and answered that they did ask somebody for help as a result of the shock name more potential ties than those who claimed that there was no one to ask. This could suggest that those who were given help updated their beliefs about who they could turn to in times of crisis upward, while others might have realized that there was nobody to help and updated their estimate downward instead. Density increases with demonetization; given the previous results on size, this suggests that instead of expanding their networks, individuals might have tried to use their existing networks more intensely. Given the strict social hierarchies and spatial segregation in rural South India, expanding one's network to new actors might have been simply unfeasible, as network size is essentially determined by the size of one's caste community.

Turning to the covariates, being a woman is associated with a smaller total network in terms of size and a lower network density. While both women and men are involved in borrowing and lending, men do both much more frequently. In our sample, only about 25 percent of those having taking out a loan are female. Further, men predominantly borrow from other men (about 90 percent); women tend to mostly borrow from men, with only a minority borrowing from other women. Due to the inherent exclusion of women from the financial system, we might thus not capture a woman's coping network fully. Still, we think that our network variable could capture at least some part of a woman's interactions with her social network as a result of demonetization: the main reason given for taking out a loan after demonetization is family expenses (such as food), which is also the main reason women take out loans according to our data.

Higher educational attainment correlates positively with network size, as individuals might have made additional connections through educational institutions. It correlates negatively with network density, though: those with a Bachelor's or postgraduate degree might have other forms than social networks to cope with shocks (more access to formal banking, for example). Further, a higher literacy

score is associated with a smaller network size and a larger network density; a better score on the numeracy test is related to a smaller network density. These two variables can capture a variety of things, from actual knowledge of the subject to motivation while taking the test. The Raven’s test score, which was also included, is not significant in either specification. We hold fluid intelligence (through the Raven’s score), educational attainment, and personality constant, so the literacy variable would then capture an additional concept, such as actual knowledge of reading and writing, or willingness to complete the full test (motivation, which would in part be captured by the conscientiousness trait). Conscientiousness does carry the same sign in both cases, but it is statistically insignificant. One could thus imagine literacy representing some form of motivation and diligence, which could relate to individuals updating their beliefs and behaviors more readily as a result of the shock. Finally, the variables capturing individual heterogeneity (the non-cognitive skills variables) show that, holding all other personality traits constant, greater emotional stability is associated with a smaller network and making use of that network less intensively. Some household characteristics are also significant, with higher food and health expenses, as well as higher household income, related to having a larger network. Household income is also related to a smaller density, as richer households are more able to cope based on their own resources.

3.6.3 Second stage results: the causal determinants of trust

Columns (2) through (4) in Tables 3.3 and 3.4 depict the second stage results, estimating the effects of the predicted network measures, \widehat{SN}_i , on the three different measures of social trust after correcting for potential endogenous network formation and use. Network size causally increases trust in kin vs. non-kin among employees: a one standard deviation increase in network size, which increases total network size from a sample average of almost 8 connections to 12 connections, increases trust in kin by 0.44 standard deviations. This is essentially equal to moving from answering somewhere between ‘sometimes’ to ‘quite often’ to solidly answering ‘quite often’ to the question “kin members are more trustworthy than non-kin members among employees”. Network size further has a similarly large but opposite effect on trust in neighbors: those with larger networks are more likely to say that they trust their neighbors less, though the effect is only significant at the 10 percent level.

Network density has a negative effect on trust in kin among employees, however: a one standard deviation increase in network density, which doubles density from a mean of 15 percent of used connections to about 33 percent of used connections, decreases trust in kin by 0.8 standard deviations. Analogously, this is equal to moving from being between ‘sometimes’ to ‘quite often’ to answering that “kin are ‘rarely’ more trustworthy than non-kin”. In line with previous results, a higher network density increases trust in neighbors. Results are in line with the OLS

regressions in Table A3.4 in terms of sign of the coefficients, but they are now significant (at the 5 and 10 percent levels). Compared to the OLS coefficients, the second stage IV coefficients for trust in neighbors and trust in kin employees (due to both network size and density) increase quite substantially. This suggests that network size and density can indeed be considered endogenous.¹⁹

These coefficients suggest that as networks get larger, people seem to place more trust in their kin than non-kin, while as usage gets denser, people place less trust in their kin (as compared to others). These explanations are not necessarily contradictory. Network size, as it is defined in this paper, does not relate to network usage. Instead, as has been illustrated in the first stage discussion, the effect of demonetization on network size is to a large degree driven by changes in the number of potential ties, due to an updating of information about who could help in times of crises. If updating has led someone to re-estimate their number of potential ties downward, they are probably insecure about whom they can really rely on. Kin, given tight social structures in rural India, seem to be a reliable option. Neighbors, though, might not be. Neighborhoods are quite homogeneous in the ‘colony’ part of the villages, which is predominantly occupied by Dalits. They are less homogeneous in the ‘ur’ part, in which middle and upper castes live side by side. As we will see later in this paper, the negative effect of network size is effectively driven by middle castes. The effect seems to be the same as for trust in kin: as network size grows, individuals are more weary about those at the weaker ends and tend to trust those more similar to themselves (kin, neighbors in homogeneous environments), which is essentially the homophily principle.

Network density, unlike network size, represents the share of used connections over all connections. The story here seems to be reversed: those who use their networks more intensely are more willing to trust outsiders relative to their own kin (or trust both equally little). As individuals make more use of their networks, they start relying on connections that are further removed from them. Relying on these weaker ties could have two effects: it could actually increase trust in outsiders, if they are willing to help contrary to expectation, or it could simply decrease trust in kin, as kin were an insufficient safety net that drove the individual to resort to weaker ties in the first place. Since the question is phrased in relative terms (kin versus non-kin), we cannot fully distinguish between the reduction being caused by an absolute decrease in trusting kin, or just a relative decrease in comparison to non-kin. Further, network density is positively related to trust in neighbors. As we will see in the next section, this result is driven by lower castes, who live in more homogeneous neighborhoods where, even if they rely on weaker ties, those ties are still very similar to them.

¹⁹Corresponding Durbin-Wu-Hausman tests confirm this, though the p-values are higher than ideal (around $p = 0.09$).

Table 3.3 – IV estimates of the determinants of trust – Network size

	First stage	Second stages		
	Network size (1)	Neighborhood (2)	Kin employees (3)	Generalized Trust (4)
Network size (std)		-0.407* (0.240)	0.444** (0.217)	0.033 (0.240)
Age	0.014*** (0.004)	0.008 (0.007)	-0.003 (0.007)	-0.001 (0.006)
Female	-0.550*** (0.099)	-0.145 (0.171)	0.386*** (0.126)	0.017 (0.137)
Middle caste	-0.367** (0.136)	0.655** (0.268)	-0.628*** (0.214)	0.998* (0.526)
Upper caste	-0.306 (0.207)	0.296 (0.254)	-0.449* (0.254)	1.145** (0.499)
Married	0.088 (0.119)	-0.041 (0.163)	-0.038 (0.134)	-0.118 (0.140)
Primary completed	0.000 (0.152)	0.079 (0.088)	-0.012 (0.125)	-0.013 (0.081)
High school (8th-10th)	0.102 (0.131)	0.552*** (0.135)	0.413*** (0.138)	-0.009 (0.152)
HSC/Diploma (11th-12th)	0.211 (0.187)	0.674*** (0.241)	0.373* (0.200)	0.251 (0.192)
Bachelors (13th-15th)	0.063 (0.257)	0.281 (0.304)	-0.044 (0.231)	-0.133 (0.207)
Post Grad	0.813*** (0.215)	1.060*** (0.379)	0.025 (0.305)	0.317 (0.300)
Raven (std)	-0.057 (0.049)	-0.056 (0.066)	0.072* (0.039)	-0.016 (0.075)
Literacy (std)	-0.121* (0.058)	-0.306*** (0.076)	-0.221*** (0.080)	-0.069 (0.090)
Numeracy (std)	0.108 (0.060)	0.080 (0.092)	0.039 (0.068)	0.075 (0.091)
OP (std)	0.094 (0.084)	0.178** (0.074)	0.002 (0.081)	0.099 (0.089)
CO (std)	-0.061 (0.055)	0.222** (0.097)	0.283*** (0.082)	0.029 (0.100)
EX (std)	0.114 (0.092)	0.306*** (0.071)	0.114 (0.074)	0.097 (0.099)
AG (std)	0.010 (0.047)	0.183** (0.077)	-0.012 (0.069)	0.698*** (0.066)
ES (std)	-0.116** (0.049)	-0.300*** (0.087)	-0.004 (0.058)	-0.357*** (0.110)
Grit (std)	0.157 (0.097)	-0.180 (0.156)	-0.095 (0.100)	0.000 (0.084)
Food expenses	0.209* (0.109)	0.029 (0.115)	-0.323*** (0.123)	-0.060 (0.149)
Health expenses	0.069** (0.024)	-0.050 (0.045)	0.008 (0.043)	-0.060 (0.049)
Ceremony expenses	-0.041 (0.057)	0.107* (0.062)	0.041 (0.061)	0.010 (0.082)
HH income	0.112*** (0.028)	-0.026 (0.072)	-0.067 (0.091)	-0.028 (0.087)
Demonetization	-0.643*** (0.172)			
F-stat	14.04			
N	663	663	663	663
R ²	0.412	0.254	0.264	0.571

Source: NEEMSIS (2016-2017); authors' computations. Notes: Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Village-area fixed effects included. Base categories: caste = Dalit, education = no completed primary, sex = male. Household expense variables are in natural logarithm.

Table 3.4 – IV estimates of the determinants of trust – Network density

	First stage	Second stages		
	Network density (1)	Neighborhood (2)	Kin employees (3)	Generalized Trust (4)
Network density (std)		0.734* (0.407)	-0.800* (0.466)	-0.059 (0.437)
Age	0.026*** (0.005)	-0.016 (0.011)	0.024*** (0.008)	0.001 (0.012)
Female	-0.637*** (0.169)	0.547** (0.273)	-0.368 (0.379)	-0.039 (0.296)
Middle caste	0.150 (0.446)	0.694*** (0.241)	-0.670** (0.295)	0.995** (0.496)
Upper caste	0.476 (0.465)	0.071 (0.359)	-0.204 (0.386)	1.163** (0.535)
Married	-0.203 (0.151)	0.072 (0.150)	-0.161 (0.130)	-0.127 (0.146)
Primary completed	0.098 (0.142)	0.007 (0.163)	0.067 (0.213)	-0.007 (0.087)
High school (8th-10th)	0.144 (0.181)	0.405** (0.175)	0.573** (0.240)	0.002 (0.170)
HSC/Diploma (11th-12th)	-0.235 (0.280)	0.760*** (0.191)	0.279 (0.421)	0.244 (0.230)
Bachelors (13th-15th)	-0.427** (0.154)	0.569 (0.450)	-0.358 (0.366)	-0.156 (0.268)
Post Grad	-0.724** (0.233)	1.260*** (0.413)	-0.193 (0.538)	0.301 (0.349)
Raven (std)	0.025 (0.025)	-0.051 (0.073)	0.067 (0.041)	-0.016 (0.076)
Literacy (std)	0.129*** (0.040)	-0.351*** (0.077)	-0.172 (0.123)	-0.065 (0.107)
Numeracy (std)	-0.175*** (0.050)	0.164 (0.109)	-0.053 (0.117)	0.068 (0.106)
OP (std)	-0.012 (0.047)	0.148* (0.076)	0.035 (0.100)	0.102 (0.092)
CO (std)	0.012 (0.082)	0.238*** (0.092)	0.265** (0.121)	0.027 (0.102)
EX (std)	0.001 (0.033)	0.259*** (0.095)	0.166* (0.100)	0.101 (0.101)
AG (std)	-0.057 (0.047)	0.221*** (0.055)	-0.053 (0.087)	0.695*** (0.069)
ES (std)	-0.117* (0.058)	-0.167 (0.117)	-0.150 (0.104)	-0.368*** (0.129)
Grit (std)	-0.018 (0.057)	-0.231* (0.121)	-0.040 (0.118)	0.004 (0.083)
Food expenses	0.042 (0.111)	-0.086 (0.130)	-0.197 (0.163)	-0.051 (0.136)
Health expenses	-0.003 (0.032)	-0.076* (0.045)	0.036 (0.037)	-0.058 (0.046)
Ceremony expenses	0.078 (0.056)	0.066 (0.078)	0.086 (0.088)	0.013 (0.094)
HH income	-0.194* (0.087)	0.071 (0.095)	-0.173* (0.093)	-0.035 (0.095)
Demonetization	0.357*** (0.056)			
F stat	41.2			
N	663	663	663	663
R ²	0.361			0.570

Source: NEEMSIS (2016-2017); authors' computations. Notes: Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Village-area fixed effects included. Base categories: caste = Dalit, education = no completed primary, sex = male. Household expense variables are in natural logarithm.

3.6.4 Heterogeneity analysis

The previous analysis was conducted for the entire matched sample. Still, important differences might exist between subgroups of the sample that are hidden in a general analysis. This is particularly important in the rural Indian context, in which strict social hierarchies along the lines of caste membership and, to a certain extent, gender have been traditionally prevalent. These segregating lines might have become less dominant, but they remain visible. For example, the vast majority of marriages still take place within the same caste, and caste membership can enhance or hinder economic and social mobility (Munshi, 2016b). Men remain the traditional household heads and tend to be the decision-makers in the household. Accordingly, the following section splits our sample into the different caste categories (Dalits, middle castes, upper castes) and along gender lines. We will further look at differential effects by employment status, as the degree to which one interacts with others in one's working life could affect network composition and usage.

Heterogeneous effects by caste membership

Table 3.5 presents our IV estimates by caste membership, split into Dalits, middle castes, and upper castes. Our identification strategy for network size seems to hold only for middle castes, as the other F-statistics are very low (2.4 for Dalits and 5.8 for upper castes), though demonetization has the same effect for all caste categories in terms of sign: having lived through demonetization decreases network size for all, presumably as people update their beliefs of who they can rely on. Among middle castes, network size decreases trust placed in neighbors and increases trust in kin among employees. For network density, the F-statistic holds for lower castes and possibly middle castes (F statistic = 6.4). Among lower castes, making use of one's network more intensely increases both trust placed in neighbors and trust in general. Among upper castes, a similar increase in density decreases trust in kin relative to non-kin employees. Both effects were visible in the previous combined sample estimations.

The differential effects that we find by caste membership relate back to the strict social hierarchies that prevail in rural South India and the importance of taking these into account for any meaningful analysis (Vijayabaskar and Kalaiyarasan, 2014). A caste can provide important economic support to its members and enable effective consumption smoothing (Munshi, 2016b). To enable consumption smoothing within a group, the group must have good information about its members and must be able to punish those that refuse to adhere to their obligations. Part of this mechanism was illustrated in the introductory quote, in which Gomathi, a 26-year-old women who is part of the Dalit community, explains that her female neighbors are there to help her in times of need, but they also expect help from her whenever need arises. While consumption smoothing through borrowing and

Table 3.5 – Estimates of determinants of trust by caste membership for both measures of social networks

(a) Lower castes

	First stage	Second stages			First stage	Second stages		
	Density (1)	Neigh (2)	Kin (3)	Gen Trust (4)	Size (5)	Neigh (6)	Kin (7)	Gen Trust (8)
Density (std)		1.342** (0.650)	-0.355 (0.757)	0.945*** (0.242)				
Size (std)						-1.158 (0.956)	0.306 (0.563)	-0.815 (0.525)
Demo	0.295** (0.093)				-0.342 (0.222)			
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	10.14				2.37			
N	299	299	299	299	299	299	299	299
R ²	0.385		0.284	0.278	0.272		0.383	0.471

Notes: Table based on NEEMSIS (2016-2017). Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

(b) Middle castes

	First stage	Second stages			First stage	Second stages		
	Density (1)	Neigh (2)	Kin (3)	Gen Trust (4)	Size (5)	Neigh (6)	Kin (7)	Gen Trust (8)
Density (std)		0.974 (0.596)	-1.786*** (0.679)	-0.637 (0.569)				
Size (std)						-0.429*** (0.137)	0.787*** (0.189)	0.281 (0.260)
Demo	0.300** (0.119)				-0.681*** (0.123)			
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	6.35				30.68			
N	288	288	288	288	288	288	288	288
R ²	0.395			0.406	0.522	0.429	0.151	0.567

Notes: Table based on NEEMSIS (2016-2017). Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

(c) Upper castes

	First stage	Second stages			First stage	Second stages		
	Density (1)	Neigh (2)	Kin (3)	Gen Trust (4)	Size (5)	Neigh (6)	Kin (7)	Gen Trust (8)
Density (std)		0.742 (0.797)	-0.681 (0.697)	-1.824** (0.852)				
Size (std)						-0.334 (0.505)	0.306* (0.160)	0.820*** (0.273)
Demo	0.500 (0.254)				-1.111* (0.461)			
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	3.88				5.80			
N	76	76	76	76	76	76	76	76
R ²	0.747	0.620	0.531		0.793	0.568	0.639	0.614

Notes: Table based on NEEMSIS (2016-2017). Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

lending thus largely occurs within castes, caste is also a significant determinant of the type of borrowing that is available to individuals. Intuitively, one can only borrow and lend from one's network if the network has the necessary resources. Looking at the study region at hand, Guérin et al. (2013) show that the financial landscape is highly fragmented along caste lines. Lower castes are less likely to borrow from social networks and more likely to borrow from ambulant lenders, though they are also more credit-constrained in general. This is also visible in our data, as descriptive statistics show that lower castes are the most likely to have asked for help but been refused it.

As a result, loans taken out by Dalits after demonetization are more likely to stem from employers and maistries (labor contractors) than prior to demonetization.²⁰ This does not hold true for middle castes, though: while their share of loans from employers also increases (from about 0 to about 8 percent), it is accompanied by an even larger increase in the share of loans coming from relatives and 'well-known people' (from 63 to 73 percent of all loans).²¹

As a result of demonetization, different castes thus answer the shock with different borrowing patterns: lower castes respond to the shock by shifting their borrowing from borrowing among their own caste prior to demonetization to borrowing also from upper castes after demonetization (generally their employers); middle castes shift from borrowing from their own caste and upper castes prior to demonetization to borrowing almost exclusively from within their own caste (90 percent of loans) after demonetization. This is in line with previous research, showing that transfers from other caste members are the preferred method of consumption smoothing in response to income fluctuations (Munshi and Rosenzweig, 2009). Almost all members of the lower caste in our data set are part of the Paraiyar community. Among middle castes, the majority are members of the Vanniyar caste (82 percent). The pattern described among the middle castes holds for the Vanniyar subgroup only, and in fact becomes even stronger.

Further, while middle castes were less likely than lower castes to say that they asked for help and were refused, they are more likely to say that there was no one around to ask for help. Unlike those in lower castes, who are more likely to be employees, members of middle castes could not ask their employers for help to cope. In light of Table 3.5, this information suggests that middle castes updated their information about their social network, leading to fewer potential ties after demonetization. Those who then still had (updated) larger social networks were able to borrow from their own caste and are more trusting in their kin in comparison

²⁰In fact, prior to demonetization, lower castes received the largest share of loans from 'well-known people' and relatives (67 percent), and about 3.6 percent of loans from employers and labor contractors. This changes to 57 percent and 23.75 percent after demonetization, respectively.

²¹Well-known people, "terinjavanga" in Tamil, is a common Tamil term referring to people that have been known to the individual or family for a long time (years or even generations). Most of the time, these people are known through networks, such as friends of relatives or removed relatives, such as a relative of a daughter married to somebody in another village.

to non-kin as a result. Borrowing also took place on similar social levels (friends and ‘well-known people’) in comparison to lower castes, who borrowed up, by borrowing from employers. Frequent interactions between different ethnic groups has been shown to decrease levels of mistrust between them (Stolle et al., 2008). In our case, however, interactions between Dalits and upper castes are probably not frequent enough for this type of mechanism to take place, so any potentially positive effect could not be big enough to observably overcome existing distrust due to pre-existing social hierarchies. Lower caste members who were able to borrow from neighbors (which still happened, though at a lower rate) still borrowed within-caste, as they tend to live in homogeneous neighborhoods. For lower castes, making use of weaker links (higher network density) still means that those links remain within the caste network and can be trusted. Middle castes, by comparison, live in more heterogeneous neighborhoods.

These findings relate back to other studies on social activities in heterogeneous communities such as Alesina and La Ferrara (2000), who find that the degree of heterogeneity in communities influences the amount of participation in groups. It also relates to other surveys, such as the World Values Survey. In addition to a question about neighbors in general, the survey includes questions about trust in neighbors speaking a different language or following a different religion. Trust levels in India significantly decline, from almost 90 percent saying they trust their neighbors to only about 55-60 percent, when asked about dissimilar neighbors (as cited in Munshi, 2016a). In the case at hand, trust in others is thus shaped strongly with relation to closeness to the self, with similar people considered more trustworthy — illustrative of the homophily principle in social networks (McPherson et al., 2001).

Heterogeneous effects by gender

Gender is an important factor in rural South India, with traditional gender roles dominating. It is therefore crucial to consider potentially differential effects by gender. Table 3.6 presents our IV results of the determinants of trust by gender.

It becomes clear that our previous results, which showed network size and density affecting levels of trust, was driven by men only. In fact, looking at men and women separately shows that the first stage holds strongly for men for both measures. It also holds for women with regard to network size, but no coefficient is significant in the second stage. With regard to network size, among men, the coefficient of trust in kin relative to non-kin shows a slightly increased effect, from -0.44 in the combined (female and male) sample in Table 3.3 to -0.55 for the male only sample, while the coefficient for trust in neighbors remains almost equal. For network density, both coefficients nearly halve, with the coefficient for trust in kin among employees gaining significance.

It is likely that our gendered results are driven by the gender roles in South

Table 3.6 – Estimates of determinants of trust by gender for both measures of social networks

(a) Men

	First stage	Second stages			First stage	Second stages		
	Density (1)	Neigh (2)	Kin (3)	Gen Trust (4)	Size (5)	Neigh (6)	Kin (7)	Gen Trust (8)
Density (std)		0.466* (0.279)	-0.549** (0.247)	-0.090 (0.228)				
Size (std)						-0.422* (0.220)	0.498** (0.239)	0.082 (0.207)
Demo	0.611*** (0.077)				-0.675*** (0.191)			
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	52.89				12.95			
N	380	380	380	380	380	380	380	380
R ²	0.458	0.262	0.227	0.578	0.454	0.297	0.299	0.579

Notes: Table based on NEEMSIS (2016-2017). Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

(b) Women

	First stage	Second stages			First stage	Second stages		
	Density (1)	Neigh (2)	Kin (3)	Gen Trust (4)	Size (5)	Neigh (6)	Kin (7)	Gen Trust (8)
Density (std)		-2.098 (2.549)	0.970 (0.840)	-0.606 (1.232)				
Size (std)						-0.601 (0.399)	0.278 (0.270)	-0.174 (0.288)
Demo	-0.190 (0.160)				-0.639*** (0.156)			
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	1.37				16.89			
N	283	283	283	283	283	283	283	283
R ²	0.446			0.518	0.419			

Notes: Table based on NEEMSIS (2016-2017). Standard errors clustered at the village level in parentheses;*** p<0.01,** p<0.05,* p<0.1.

India. In our sample, men are more likely to say that they have asked somebody for help as a result of demonetization, which would lead to information updating, inducing the sort of intensification of network usage that we have in mind (13 percent of men said they asked for help vs. 6 percent of women). In fact, women are more likely to say that they did not need to ask anybody for help (74 percent of women answered this way compared to 67 percent of men). This could suggest several things: that women do not feel like they need to ask for help as this is assumed to be part of a man's role, that informal lending is socially inaccessible for women, or that women were able to overcome this period of cash shortage with the help of, for example, hidden cash reserves. The literature supports all three hypotheses. Indeed, women and members of lower castes in north Tamil

Nadu have more difficulties accessing informal lending, paying on average more and borrowing primarily for consumption (Harriss-White and Colatei, 2004). Further, there is abundant evidence that women do not share their entire income with their husband, often putting some of it away. The vast majority of women does not have access to the banking system, and only via joint accounts with their husbands for those that do. Saving some money in private cash hoards thus provides the only way for women to guard it from the males in their households (who might prefer to spend it on demerit goods) or to save money for their children. Indeed, in the survey area, about 70 percent of women claim to secretly save some cash (Guérin, 2008). Women were then doubly hit by the demonetization shock: standing in long queues to exchange the money could be considered inappropriate while, at the same time, the pure revelation of a secret cash hoard to husbands could have negative repercussions, potentially leading some women to lose control of their reserves (Ghosh et al., 2017). A deeper dive into these different channels of a gendered analysis of the demonetization shock could be a rather fruitful avenue for future research.

Heterogeneous effects by employment status

Lastly, we will consider differential effects by employment status. Demonetization did not have an effect on labor force participation in the study region: prior to the shock, 93 percent of people reported having worked in the past 7 days, which did not change as a result of the policy. We therefore look at employment status instead, as those with a less secure employment and fewer interactions in their workplace might be more vulnerable to shocks and therefore rely more heavily on their networks to cope. For example, a salaried worker could ask his employer for an advance, while a self-employed worker does not have this option. At the same time, those most in need of coping could also be disadvantaged by their networks; if networks are formed among similar people, it is well possible that those with a less secure employment status might also only have people in their network with similarly insecure employment, in which case we would not expect to see any significant results. Further, there is ample evidence that occupational segregation exists along the lines of caste-based networks, reinforced by labor contractors engaging in caste-based hiring (Munshi, 2016b, and references therein).

Workers in the study region are employed in four main categories: non-salaried agricultural work on their own farm or self-employment on one hand, and salaried non-agricultural or agricultural work on someone else's farm on the other.²² Given that the sample is located in the rural area, the last category (salaried non-agricultural workers) also includes workers employed primarily under the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA). We introduce a dummy variable in all employment status regressions, indicating whether the

²²Few workers are unpaid, which is why this category will be disregarded for our purpose.

worker declares her main occupation to be MGNREGA.²³

As has become clear from the previous analysis, caste membership remains a decisive factor in rural South India. This also holds true with regards to occupational choice, as illustrated by Table 3.7. Lower castes tend to be employed as either salaried agricultural workers (agricultural coolies) or salaried non-agricultural workers, middle castes are more likely to be land-owners and therefore work on their own land, and upper castes predominantly work in self-employed jobs or as salaried non-agricultural workers. The large share of middle caste landowners is the result of the restructuring of land that occurred in South India over the last three decades as upper castes (the traditional landowners) moved to urban areas (Guérin et al., 2015).

Table 3.7 – Distribution of employment status by caste membership

	Dalit	Middle caste	Upper caste	Total shares
Agricultural activity	5.27	33.57	6.43	17.9
Self-employed	11.11	20.59	48.93	19.37
Salaried job (agri)	34.58	3.47	1.17	17.23
Salaried job (non-agri)	49.04	42.37	43.47	45.49

Notes: Table based on NEEMIS (2016-2017). Unpaid workers are disregarded.

Table A3.5 displays our results for non-salaried occupations (agricultural work on one’s own farm and self-employment) and Table A3.6 shows results for salaried occupations (agricultural and non-agricultural). Our identification strategy holds for agricultural workers who work on their own farm (these are mostly members of middle castes, for whom we have consistently found an effect throughout this paper) and for salaried agricultural occupations (90 percent of whom are Dalits). For own-farm agricultural workers, we once again find that a higher network density leads to less trust in one’s kin and now also less trust in people in general, while a larger network size leads to more trust in kin and more trust in people in general. For salaried agricultural workers, we now see that a larger network size relates to less trust in neighbors (the opposite result of what we found in Table 3.5 when looking at the entire Dalit community, no matter their employment status). This result is still in line with our story: Dalits who are working in salaried agricultural occupations are the ones who could ostensibly diversify their borrowing and lending and borrow upward (from employers). They are thus not broadening their network size towards others within the neighborhood but towards superiors. This then leads those with larger networks to trust their neighbors less, as they were unable to rely on them and had to engage in more hierarchically unequal types of borrowing.

Self-employed workers, working mostly in small business, were probably unable

²³MGNREGA participation in the study region is actually lower after demonetization than before. MGNREGA offers guaranteed employment in rural areas conducting unskilled manual labor for at least 100 days per financial year. In the study region, the MGNREGA work was stopped from January until March 2017 due to the cash shortages. In other regions, MGNREGA participation has been shown to have increased as a result of the demonetization policy (Ghosh et al., 2017).

to intensify their network relationship and realized that the only help they could get as a result of the shock came from professional money lenders (as in the qualitative example of Saleem Basha mentioned earlier, who had to take out a loan from a money lender to keep his small grocery shop stocked). Indeed, research from India as a whole shows that the self-employed have intensified their credit relationships with customers (Ghosh et al., 2017). One could thus think that we should also see an effect for the self-employed, as selling to others on credit is by itself an act that requires the seller to trust that the buyer will eventually repay. One possible explanation for this lack of effect is that selling on credit is a necessary requirement to retain customers – as in the qualitative example mentioned, not a voluntary choice. This also does not intensify or increase the quantity of social interactions, as a seller’s aim in extending credit is to retain an existing customer base, not to build up a new one. Alternatively, the trust formation in our context might only happen if the person was able to receive help when in need, and not if the person was able to give help. This could then explain the insignificant results for upper castes throughout the paper, as upper castes were likely to have either been unaffected by the shock due to sufficient resources or to have increased their lending to help others.

3.6.5 Robustness checks

Days passed since demonetization

The survey was conducted over two time periods, one prior to and one after demonetization. The post-demonetization survey collection started in January 2017 (about 2 months after the shock) and ended at the end of March 2017, meaning that the last interview was conducted about 5 months after the shock.²⁴ According to the mechanism at hand – that trust is built through social interactions – time passed since demonetization is likely to have a positive effect on our measures of trust, as people have more time to actually interact. In this section, we explore the time dimension of the survey, looking at the effect of time passed since demonetization. As expected, network density and days passed since demonetization correlate positively, though not particularly strongly ($r = 0.18, p < 0.000$). Similarly, network size and days passed since demonetization correlate negatively ($r = -0.30, p < 0.000$). Figure A3.2 illustrates those correlations. Table A3.7 provides the IV specification, using days passed as the instrument instead of the demonetization dummy used in Table 3.4. Results mirror the results obtained with the demonetization dummy. The robustness checks suggest a minor role for the amount of time passed between the demonetization shock and the date of interview.

²⁴The survey started in August 2016 and was interrupted a week before the demonetization shock took place. This interruption was related not to the demonetization shock but to logistical survey constraints (enumerator payment and technical issues with the digital tablets we used to collect the data).

Lowering the trustworthiness threshold

A potential second channel that could affect our results is that instead of changing beliefs about others' trustworthiness, the shock might have lowered the trustworthiness threshold at which individuals are willing to interact. Instead of revealing information about the ties (if one can rely on them or not), the individual might in this scenario simply be willing to interact with anybody who could help, even if those people are not necessarily trustworthy. Rather than information updating, the mechanism would then be driven by a lower trustworthiness threshold that has to be overcome to facilitate interaction. If this channel was at play, we would expect individuals to increase their number of ties after the shock (having a larger network size), as more ties would pass the lowered trustworthiness threshold. We would also expect individuals to increase their network density (using more of their ties), as, again, a larger share of the network surpasses the threshold. Instead of a larger network size, however, we observe a smaller total network size, driven by a lower number of elicited potential ties. Network density is indeed larger after the shock, but this is also driven by a reduction in the number of potential ties (part of the denominator). This leads us to conclude that demonetization did in fact affect people's beliefs about others, as they elicit fewer potential ties after the shock than before.

Estimations without the agreeableness dimension

One potential concern about including the Big Five in our second stage IV regressions is simultaneity bias, which would be problematic if Big Five dimensions determine trust and trust determines Big Five scores. The dimension for which this is most likely is agreeableness, the tendency to act cooperatively. It relates in fact to the facets altruism and trust, as cooperative interactions require trust between interacting parties. We estimate our IV regressions without the agreeableness dimension. Table A3.8 displays the second stages for both network density and network size.²⁵ Table A3.8 shows that our main results hold, even when estimated without the agreeableness dimension. In fact, our coefficients for the effect of network size on trust in neighbors and trust in kin relative to non-kin are barely affected, with the coefficient for trust in neighbors increasing slightly (from 0.73 to 0.87). The changes in coefficients for the effect of network density on trust in neighbors and kin are even smaller. The largest difference is seen for the generalized trust measure, though results remain insignificant, as they have been throughout the paper. This leads us to conclude that the estimations are still valid with the agreeableness dimension.

²⁵First stages are not reported for simplicity but hold as well. The corresponding F-statistics are 38.15 for network density and 14.15 for network size.

Migrating households

A fourth channel that could affect our results is migration. There has been evidence throughout the country that migrants were forced to return home after the shock, as employers were often unable to continue paying wages (Ghosh et al., 2017). More than half of the households in our sample are migrant households (meaning that at least one member migrates temporarily for work). While survey collection was essentially random and the break in the survey was not related to demonetization, the survey team did decide to interview migrating households later during the survey timeframe to have a better chance of interviewing them.²⁶ In order to check that our results are not driven by migrating households, we run the IV estimations while also including a dummy variable that takes the value of 1 if somebody in the household is a migrant (even if it's not the person interviewed) and an additional dummy taking the value of 1 if the individual himself is a migrant. Table A3.9 shows that the migrant household dummy is only significant in column (4) (trust in kin among employees, with network size being the endogenous variable). Results of the effect of network size and network density on trust in employees remain; while the coefficients' signs for the results for trust in neighbors stay the same, however, they are no longer statistically significant (though they were only significant at the 10 percent level in Tables 3.4 and 3.3 anyway).

Poverty

One additional potential confounder of our results could be poverty, as poverty could make people less trusting in general. Lack of trust has been linked to low socio-economic status and lack of material wealth (e.g. Delhey and Newton, 2003; Knack and Keefer, 1997; Uslaner, 2002), with one possible channel that the risk of trusting may be too great for the most deprived, as they have a greater share of their total wealth to lose if their trust is betrayed (Putnam, 2001).

In the previous estimations, we already control for households expenses and household income, in addition to basing the estimations on a sample matched based on these households' characteristics and characteristics of the dwelling, which should capture part of this aspect. Still, to make sure that our results hold for poorer and richer households, we build an asset index based on items that are likely to change as a result of the shock (household income and expenses) and items that better capture an enduring poverty status (goods owned by the household and participation in government schemes targeted to the poor).²⁷ We reduce these items to a composite index using principal component analysis and keep

²⁶Migrants tend to travel to their home villages for festivals.

²⁷The exact list of variables included in the poverty index are: household expenses on food, ceremonies, and health; household income; whether the household owns a fridge, expensive furniture, a car, a cell phone, a landline, or a computer; whether the household benefits from the ration card, free housing, free cow and goats, or free gas government schemes; and characteristics of the dwelling (water access, electricity, toilet facilities, and type of house).

the first component. Table A3.10 displays the results for the poorest and richest households. Results hold for the richest households (columns (4) through (6)), though we cannot find any significant results for network density among the poorest households, finding only that network size decreases trust in neighbors and trust in people in general among this group. Our results for trust in kin among employees are thus driven by individuals from comparatively wealthy households. One has to keep in mind that poverty status varies with caste membership, as lower castes score lower on the poverty index than upper castes. The tertile declared as ‘rich’ in this section is then made up of upper castes and middle castes, though even some Dalits are in this group (22 percent Dalits, 58 percent middle castes, and 20 percent upper castes); the tertile declared as ‘poor’, on the other hand, is 53 percent Dalits and 44 percent middle castes. Indeed, results for the richest tertile look similar to those obtained for middle castes when splitting by caste membership in Table 3.5. Results for the poorest tertile are different from those obtained for lower castes, though, as network size now significantly decreases trust in neighbors and people in general. In the estimations for lower castes, those coefficients carried the same sign but were statistically insignificant. As the poorest tertile contains a significant number of middle castes (almost half) for whom the negative effect was also visible in Table 3.5, it is possible that they drive these results. Given the sorting of castes along the income distribution, we conclude that caste membership is a stronger driver of trust formation than poverty.

Happiness

The last channel for which we consider robustness checks is happiness. People could be unhappy about the demonetization shock and its consequences and therefore less likely to trust others. Table 3.2 in the exclusion restriction section provides evidence that people did not think that demonetization itself changed their levels of trust in neighbors and kin (about 74 percent of the demonetization sample).²⁸ Other research provides evidence that demonetization was mostly perceived positively, even among those who suffered (Ghosh et al., 2017). Unfortunately, the survey does not contain questions about life satisfaction, the most common variable with which happiness can be captured.

Given the lack of a life satisfaction variable, we will rely on the Big Five dimension emotional stability, which has been most strongly related to concepts of happiness (Hills and Argyle, 2001). The descriptive statistics for the balanced sample (displayed in Table A3.3) show that emotional stability does indeed remain different between the pre- and post demonetization samples, even after matching, with the post demonetization sample slightly more emotionally stable (*p-value of 0.05*). We estimate separate IV regressions for those who score in the lowest tertile

²⁸Demonetization led people to put less trust in banks, with 22 percent saying that demonetization made them trust banks less.

of the emotional stability dimension and those who score in the highest.²⁹ Table A3.11 shows that the effect for trust in kin relative to non-kin members is the same, no matter the position within the emotional stability dimension. Thus, to the extent that it can be approximated by emotional stability, our results do not appear to be driven by happiness.

3.7 Conclusion

Trust in other people, an essential component of social capital, is particularly crucial in developing countries, where a large share of transactions are informal and take place within social networks. But trust is to a large extent endogenous, as it is “an outcome and an antecedent of relationships” (Nooteboom, 2007, p. 33). This paper aims to disentangle this relationship between social networks and trust by exploring an exogenous variation that directly affected people’s information about their social network but did not have a direct effect on interpersonal trust. The exogenous shock explored, the 2016 demonetization policy in India, reduced money supply overnight, inducing individuals to rely on their social networks for everyday transactions.

We use novel quantitative and qualitative data from rural Tamil Nadu, collected by the authors, to provide causal estimates of the effects of two measures of social networks (size and density) and three measures of trust (trust within a neighborhood, trust in kin versus non-kin among employees, and generalized trust). We use an IV approach with the shock introduced by the demonetization policy in November 2016 as an instrument that had a significant effect on network measures but did not directly affect trust placed in other people. This presupposed channel is also visible in first-hand qualitative data collected by the authors to understand how demonetization impacted people’s lives in rural South India, thereby offering convincing evidence for the exclusion restriction. We control for a large variety of individual characteristics that could affect trust formation, such as cognitive ability and personality traits, which in other cases have been considered unobservable or required panel data to be purged from the estimates. We use network data collected as part of the survey to construct two measures of social networks: network size, the sum of all potential and actual ties, and network density, the share of ties activated. Both measures are mostly reliant on economic interactions (loans and access to labor).

We find that network density causally increases levels of trust placed in neighbors and decreases trust placed in kin among employees, while network size decreases trust in neighbors and increases trust placed in kin employees. Heterogeneity analyses illustrate that these results for the entire sample hide important differences. Most notably, our results only hold for men, as strong gender roles both reduce

²⁹We also tried quartiles and results are similar.

women’s ability to interact in the way that we are capturing interactions and might mean that women have different strategies of coping with shocks (such as cash hoarding), not reflected by our data. Further, we find different results by caste membership. We do not find any significant results for upper castes, though due to the stratified way in which the survey was collected, our sample for upper castes is also quite small. Comparing lower castes (Dalits) and middle castes still reveals important differences regarding the levels and types of interactions that occurred as a result of the shock. Lower castes coped by taking out loans from those around them (in homogeneous neighborhoods) and from their employers. Among Dalits, who are oftentimes employed as salaried agricultural laborers in the study region, we find that making use of one’s network more intensely (increased network density) leads to higher trust in neighbors. For middle castes, though, the story is different. Middle castes live in more heterogeneous environments and often work on their own agricultural land, as the exodus of upper castes to urban areas has enabled a reallocation of land to the middle castes. As a result of the shock, they coped by borrowing from other caste members, or ‘well-known people’. Among middle castes, a larger number of ties (network size) leads to more trust in kin members in comparison to non-kin members and lower trust in neighbors, who could be more dissimilar to the self. For this group, higher network density, making use of one’s network more intensely, leads to lower trust placed in kin-employees. As middle castes have to expand their networks to cope, they then rely on weaker ties of which they are arguably more dubious, driving the reduction in trust levels.

This paper illustrates that a common shock can have differential effects on levels of trust in a society, given the type of interactions that take place as a result of the shock. Notably, it demonstrates homophily in networks in rural South India, where interactions that happen within a homogeneous group (neighborhoods for lower castes, kin and other caste members for upper castes) foster trust, while outside interactions or relying on marginal ties decrease it. This is in line with previous research demonstrating that Indians tend to trust those who are similar to themselves but not other linguistic or religious groups (Munshi, 2016a). The paper also showcases the importance of not relying only on broad measures of trust, such as generalized trust, when examining an environment characterized by tightly knit social groups. We do not find any results for our measure of generalized trust, but results turn significant once we consider measures of trust that more clearly define an in-group in comparison to an out-group (neighbors and non-neighbors, kin among employees and non-kin among employees). The paper further presents evidence that caste membership remains a significant determinant of social and economic outcomes in today’s rural India.

Appendix for Chapter 3: Tables and Figures

Table A3.1 – Descriptive statistics of the individual level sample

	N	Mean	SD	Min	Max
Age	885	42.65	13.67	18	81
Married	885	0.82	0.39	0	1
Male	885	0.56	0.50	0	1
Dalit	885	0.48	0.50	0	1
Middle caste	885	0.42	0.49	0	1
Upper caste	885	0.10	0.30	0	1
Below primary	885	0.38	0.49	0	1
Primary completed	885	0.21	0.40	0	1
High school (8th-10th)	885	0.26	0.44	0	1
HSC/Diploma (11th-12th)	885	0.07	0.26	0	1
Bachelors (13th-15th)	885	0.06	0.23	0	1
Post Grad	885	0.02	0.14	0	1
Raven	885	13.21	8.84	0	36
Literacy	885	1.72	1.72	0	4
Numeracy	885	1.79	1.30	0	4
OP	885	2.76	0.67	1.1	4.71
CO	885	3.47	0.67	1.6	5.00
EX	885	3.36	0.59	1.4	4.71
AG	885	3.35	0.39	2.3	5.00
ES	885	3.29	0.48	2.0	5.00
Grit	885	3.06	0.58	1.3	5.00
Trust in neighborhood	885	3.92	1.02	1	5
Trust in kin	885	3.45	0.95	1	5
Generalized Trust	885	3.20	0.81	1	5
Actual SN size	885	1.39	1.75	0	10
Potential SN size	885	6.83	3.61	1	21
Total SN size	885	8.22	4.34	1	26
SN density	885	0.15	0.17	0	0.77
Villages					
ELA	885	0.11	0.31	0	1
GOV	885	0.10	0.29	0	1
KAR	885	0.09	0.29	0	1
KOR	885	0.09	0.28	0	1
KUV	885	0.11	0.31	0	1
MAN	885	0.10	0.30	0	1
MANAM	885	0.09	0.28	0	1
NAT	885	0.11	0.31	0	1
ORA	885	0.10	0.30	0	1
SEM	885	0.11	0.32	0	1

Source: NEEMSIS (2016-2017); authors' computations. The data is available on the NEEMSIS webpage <https://neemsis.hypotheses.org/>.

Notes: Sample restricted to those with non-missing cognitive and non-cognitive skills, and trust variables. The raw individual level data contains 954 individuals.

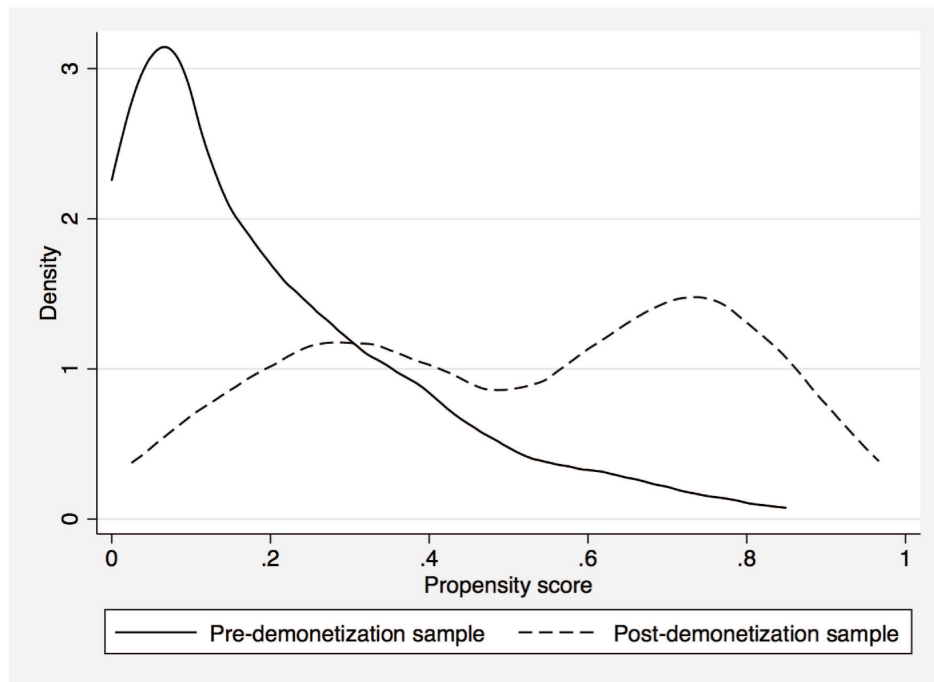
Table A3.2 – Balance checks - unmatched samples

	Before demonetization			After demonetization			Diff	P-value
	N	Mean	SD	N	Mean	SD		
Age	630	43.13	13.92	255	41.48	13.01	1.64	0.12
Married	630	0.82	0.38	255	0.80	0.40	0.02	0.44
Male	630	0.53	0.50	255	0.63	0.49	-0.10	0.01
Dalit	630	0.49	0.50	255	0.46	0.50	0.04	0.37
Middle caste	630	0.40	0.49	255	0.45	0.50	-0.04	0.25
Upper caste	630	0.10	0.30	255	0.09	0.29	0.01	0.69
Below primary	630	0.40	0.49	255	0.33	0.47	0.07	0.06
Primary completed	630	0.21	0.41	255	0.19	0.39	0.02	0.53
High school (8th-10th)	630	0.26	0.44	255	0.29	0.45	-0.03	0.35
HSC/Diploma (11th-12th)	630	0.06	0.24	255	0.10	0.30	-0.04	0.07
Bachelors (13th-15th)	630	0.05	0.22	255	0.07	0.25	-0.02	0.35
Post Grad	630	0.02	0.13	255	0.02	0.15	-0.01	0.55
Raven	630	12.52	8.30	255	14.89	9.86	-2.36	0.00
Literacy	630	1.61	1.71	255	1.99	1.70	-0.37	0.01
Numeracy	630	1.71	1.32	255	1.96	1.23	-0.25	0.00
OP	630	2.90	0.57	255	2.96	0.48	-0.06	0.13
CO	630	3.56	0.66	255	3.77	0.71	-0.21	0.00
EX	630	3.49	0.49	255	3.56	0.45	-0.07	0.06
AG	630	3.47	0.36	255	3.58	0.39	-0.10	0.00
ES	630	3.41	0.61	255	3.52	0.62	-0.10	0.02
Grit	630	3.12	0.58	255	3.47	0.59	-0.36	0.00
Food expenses (HH)	630	7.02	0.50	255	6.73	0.44	0.30	0.00
Health expenses (HH)	630	9.12	1.09	255	9.00	1.00	0.13	0.11
Ceremonies expenses (HH)	629	9.16	0.83	255	9.04	0.58	0.12	0.04
Total HH income	629	11.49	0.95	255	11.87	0.70	-0.38	0.00
Trust in neighborhood	630	3.89	1.06	255	4.00	0.91	-0.11	0.15
Trust in kin	630	3.50	0.98	255	3.32	0.85	0.18	0.01
Generalized Trust	630	3.16	0.84	255	3.30	0.74	-0.14	0.02
Actual ties	630	1.42	1.84	255	1.33	1.50	0.09	0.51
Potential ties	630	7.40	3.81	255	5.42	2.56	1.98	0.00
Total SN size	630	8.82	4.64	255	6.75	3.02	2.07	0.00
SN density	630	0.14	0.16	255	0.18	0.19	-0.04	0.00

Source: NEEMIS (2016-2017); authors' computations. Sample contains only those with non-missing cognitive and non-cognitive variables.

Notes: OP = Openness to Experience, CO = Conscientiousness, EX = Extraversion, AG = Agreeableness, ES = Emotional Stability. Personality traits are acquiescence corrected. Household expenses are in natural logarithms.

Figure A3.1 – Common support for matching



Source: NEEMSI (2016-17); based on authors' calculations.

Table A3.3 – Balance checks - matched samples

	Before demonetization			After demonetization			Diff	P-value
	N	Mean	SD	N	Mean	SD		
Age	425	39.38	13.67	238	41.77	13.02	-2.39	0.12
Married	425	0.75	0.39	238	0.80	0.40	-0.05	0.32
Male	425	0.68	0.50	238	0.61	0.49	0.08	0.14
Dalit	425	0.44	0.50	238	0.45	0.50	-0.01	0.87
Middle caste	425	0.44	0.50	238	0.45	0.50	-0.01	0.89
Upper caste	425	0.12	0.33	238	0.10	0.30	0.02	0.59
Below primary	425	0.31	0.49	238	0.33	0.47	-0.02	0.75
Primary completed	425	0.19	0.39	238	0.20	0.40	-0.01	0.83
High school (8th-10th)	425	0.22	0.45	238	0.29	0.45	-0.07	0.13
HSC/Diploma (11th-12th)	425	0.19	0.26	238	0.10	0.30	0.09	0.08
Bachelors (13th-15th)	425	0.06	0.23	238	0.06	0.24	0.00	0.99
Post Grad	425	0.03	0.14	238	0.03	0.16	0.00	0.99
Raven	425	1.96	1.75	238	2.00	1.70	-0.04	0.84
Literacy	425	1.94	1.34	238	1.96	1.22	-0.01	0.92
Numeracy	425	13.99	8.36	238	14.50	9.90	-0.52	0.64
OP	425	3.03	0.55	238	2.97	0.47	0.06	0.29
CO	425	3.53	0.68	238	3.73	0.71	-0.20	0.01
EX	425	3.45	0.49	238	3.56	0.46	-0.11	0.07
AG	425	3.42	0.37	238	3.56	0.39	-0.13	0.01
ES	425	3.34	0.61	238	3.49	0.62	-0.14	0.05
Grit	425	3.34	0.56	238	3.43	0.57	-0.09	0.13
Food expenses (HH)	425	6.72	0.46	238	6.75	0.44	-0.03	0.68
Health expenses (HH)	425	9.03	1.13	238	9.00	1.00	0.03	0.83
Ceremonies expenses (HH)	425	9.10	0.88	238	9.04	0.59	0.06	0.63
Total HH income	425	11.87	0.91	238	11.84	0.71	0.02	0.76
Trust in neighborhood	425	3.66	1.06	238	4.02	0.93	-0.36	0.00
Trust in kin	425	3.46	0.99	238	3.31	0.86	0.15	0.17
Generalized Trust	425	3.13	0.84	238	3.28	0.75	-0.15	0.23
Actual ties	425	1.24	1.76	238	1.32	1.50	-0.08	0.66
Potential ties	425	7.90	3.84	238	5.40	2.62	2.50	0.00
Total SN size	425	9.14	4.68	238	6.72	3.09	2.42	0.00
SN density	425	0.12	0.15	238	0.18	0.19	-0.06	0.00

Source: NEEMSI (2016-2017); authors' computations. Sample contains only those with non-missing cognitive and non-cognitive variables.

Notes: OP = Openness to Experience, CO = Conscientiousness, EX = Extraversion, AG = Agreeableness, ES = Emotional Stability. Personality traits are acquiescence corrected. Household expenses are in natural logarithms.

Table A3.4 – OLS estimates of the determinants of trust

Trust measures	Neighborhood	Neighborhood	Kin employees	Kin employees	Gen. Trust	Gen. Trust
	(1)	(2)	(3)	(4)	(5)	(6)
Network density (std)	0.005 (0.042)		-0.055 (0.068)		-0.039 (0.044)	
Network size (std)		-0.021 (0.074)		0.173*** (0.045)		0.064 (0.062)
Age	0.002 (0.005)	0.003 (0.005)	0.005 (0.004)	0.001 (0.005)	0.000 (0.006)	-0.002 (0.006)
Female	0.087 (0.102)	0.072 (0.123)	0.101 (0.107)	0.234** (0.077)	-0.026 (0.060)	0.034 (0.082)
Middle caste	1.044*** (0.174)	1.026*** (0.151)	-1.027*** (0.094)	-0.888*** (0.110)	0.985* (0.519)	1.028* (0.540)
Upper caste	0.659** (0.224)	0.644*** (0.194)	-0.805*** (0.158)	-0.694*** (0.188)	1.147** (0.499)	1.173* (0.522)
Married	-0.077 (0.193)	-0.076 (0.193)	-0.009 (0.141)	-0.013 (0.144)	-0.122 (0.162)	-0.120 (0.156)
Primary completed	0.066 (0.101)	0.067 (0.097)	0.007 (0.189)	-0.003 (0.166)	-0.009 (0.088)	-0.014 (0.090)
High school (8th-10th)	0.514*** (0.133)	0.517*** (0.133)	0.462** (0.182)	0.438** (0.165)	-0.001 (0.164)	-0.012 (0.166)
HSC/Diploma (11th-12th)	0.548** (0.173)	0.553** (0.183)	0.496* (0.234)	0.458* (0.218)	0.250 (0.189)	0.241 (0.180)
Bachelors (13th-15th)	0.249 (0.351)	0.249 (0.348)	-0.031 (0.236)	-0.021 (0.227)	-0.148 (0.223)	-0.136 (0.226)
Post Grad	0.698** (0.298)	0.713* (0.323)	0.381 (0.253)	0.268 (0.239)	0.316 (0.332)	0.289 (0.312)
Raven (std)	-0.020 (0.075)	-0.022 (0.070)	0.036 (0.048)	0.049 (0.048)	-0.017 (0.093)	-0.013 (0.089)
Literacy (std)	-0.253*** (0.076)	-0.255*** (0.066)	-0.273*** (0.072)	-0.257*** (0.068)	-0.068 (0.094)	-0.065 (0.087)
Numeracy (std)	0.029 (0.111)	0.031 (0.103)	0.085 (0.070)	0.073 (0.059)	0.072 (0.102)	0.071 (0.097)
OP (std)	0.120** (0.048)	0.123** (0.046)	0.063 (0.089)	0.041 (0.087)	0.102 (0.105)	0.095 (0.101)
CO (std)	0.253*** (0.077)	0.251** (0.079)	0.250* (0.116)	0.262** (0.103)	0.027 (0.114)	0.031 (0.116)
EX (std)	0.270** (0.083)	0.272*** (0.080)	0.154 (0.090)	0.138 (0.084)	0.101 (0.109)	0.094 (0.107)
AG (std)	0.195** (0.073)	0.194** (0.075)	-0.027 (0.073)	-0.020 (0.071)	0.696*** (0.072)	0.699*** (0.071)
ES (std)	-0.249** (0.109)	-0.252** (0.100)	-0.066 (0.082)	-0.038 (0.073)	-0.365** (0.124)	-0.353** (0.118)
Grit (std)	-0.244* (0.122)	-0.241* (0.120)	-0.026 (0.116)	-0.052 (0.106)	0.005 (0.092)	-0.005 (0.091)
Food expenses	-0.044 (0.100)	-0.040 (0.104)	-0.241* (0.109)	-0.275** (0.117)	-0.052 (0.149)	-0.066 (0.147)
Health expenses	-0.078 (0.043)	-0.076 (0.046)	0.037 (0.038)	0.026 (0.038)	-0.058 (0.050)	-0.062 (0.051)
Ceremony expenses	0.124** (0.054)	0.123* (0.055)	0.027 (0.087)	0.030 (0.081)	0.012 (0.089)	0.011 (0.090)
HH income	-0.076 (0.093)	-0.074 (0.098)	-0.023 (0.109)	-0.033 (0.115)	-0.031 (0.101)	-0.032 (0.107)
N	663	663	663	663	663	663
R ²	0.344	0.344	0.294	0.311	0.570	0.572

Source: NEEMSIS (2016-2017); authors' computations based on the matched samples.

Notes: Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Village-area fixed effects included. Base categories: caste = Dalit, education = no completed primary, sex = male. Personality traits are acquiescence corrected and standardized. Household expense variables are in natural logarithm.

Table A3.5 – Estimates of determinants of trust by employment status for both measures of social networks – Non-salaried occupations

(a) Agriculture, own farm

	First stage	Second stages			First stage	Second stages		
	Density (1)	Neigh (2)	Kin (3)	Gen Trust (4)	Size (5)	Neigh (6)	Kin (7)	Gen Trust (8)
Density (std)		0.153 (0.094)	-0.562*** (0.153)	-0.564*** (0.165)				
Size (std)						-0.288 (0.225)	1.058* (0.546)	1.062** (0.486)
Demo	1.204*** (0.315)				-0.639* (0.325)			
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	14.65				3.88			
N	135	135	135	135	135	135	135	135
R ²	0.565	0.551	0.512	0.580	0.523	0.489	0.126	0.217

Notes: Table based on NEEMSIS (2016-2017). Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Apart from the usual controls, a dummy for whether or not the person's main occupation is MGNREGA is included.

(b) Self-employed

	First stage	Second stages			First stage	Second stages		
	Density (1)	Neigh (2)	Kin (3)	Gen Trust (4)	Size (5)	Neigh (6)	Kin (7)	Gen Trust (8)
Density (std)		0.416 (0.577)	0.877 (0.599)	-0.424 (0.446)				
Size (std)						-0.189 (0.334)	-0.399* (0.232)	0.193 (0.269)
Demo	0.322 (0.243)				-0.709 (0.550)			
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	1.76				1.66			
N	115	115	115	115	115	115	115	115
R ²	0.462	0.513	0.101	0.734	0.585	0.522	0.580	0.830

Notes: Table based on NEEMSIS (2016-2017). Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Apart from the usual controls, a dummy for whether or not the person's main occupation is MGNREGA is included.

Table A3.6 – Estimates of determinants of trust by employment status for both measures of social networks – Salaried occupations

(a) Salaried agricultural occupations

	First stage	Second stages			First stage	Second stages		
	Density (1)	Neigh (2)	Kin (3)	Gen Trust (4)	Size (5)	Neigh (6)	Kin (7)	Gen Trust (8)
Density (std)		4.843 (3.813)	-1.188 (1.957)	-0.785 (2.145)				
Size (std)						-1.386*** (0.319)	0.340 (0.498)	0.225 (0.464)
Demo	0.203 (0.232)				-0.709*** (0.205)			
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	0.77				11.90			
N	115	115	115	115	115	115	115	115
R ²	0.660		0.122	0.779	0.650	0.434	0.677	0.884

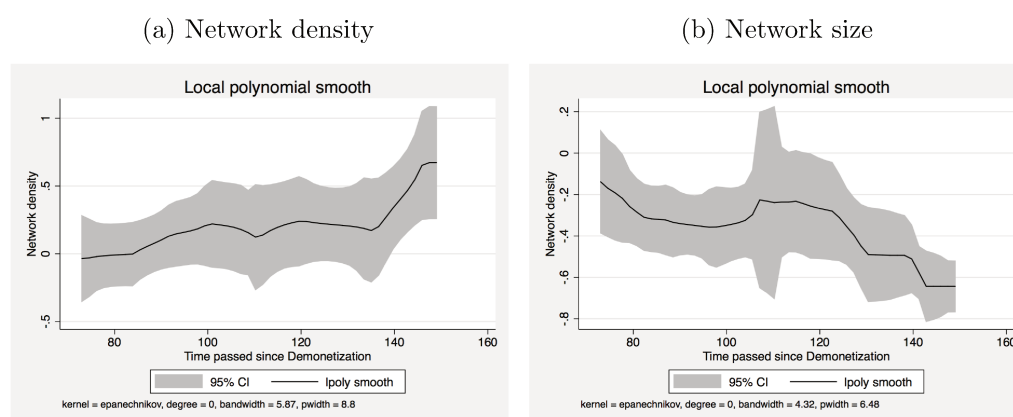
Notes: Table based on NEEMSIS (2016-2017). Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Apart from the usual controls, a dummy for whether or not the person's main occupation is MGNREGA is included.

(b) Salaried non-agricultural occupations

	First stage	Second stages			First stage	Second stages		
	Density (1)	Neigh (2)	Kin (3)	Gen Trust (4)	Size (5)	Neigh (6)	Kin (7)	Gen Trust (8)
Density (std)		1.704 (1.522)	-0.745 (0.587)	-0.753 (0.863)				
Size (std)						-0.795* (0.472)	0.347** (0.170)	0.351 (0.256)
Demo	0.270 (0.232)				-0.578* (0.275)			
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	1.35				4.42			
N	261	261	261	261	261	261	261	261
R ²	0.357		0.018	0.360	0.346	0.276	0.393	0.641

Notes: Table based on NEEMSIS (2016-2017). Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Apart from the usual controls, a dummy for whether or not the person's main occupation is MGNREGA is included.

Figure A3.2 – Relationship between days passed since demonetization and networks



Source: NEEMSI (2016-17); based on authors' calculations.

Table A3.7 – Robustness: IV estimates of the determinants of trust, using days passed since demonetization as instrument

	First stage	Second stages			First stage	Second stages		
	Density (1)	Neigh (2)	Kin (3)	Gen Trust (4)	Size (5)	Neigh (6)	Kin (7)	Gen Trust (8)
Density (std)		0.426 (0.340)	-0.930** (0.386)	-0.139 (0.341)				
Size (std)						-0.263 (0.215)	0.575*** (0.209)	0.086 (0.208)
Days passed	0.004*** (0.001)				-0.006*** (0.001)			
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	57.36				26.90			
N	663	663	663	663	663	663	663	663
R ²	0.369	0.218		0.564	0.419	0.309	0.206	0.571

Notes: Table based on NEEMSI (2016-2017). Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Village-area fixed effects included. Base categories: caste = Dalit, education = no completed primary, sex = male.

Table A3.8 – Robustness: IV estimates of the determinants of trust, without controlling for agreeableness

Trust measures	Neighborhood	Kin employees	Gen. Trust	Neighborhood	Kin employees	Gen. Trust
	(1)	(2)	(3)	(4)	(5)	(6)
Network density (std)	0.866** (0.370)	-0.832* (0.464)	0.359 (0.621)			
Network size (std)				-0.466** (0.225)	0.447** (0.207)	-0.193 (0.349)
Age	-0.019 (0.012)	0.025*** (0.008)	-0.007 (0.020)	0.010 (0.007)	-0.003 (0.007)	0.004 (0.007)
Female	0.653*** (0.236)	-0.393 (0.381)	0.298 (0.299)	-0.160 (0.165)	0.387*** (0.123)	-0.039 (0.236)
Middle caste	0.695*** (0.213)	-0.670** (0.292)	0.996** (0.468)	0.650** (0.306)	-0.627*** (0.216)	0.977** (0.391)
Upper caste	0.039 (0.328)	-0.197 (0.388)	1.062** (0.522)	0.302 (0.301)	-0.450* (0.258)	1.172*** (0.322)
Married	0.113 (0.166)	-0.171 (0.135)	0.004 (0.210)	-0.024 (0.160)	-0.039 (0.136)	-0.053 (0.168)
Primary completed	0.008 (0.164)	0.067 (0.209)	-0.006 (0.123)	0.090 (0.102)	-0.012 (0.126)	0.029 (0.123)
High school (8th-10th)	0.411** (0.189)	0.572** (0.243)	0.021 (0.212)	0.579*** (0.133)	0.411*** (0.141)	0.090 (0.164)
HSC/Diploma (11th-12th)	0.811*** (0.198)	0.266 (0.431)	0.403** (0.196)	0.701*** (0.248)	0.371* (0.201)	0.357* (0.214)
Bachelors (13th-15th)	0.623 (0.457)	-0.370 (0.364)	0.011 (0.315)	0.282 (0.306)	-0.044 (0.231)	-0.130 (0.228)
Post Grad	1.427*** (0.390)	-0.233 (0.532)	0.826* (0.476)	1.165*** (0.366)	0.018 (0.290)	0.718* (0.375)
Raven (std)	-0.058 (0.086)	0.069 (0.044)	-0.038 (0.120)	-0.062 (0.074)	0.073* (0.039)	-0.040 (0.111)
Literacy (std)	-0.351*** (0.067)	-0.172 (0.123)	-0.064 (0.116)	-0.299*** (0.075)	-0.222*** (0.081)	-0.042 (0.116)
Numeracy (std)	0.178** (0.090)	-0.056 (0.117)	0.111 (0.112)	0.078 (0.084)	0.039 (0.069)	0.070 (0.085)
OP (std)	0.089 (0.083)	0.049 (0.095)	-0.086 (0.118)	0.134* (0.069)	0.005 (0.079)	-0.067 (0.102)
CO (std)	0.291*** (0.102)	0.252** (0.128)	0.194 (0.151)	0.262*** (0.096)	0.280*** (0.088)	0.182 (0.135)
EX (std)	0.356*** (0.099)	0.142 (0.091)	0.409*** (0.141)	0.391*** (0.083)	0.109* (0.060)	0.424*** (0.144)
ES (std)	-0.135 (0.111)	-0.157 (0.097)	-0.266* (0.160)	-0.294*** (0.080)	-0.005 (0.058)	-0.332*** (0.112)
Grit (std)	-0.204* (0.124)	-0.046 (0.112)	0.089 (0.090)	-0.151 (0.159)	-0.097 (0.093)	0.111 (0.082)
Food expenses	-0.171 (0.158)	-0.177 (0.172)	-0.318*** (0.099)	-0.022 (0.127)	-0.320** (0.131)	-0.256* (0.135)
Health expenses	-0.066 (0.052)	0.033 (0.036)	-0.026 (0.076)	-0.038 (0.050)	0.007 (0.041)	-0.014 (0.073)
Ceremony expenses	0.068 (0.063)	0.085 (0.088)	0.017 (0.165)	0.114** (0.054)	0.041 (0.061)	0.037 (0.151)
HH income	0.095 (0.090)	-0.178* (0.100)	0.040 (0.130)	-0.021 (0.071)	-0.067 (0.093)	-0.008 (0.090)
N	663	663	663	663	663	663
R ²			0.208	0.204	0.262	0.284

Notes: Table based on NEEMSIS (2016-2017). Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Village-area fixed effects included. Base categories: caste = Dalit, education = no completed primary, sex = male. Household expense variables are in natural logarithm. First stage F-statistics are for 38.15 network density and 14.15 for network size.

Table A3.9 – Robustness: Second stage IV estimations with migration dummies

	(1) Neigh	(2) Neigh	(3) Kin employees	(4) Kin employees	(5) Gen. trust	(6) Gen. trust
Network density (std)	0.416 (0.677)		-1.350*** (0.520)		-0.238 (0.555)	
Network size (std)		-0.219 (0.331)		0.711** (0.307)		0.125 (0.284)
Migrant HH	0.426 (0.293)	0.381 (0.301)	0.395 (0.264)	0.538** (0.233)	0.193 (0.208)	0.218 (0.223)
Indv is migrant	-0.459* (0.249)	-0.341 (0.239)	-0.025 (0.354)	-0.408*** (0.155)	-0.170 (0.247)	-0.238 (0.199)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	663	663	663	663	663	663
R ²	0.219	0.332		0.091	0.546	0.573

Notes: Table based on NEEMSI (2016-2017). Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Village-area fixed effects included. The usual control variables are included. First stage F-statistics are 17.62 for network density and 22.47 for network size.

Table A3.10 – Robustness: Second stages of IV estimations of determinants of trust across the poverty distribution

	Lowest Poverty tertile			Highest Poverty tertile		
	Neigh	Kin employees	Gen. trust	Neigh	Kin employees	Gen. trust
	(1)	(2)	(3)	(4)	(5)	(6)
Network density	-23.581 (94.281)	-6.099 (26.842)	-12.509 (48.851)	0.348 (0.313)	-1.452** (0.622)	-0.840** (0.335)
F-statistic	0.1			12.9		
N	220	220	220	222	222	222
R ²				0.329		0.385
Network size	-2.061*** (0.552)	-0.533 (0.529)	-1.093*** (0.384)	-0.155 (0.170)	0.644*** (0.200)	0.373* (0.210)
F-statistic	17.3			9.2		
N	222	222	222	220	220	220
R ²	0.384	0.324	0.538		0.350	0.250

Notes: Table based on NEEMSI (2016-2017). Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Village-area fixed effects included. The usual control variables are included.

Table A3.11 – Robustness: Second stages of IV estimations of determinants of trust across the emotional stability distribution

	Lowest ES tertile			Highest ES tertile		
	Neigh (1)	Kin employees (2)	Gen. trust (3)	Neigh (4)	Kin employees (5)	Gen. trust (6)
Network density	0.267 (0.698)	-0.621** (0.294)	0.224 (0.536)	0.563 (0.385)	-1.075*** (0.267)	-0.621** (0.294)
F-statistic	5.28			4.34		
N	219	224	219	224	224	224
R^2	0.472	0.682	0.654	0.327		0.682
Network size	-0.167 (0.405)	0.902* (0.461)	-0.140 (0.310)	-0.311 (0.205)	0.594*** (0.180)	0.343** (0.174)
F-statistic	16.9			7.8		
N	219	219	219	224	224	224
R^2	0.474	0.125	0.654	0.507	0.317	0.821

Notes: Table based on NEEMSIS (2016-2017). Standard errors clustered at the village level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Village-area fixed effects included. The usual control variables are included.

Conclusion

Main results and policy implications

The goal of this thesis is to shed light on the role of cognitive and non-cognitive skills in a context in which they have been rarely looked at so far, developing countries, but in which they might be the most relevant. I have shown that self-assessments suffer from response bias, but their internal validity can be improved by correcting for those. Observational exercises are a promising way to capture the results of programs explicitly aimed at fostering non-cognitive skills, programs that are arguably even more important in developing countries than in developed ones (though current programs seem to mostly start at an age at which malleability of skills is more difficult). I have further shown that non-cognitive skills carry a substantial return in the formal labor market in a developing country, Bangladesh, contributing to the small but growing literature on returns to non-cognitive skills in developing countries. Further, cognitive and non-cognitive skills have proven useful as covariates when exploring the causal determinants of social trust, identified through the demonetization policy in India.

Given the importance of non-cognitive skills, including measures of non-cognitive skills in school surveys could be a crucial tool to better understand current levels among students in developing countries. However, they are difficult to use for accountability purposes, as they can be prone to response bias (as shown in this thesis), faking (if school funding is tied to good scores, for example), or reference bias (students with better peers might rate themselves differently than those with worse peers). In developed countries, recommendations have been to rely on administrative data instead, such as school attendance. This might be difficult in developing countries, where absences might be due to more factors than just a lack of non-cognitive skills, such as teacher absenteeism, social norms preventing school attendance, or a child's need to contribute to the family's income. Relying on school attendance data to measure the impact of non-cognitive skills on later life outcomes also requires reliable tracking of students after leaving school, which could be particularly difficult among the most vulnerable students. Observational exercises during classroom hours with evaluations by trained teachers could be a potential solution. Still, in the case of Mozambique, it remains to be seen which instrument (self-assessment or observational exercise) is actually better able

to predict labor and life outcomes and to assess the long-term sustainability of programs such as *MUVA* Attitude.

Limitations and further research

Lack of data sources capturing cognitive and non-cognitive skills reliably in a developing country remains a challenge. Promising future strands of research in economics certainly relate to establishing causal estimates of the effect of cognitive and non-cognitive skills on life outcomes in developing countries. This would require, ideally, panel data collected during childhood or adulthood (or both if one wishes to see how these skills are being developed over the life course). Chapter 2 makes the assumption that non-cognitive skills do not change after entering a particular job. While this assumption has been validated in the literature, a formal test in developing countries is needed. One could imagine these skills to be stable in a highly developed country like Australia (the country studied in Cobb-Clark and Tan, 2011)); however, if workers' level of skills is low before starting a job, a development while employed is certainly imaginable. Our results on the determinants of wage growth in Bangladesh (chapter 2) indeed seem to suggest that conscientiousness and grit develop during working life.

In the same spirit, additional research is needed for different segments of the labor market. I am only studying the formal sector in Bangladesh, for example, even though this is a small part of the Bangladeshi labor market. Focusing on sectors that are more important in developing countries, such as the informal sector or agriculture and extending the analysis to women and oftentimes excluded groups is necessary. The literature has shown that rewards to skills differ by gender, and oftentimes reflect pre-existing stereotypes (women being rewarded for being more observant and agreeable, men being rewarded for taking risk and speaking up). This could be particularly important in the developing world, where gender roles remain much more traditional. Chapter 3 has shown that existing social hierarchies determine outcomes, in the chapter studied, levels of trust. Extending the analysis of the effect of non-cognitive skills to traditionally excluded groups could thus help understand whether individual skills enable these individuals to overcome hierarchies, or whether the latter predominates.

The biggest implication of this thesis is the need to develop better measurement instruments, ideally with the help of psychologists from local environments. The internal validity of most constructs used is lower than those obtained for developed countries, hampering the constructs ability to draw meaningful results; this holds especially true for the instruments included in chapters 1 and 2, which were taken as given. Field work in India as part of chapter 3 highlighted the difficulties of translating Western concepts of personality into local languages and making it accessible to low-literacy populations, while still making the resulting scales comparable to other countries. Resulting internal validity for these constructs

is higher than in the other chapters, showcasing that it does seem possible to conduct these types of surveys in developing countries, if emphasis is being placed on careful translation and discussions of cultural differences captured in language with the local survey team. While the literature has claimed that instruments such as the Big Five are valid across cultures, most of these cultural validates stem from highly developed countries. Engaging more with local contexts to assure cross-cultural validity of the constructs used is necessary, but also limits cross-country comparisons.

Bibliography

- Abay, K. A., Blalock, G., and Berhane, G. (2017). Locus of control and technology adoption in developing country agriculture: Evidence from Ethiopia. *Journal of Economic Behavior & Organization*, 143:98–115.
- Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High Wage Workers and High Wage Firms. *Econometrica*, 67(2):251–333.
- Acevedo, P., Cruces, G., Gertler, P., and Martinez, S. (2017). Living Up to Expectations: How Job Training Made Women Better Off and Men Worse Off. Technical report, National Bureau of Economic Research.
- Achenbach, T. M., McConaughy, S. H., and Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, 101(2):213.
- Acosta, P. A., Muller, N., and Sarzosa, M. (2015). Beyond qualifications: returns to cognitive and socio-emotional skills in Colombia. *World Bank Policy Research Working Paper no. 7430*.
- Alesina, A. and La Ferrara, E. (2000). Participation in Heterogeneous Communities. *The Quarterly Journal of Economics*, 115(3):847–904.
- Algan, Y., Bouguen, A., Charpentier, A., Chevallier, C., Huillery, E., and Solnon, A. (2016). Évaluation de l’impact du programme Énergie Jeunes. Rapport intermédiaire 2016 – Impact sur les élèves du 6eme. Technical report, available at: http://www.energiejeunes.fr/Etude_impact_Energie_Jeunes_2016.pdf.
- Allport, G. W. and Odbert, H. S. (1936). Traitnames. A Psycho-Lexical Study. *Psychological Monographs*, 47(171).
- Almlund, M., Duckworth, A. L., Heckman, J., and Kautz, T. (2011). Personality psychology and economics. In *Handbook of the Economics of Education*, volume 4, pages 1–181. Elsevier.
- Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist’s companion*. Princeton University Press.
- Asadullah, M. N. (2006). Returns to Education in Bangladesh. *Education Economics*, 14(4):453–468.
- Asian Development Bank and Bangladesh Bureau of Statistics (2012). *The Informal Sector and Informal Employment in Bangladesh*. Asian Development Bank, Manila.
- Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, s., and Rubio-Codina, M. (2015). Estimating the production function for human capital: Results from a randomized control trial in Colombia. Technical report, National Bureau of Economic Research.
- AVSI (2016). Comportamento dos jovens e desafios da empregabilidade. Presentation.

- Bandura, A. (1999). Social Cognitive Theory of Personality. In John, O. P., Robins, R. W., and Pervin, L. A., editors, *Handbook of Personality: Theory and Research. Third Edition*, pages 154–196.
- Beaman, L. (2016). Social Networks and the Labor Market. In Bramoullé, Y., Galeotti, A., and Rogers, B., editors, *The Oxford Handbook of the Economics of Networks*. Oxford University Press.
- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Columbia University Press, NY.
- Bentolila, S., Michelacci, C., and Suarez, J. (2010). Social contacts and occupational choice. *Economica*, 77(305):20–45.
- Berardi, N. (2013). Social networks and wages in Senegal’s labor market. *IZA Journal of Labor & Development*, 2(1):3.
- Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1):122–142.
- Blom, A. and Saeki, H. (2011). Employability and skill set of newly graduated engineers in India. *World Bank Working Paper no. 5640*.
- Borghans, L., Duckworth, A. L., Heckman, J. J., and ter Weel, B. (2008a). The Economics and Psychology of Personality Traits. *Journal of Human Resources*, 43(4):972–1059.
- Borghans, L., Meijers, H., and Ter Weel, B. (2008b). The role of noncognitive skills in explaining cognitive test scores. *Economic Inquiry*, 46(1):2–12.
- Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450.
- Bourdieu, P. (1980). Le capital social. *Actes de la recherche en sciences sociales*, 31(1):2–3.
- Bourguignon, F., Fournier, M., and Gurgand, M. (2007). Selection Bias Corrections Based on the Multinomial Logit Model: Monte-Carlo Comparisons. *Journal of Economic Surveys*, 21(1):174–205.
- Braeken, J. and Van Assen, M. A. (2017). An empirical Kaiser criterion. *Psychological Methods*, 22(3):450.
- Breman, J. (1996). *Footloose labour: working in India’s informal economy*, volume 2. Cambridge University Press.
- Bryk, A. S. and Raudenbush, S. W. (1992). *Hierarchical linear models: applications and data analysis methods*. Sage Publications, Inc.
- Burks, S. V., Cowgill, B., Hoffman, M., and Housman, M. (2015). The Value of Hiring Through Employee Referrals. *The Quarterly Journal of Economics*, 130(2):805–839.

- Caliendo, M., Cobb-Clark, D., and Uhlenhorff, A. (2015). Locus of control and job search strategies. *Review of Economics and Statistics*, 97(1):88–103.
- Cameron, A. C. and Miller, D. L. (2015). A practitioner’s guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372.
- Card, D. (1999). The causal effect of education on earnings. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3, pages 1801–1863.
- Card, D., Ibararán, P., Regalia, F., Rosas-Shady, D., and Soares, Y. (2011). The labor market impacts of youth training in the Dominican Republic. *Journal of Labor Economics*, 29(2):267–300.
- Cattell, R. B. (1966). The Scree Test for the Number of Factors. *Multivariate Behavioral Research*, 1(2):245–276.
- Cawley, J., Heckman, J., and Vytlačil, E. (2001). Three observations on wages and measured cognitive ability. *Labour Economics*, 8(4):419–442.
- Cobb-Clark, D. A. and Schurer, S. (2012). The stability of big-five personality traits. *Economics Letters*, 115(1):11–15.
- Cobb-Clark, D. A. and Tan, M. (2011). Noncognitive skills, occupational attainment, and relative wages. *Labour Economics*, 18(1):1–13.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4):213.
- Costa, P. T. and McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, 13(6):653–665.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297–334.
- Crowe, L., Beauchamp, M., Catroppa, C., and Anderson, V. (2011). Social function assessment tools for children and adolescents: A systematic review from 1988 to 2010. *Clinical Psychology Review*, 31(5):767–785.
- Cunha, F. and Heckman, J. J. (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources*, 43(4):738–782.
- Cunningham, W., Acosta, P., and Muller, N. (2016). *Minds and Behaviors at Work: Boosting Socioemotional Skills for Latin America’s Workforce*. World Bank Publications.
- Cunningham, W. V. and Villaseñor, P. (2016). Employer voices, employer demands, and implications for public skills development policy connecting the labor and education sectors. *The World Bank Research Observer*, 31(1):102–134.
- Delattre, E. and Sabatier, M. (2007). Social capital and wages: An econometric evaluation of social networking’s effects. *Labour*, 21(2):209–236.
- Delhey, J. and Newton, K. (2003). Who trusts?: The origins of social trust in seven societies. *European Societies*, 5(2):93–137.

- Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics*, 132(4):1593–1640.
- Díaz, J., Arias, O., and Tudela, D. (2013). Does Perseverance Pay as Much as Being Smart? The Returns to Cognitive and Non-Cognitive Skills in Urban Peru. *Unpublished paper, World Bank, Washington, D.C.*
- Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2008). Representative trust and reciprocity: Prevalence and determinants. *Economic Inquiry*, 46(1):84–90.
- Doll, B. and Elliott, S. N. (1994). Representativeness of observed preschool social behaviors: How many data are enough? *Journal of Early Intervention*, 18(2):227–238.
- Drago, F. (2011). Self-esteem and earnings. *Journal of Economic Psychology*, 32(3):480–488.
- Duckworth, A. L., Peterson, C., Matthews, M. D., and Kelly, D. R. (2007). Grit: Perseverance and Passion for Long-Term Goals. *Journal of Personality and Social Psychology*, 92(6):1087.
- Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., and Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences*, 108(19):7716–7720.
- Duckworth, A. L. and Seligman, M. E. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, 16(12):939–944.
- Elkins, I. J., King, S. M., McGue, M., and Iacono, W. G. (2006). Personality traits and the development of nicotine, alcohol, and illicit drug disorders: Prospective links from adolescence to young adulthood. *Journal of Abnormal Psychology*, 115(1):26.
- Elkins, R. K., Kassenboehmer, S. C., and Schurer, S. (2017). The stability of personality traits in adolescence and young adulthood. *Journal of Economic Psychology*, 60:37–52.
- Elliott, S., Frey, J., and Davies, M. (2015). Systems for assessing and improving students’ social skills to achieve academic competence. *Handbook of Social and Emotional Learning: Research and Practice*, pages 301–319.
- Eysenck, H. J., Eysenck, M. W., et al. (1987). *Personality and Individual Differences*. Plenum New York, NY.
- Fafchamps, M. (2006). Development and Social Capital. *The Journal of Development Studies*, 42(7):1180–1198.
- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global Evidence on Economic Preferences. *The Quarterly Journal of Economics*, forthcoming.
- Fearon, J. D., Humphreys, M., and Weinstein, J. M. (2009). Can development aid contribute to social cohesion after civil war? Evidence from a field experiment in post-conflict Liberia. *The American Economic Review*, 99(2):287–291.

- Fehr, E. (2009). On the Economics and Biology of Trust. *Journal of the European Economic Association*, 7(2-3):235–266.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378.
- Ford, J. K., MacCallum, R. C., and Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, 39(2):291–314.
- Fortin, N. M. (2008). The Gender Wage Gap among Young Adults in the United States: The Importance of Money versus People. *Journal of Human Resources*, 43(4):884–918.
- Freitag, M. and Bauer, P. C. (2016). Personality traits and the propensity to trust friends and strangers. *The Social Science Journal*, 53(4):467–476.
- Fukuyama, F. (1995). *Trust: The social virtues and the creation of prosperity*. Free Press Paperbacks.
- Ganimian, A., Barrera-Osorio, F., Loreto Biehl, M., and Cortelezzi, M. (2018). Hard Cash and Soft Skills: Experimental Evidence on Combining Scholarships and Mentoring in Argentina. Technical report, available at: <https://www.aeaweb.org/conference/2018/preliminary/paper/Ga3K3drH>.
- Ghosh, J., Chandrasekhar, C., and Patnaik, P. (2017). *Demonetisation Decoded: A Critique of India's Currency Experiment*. Routledge India.
- Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., and Soutter, C. L. (2000). Measuring Trust. *The Quarterly Journal of Economics*, 115(3):811–846.
- Glanville, J. L., Andersson, M. A., and Paxton, P. (2013). Do social connections create trust? An examination using new longitudinal data. *Social Forces*, 92(2):545–562.
- Glewwe, P., Huang, Q., and Park, A. (2017). Cognitive skills, noncognitive skills, and school-to-work transitions in rural China. *Journal of Economic Behavior & Organization*, 134:141–164.
- Golsteyn, B. and Schildberg-Hörisch, H. (2017). Challenges in research on preferences and personality traits: Measurement, stability, and inference. *Journal of Economic Psychology*, (60):1–6.
- Granovetter, M. (1973). The Strength of Weak Ties. *The American Journal of Sociology*, 78(6):1360–1380.
- Granovetter, M. (1995). *Getting a job: A study of contacts and careers*. Chicago University Press.
- Griliches, Z. (1977). Estimating the Returns to Schooling: Some Econometric Problems. *Econometrica*, 45(1):1–22.
- Groshen, E. L. (1991). Sources of Intra-Industry Wage Dispersion: How Much Do Employers Matter? *The Quarterly Journal of Economics*, 106(3):869–884.

- Guérin, I. (2008). Poor women and their money: Between daily survival, private life, family obligations and social norms. *Rural Microfinance and Employment Project (RUME) Working Paper Series*.
- Guérin, I., D'Espallier, B., and Venkatasubramanian, G. (2013). Debt in rural South India: Fragmentation, social regulation and discrimination. *The Journal of Development Studies*, 49(9):1155–1171.
- Guérin, I., Lanos, Y., Michiels, S., Nordman, C. J., and Venkatasubramanian, G. (2017). Insights on Demonetisation from Rural Tamil Nadu: Understanding Social Networks and Social Protection. *Economic and Political Weekly*, 52(52).
- Guérin, I., Michiels, S., and Venkatasubramanian, G. (2015). Labour in Contemporary South India. In Harriss-White, B. and Heyer, J., editors, *Indian Capitalism in Development*, pages 118–36. Routledge.
- Guiso, L., Sapienza, P., and Zingales, L. (2008). Social Capital as Good Culture. *Journal of the European Economic Association*, 6(2-3):295–320.
- Haddad, L. and Maluccio, J. A. (2003). Trust, membership in groups, and household welfare: Evidence from KwaZulu-Natal, South Africa. *Economic Development and Cultural Change*, 51(3):573–601.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L., et al. (1998). *Multivariate Data Analysis*, volume 5. Prentice Hall, NJ.
- Hanushek, E. A., Schwerdt, G., Wiederhold, S., and Woessmann, L. (2015). Returns to Skills Around the World: Evidence from PIAAC. *European Economic Review*, 73:103–130.
- Hanushek, E. A. and Woessmann, L. (2008). The Role of Cognitive skills in Economic Development. *Journal of Economic Literature*, 46(3):607–668.
- Hanushek, E. A. and Zhang, L. (2009). Quality-consistent estimates of international schooling and skill gradients. *Journal of Human Capital*, 3(2):107–143.
- Harriss-White, B. and Colatei, D. (2004). Rural Credit and the Collateral Question. In Harris-White, B. and Janakarajan, S., editors, *Rural India Facing the 21st Century. Essays on Long Term Change and Recent Development Policy*, pages 252–283. Anthem South Asian Studies.
- Hayton, J. C., Allen, D. G., and Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2):191–205.
- Heath, R. (2018). Why do firms hire using referrals? Evidence from Bangladeshi Garment Factories. *Journal of Political Economy*. forthcoming.
- Heckman, J. J. and Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, 91(2):145–149.
- Heckman, J. J. J., Stixrud, J., and Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24(3):411–482.

- Heineck, G. and Anger, S. (2010). The Returns to Cognitive Abilities and Personality Traits in Germany. *Labour Economics*, 17(3):535–546.
- Herrmann, B., Thöni, C., and Gächter, S. (2008). Antisocial Punishment Across Societies. *Science*, 319(5868):1362–1367.
- Hills, P. and Argyle, M. (2001). Emotional stability as a major dimension of happiness. *Personality and Individual Differences*, 31(8):1357–1364.
- Holzer, H. J. (1988). Search Method Use by Unemployed Youth. *Journal of Labor Economics*, 6(1):1.
- Hong, K. and Bohnet, I. (2007). Status and distrust: The relevance of inequality and betrayal aversion. *Journal of Economic Psychology*, 28(2):197–213.
- Hopkins, D. J. and King, G. (2010). Improving anchoring vignettes: Designing surveys to correct interpersonal incomparability. *Public Opinion Quarterly*, 74(2):201–222.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2):179–185.
- Humphrey, N., Kalambouka, A., Wigelsworth, M., Lendrum, A., Deighton, J., and Wolpert, M. (2011). Measures of social and emotional skills for children and young people: A systematic review. *Educational and Psychological Measurement*, 71(4):617–637.
- Humphries, J. E. and Kosse, F. (2017). On the interpretation of non-cognitive skills—What is being measured and why it matters. *Journal of Economic Behavior & Organization*, 136:174–185.
- Ibarrarán, P., Kluve, J., Ripani, L., and Rosas, D. (2015). Experimental evidence on the long-term impacts of a youth training program. Technical report, Inter-American Development Bank.
- John, O. P. and Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and research*, 2(1999):102–138.
- Johny, J., Wichmann, B., and Swallow, B. M. (2017). Characterizing social networks and their effects on income diversification in rural Kerala, India. *World Development*, 94:375–392.
- Jones, G. (2008). Are smarter groups more cooperative? Evidence from prisoner’s dilemma experiments, 1959–2003. *Journal of Economic Behavior & Organization*, 68(3):489–497.
- Judge, T. A., Higgins, C. A., Thoresen, C. J., and Barrick, M. R. (1999). The Big Five Personality Traits, General Mental Ability, and Career Success across the Life Span. *Personnel Psychology*, 52(3):621–652.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200.

- Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., and Borghans, L. (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. Technical report, OECD Education Working Papers, No. 110, OECD Publishing, Paris.
- Knack, S. and Keefer, P. (1997). Does social capital have an economic payoff? A cross-country investigation. *The Quarterly Journal of Economics*, 112(4):1251–1288.
- Krishnan, P. and Krutikova, S. (2013). Non-cognitive skill formation in poor neighbourhoods of urban India. *Labour Economics*, 24:68–85.
- Kuépié, M., Nordman, C. J., and Roubaud, F. (2009). Education and earnings in urban West Africa. *Journal of Comparative Economics*, 37(3):491–515.
- Kugler, A. D. (2003). Employee Referrals and Efficiency Wages. *Labour Economics*, 10(5):531–556.
- Laaajaj, R. and Macours, K. (2017). Measuring skills in developing countries. *World Bank Policy Research Working Paper 8000*.
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, pages 159–174.
- Lazarsfeld, P. F. and Merton, R. K. (1954). Friendship as a social process: A substantive and methodological analysis. In Berger, M., Abel, T., and Page, C., editors, *Freedom and Control in Modern Society*, volume 18, pages 18–66.
- Lee, J. and Newhouse, D. (2012). Cognitive Skills and Youth Labor Market Outcomes. *Background Paper for the 2012 World Development Report. The World Bank*.
- Lee, L., Li, J., and Lin, X. (2014). Binary Choice Models with Social Network under Heterogeneous Rational Expectations. *Review of Economics and Statistics*, 96(3):402–417.
- Lin, N. (1999). Building a Network Theory of Social Capital. *Connections*, 22(1):28–51.
- Lin, N. (2001). *Social capital: A theory of social structure and action*. Cambridge university press.
- Lindqvist, E. and Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. *American Economic Journal: Applied Economics*, 3(1):101–28.
- Marsden, P. V. (2005). Recent Developments in Network Measurement. In Carrington, P. J., Scott, J., and Wasserman, S., editors, *Models and methods in social network analysis*, pages 8–30. Cambridge University Press.
- McCallister, L. and Fischer, C. S. (1978). A Procedure for Surveying Personal Networks. *Sociological Methods & Research*, 7(2):131–148.
- McConaughy, S. H. and Ritter, D. R. (1995). Best practices in multidimensional assessment of emotional or behavioral disorders. In Thomas, A. and Grimes, J., editors, *Best practices in School Psychology - III*, pages 865–877. National Association of School Psychologists Washington, DC.

- McCrae, R. and Costa, P. T. (2008). A five-factor theory of personality. In John, O. P., Robins, R. W., and Pervin, L. A., editors, *Handbook of Personality: Theory and Research (3rd edition)*. The Guilford Press.
- McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27:415–444.
- Meng, X. (2004). Gender earnings gap: The role of firm specific effects. *Labour Economics*, 11(5):555–573.
- Merrell, K. W. (2003). *Behavioral, Social, and Emotional Assessment of Children and Adolescents*. Psychology Press.
- Messick, S. (1978). Potential uses of noncognitive measurement in education. *Journal of Educational Psychology*, 71(3):281.
- Miller, A. S. and Mitamura, T. (2003). Are Surveys on Trust Trustworthy? *Social Psychology Quarterly*, pages 62–70.
- Mincer, J. (1974). *Schooling, Experience, and Earnings*. Columbia University Press, NY.
- Mischel, W. and Shoda, Y. (1999). Integrating dispositions and processing dynamics within a unified theory of personality: The cognitive-affective personality system. In John, O. P., Robins, R. W., and Pervin, L. A., editors, *Handbook of personality: Theory and Research. Third Edition*, pages 197–218.
- Mischel, W. and Shoda, Y. (2008). Toward a unified theory of personality: Integrating dispositions and processing dynamics within the cognitive-affective processing system. In John, O. P., Robins, R. W., and Pervin, L. A., editors, *Handbook of Personality: Theory and Research (3rd edition)*. Guilford Press, New York.
- Montgomery, J. (1991). Social networks and labor-market outcomes: Toward an economic analysis. *The American Economic Review*, 81(5):1408–1418.
- Mueller, G. and Plug, E. (2006). Estimating the effect of personality on male and female earnings. *Industrial & Labor Relations Review*, 60(1):3–22.
- Muldrew, C. (1998). *The economy of obligation: The culture of credit and social relations in early modern England*. Springer.
- Munshi, K. (2016a). Caste and the Indian Economy. Technical report, available at: http://www.histecon.magd.cam.ac.uk/km/Munshi_JEL1.pdf.
- Munshi, K. (2016b). Caste networks in the modern Indian economy. In *Development in India*, pages 13–37. Springer.
- Munshi, K. and Rosenzweig, M. (2009). Why is mobility in India so low? Social insurance, inequality, and growth. Technical report, National Bureau of Economic Research.
- Munshi, K. and Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(1):46–98.

- Nakata, S. and Nomura, S. (2015). Informal Job Referral and Wages in Formal Sector Labor market in Bangladesh. *Unpublished paper, World Bank, Washington, D.C.*
- Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51(2):77.
- Nomura, S., Hong, S. Y., Nordman, C. J., Sarr, L. R., and Vawda, A. Y. (2013). *An assessment of skills in the formal sector labor market in Bangladesh : a technical report on the enterprise-based skills survey 2012*. The World Bank, Washington, D.C.
- Nooteboom, B. (2007). Social Capital, Institutions and Trust. *Review of Social Economy*, 65(1):29–53.
- Nordman, C. J., Sarr, L. R., and Sharma, S. (2015). Cognitive, Non-Cognitive Skills and Gender Wage Gaps: Evidence from Linked Employer-Employee Data in Bangladesh. *IZA Discussion Paper No. 9132*.
- Nordman, C. J. and Wolff, F. C. (2009). Is there a glass ceiling in Morocco? Evidence from matched worker - firm data. *Journal of African Economies*, 18(4):592–633.
- Nyhus, E. K. and Pons, E. (2005). The effects of personality on earnings. *Journal of Economic Psychology*, 26(3):363–384.
- OECD (2015). *Skills for Social Progress*. OECD Publishing, Paris.
- Pierre, G., Sanchez Puerta, M. L., Valerio, A., and Rajadel, T. (2014). STEP skills measurement surveys: innovative tools for assessing skills.
- Pollet, T. V., Roberts, S. G., and Dunbar, R. I. (2011). Extraverts have larger social network layers: But do not feel emotionally closer to individuals at any layer. *Journal of Individual Differences*, 32(3):161–169.
- Primi, R., Zanon, C., Santos, D., De Fruyt, F., and John, O. P. (2016). Anchoring vignettes. Can They Make Adolescent Self-Reports of Social-Emotional Skills More Reliable, Discriminant, and Criterion-Valid? *European Journal of Psychological Assessment*.
- Psacharopoulos, G. and Patrinos, H. A. (2004). Returns to Investment in Education: A Further Update. *Education Economics*, 12(2):111–134.
- Putnam, R. D. (2001). *Bowling alone: The collapse and revival of American community*. Simon and Schuster.
- Rammstedt, B. and Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological Assessment*, 25(4):1137.
- Rammstedt, B., Kemper, C. J., and Borg, I. (2013). Correcting Big Five personality measurements for acquiescence: An 18-country cross-cultural study. *European Journal of Personality*, 27(1):71–81.
- Ravallion, M. (2012). Poor, or just feeling poor? On using subjective data in measuring poverty.

- Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, 43(2):137–145.
- Roberts, B. W. and DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological Bulletin*, 126(1):3.
- Rohner, D., Thoenig, M., and Zilibotti, F. (2013). Seeds of distrust: Conflict in Uganda. *Journal of Economic Growth*, 18(3):217–252.
- Salomon, J. A., Tandon, A., and Murray, C. J. (2004). Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes. *The BMJ*, 328(7434):258.
- Sanchez Puerta, M. L., Valerio, A., and Gutiérrez Bernal, M. (2016). *Taking Stock of Programs to Develop Socioemotional Skills*. Washington, DC: World Bank.
- Serneels, P. (2008). Human capital revisited: The role of experience and education when controlling for performance and cognitive skills. *Labour Economics*, 15(6):1143–1161.
- Shapiro, E. S. and Kratochwill, T. R. (2000). *Behavioral assessment in schools: Theory, research, and clinical foundations*. Guilford Press.
- Simon, C. J. and Warner, J. T. (1992). Matchmaker, Matchmaker: The Effect of Old Boy Networks on Job Match Quality, Earnings, and Tenure. *Journal of Labor Economics*, 10(3):306.
- Slemrod, J. and Katuščák, P. (2005). Do trust and trustworthiness pay off? *Journal of Human Resources*, 40(3):621–646.
- Söderbom, M., Teal, F., Wambugu, A., and Kahyarara, G. (2006). The dynamics of returns to education in Kenyan and Tanzanian manufacturing. *Oxford Bulletin of Economics and Statistics*, 68(3):261–288.
- Soto, C. J., John, O. P., Gosling, S. D., and Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology*, 94(4):718.
- Stock, J. H., Wright, J. H., and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20(4):518–529.
- Stolle, D., Soroka, S., and Johnston, R. (2008). When does diversity erode trust? Neighborhood diversity, interpersonal and the mediating effect of social interactions. *Political Studies*, 56(1):57–75.
- Topa, G. (2010). Labor markets and referrals. In Benhabib, J., Bisin, A., and Jackson, M., editors, *Handbook of Social Economics, Volume 1B*, pages 1193–1221. Elsevier.
- Uher, J. and Asendorpf, J. B. (2008). Personality assessment in the Great Apes: Comparing ecologically valid behavior measures, behavior ratings, and adjective ratings. *Journal of Research in Personality*, 42(4):821–838.

- Uslaner, E. M. (2002). *The Moral Foundations of Trust*. Cambridge University Press.
- Vignoles, A. and McIntosh, S. (2001). Measuring and assessing the impact of basic skills on labour market outcomes. *Oxford Economic Papers*, 53(3):453–481.
- Vijayabaskar, M. and Kalaiyarasan, A. (2014). Caste as Social Capital. *Economic & Political Weekly*, 49(10):35.
- West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F., and Gabrieli, J. D. (2016). Promise and paradox: Measuring students’ non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis*, 38(1):148–170.
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press.
- World Bank (2013). *Seeding fertile ground : education that works for Bangladesh*. World Bank, Washington, D.C.
- World Bank (2015). *Bangladesh - More and Better Jobs to Accelerate Shared Growth and Extreme Poverty. A Systemic Country Diagnostic*. The World Bank, Washington, D.C.
- World Bank (2017). *Mozambique Economic Update: A two-speed economy*. The World Bank, Washington, DC.
- World Bank (2018). Skills for Tomorrow’s Jobs:Preparing youth for a fast-changing economy. Technical report, World Bank.
- Wu, P.-C., Foo, M.-D., and Turban, D. B. (2008). The role of personality in relationship closeness, developer assistance, and career success. *Journal of Vocational Behavior*, 73(3):440–448.

List of Tables

A0.1	Description of the Big Five personality model	11
1.1	Sample composition	21
1.2	Mean acquiescence scores by cycle, period, and gender	30
1.3	Number of factors to retain according to different criteria	31
1.4	EFA - Rotated factor loadings at baseline	33
1.5	Comparing aggregations of the self-assessment - Cronbach's α	34
1.6	Inter-rater reliability for observational exercises	36
1.7	Cronbach's α for the combined observational exercises	38
1.8	Correlations between self-assessment and observations	39
1.9	Correlations between skills over time	40
1.10	Average rank of skills	42
A1.1	Descriptive statistics for panel sample, cycle 1 and cycle 2	47
A1.2	Observational exercise assessment criteria	48
A1.3	Exploratory factor analysis - Rotated factor loadings at Midline	50
A1.4	Exploratory factor analysis - Rotated factor loadings at Endline	51
A1.5	Number of raters per period, % of sample	52
A1.6	Deviation from mean in observational scores	52
A1.7	Correlations between different measures of non-cognitive skills	53
A1.8	Correlations of skills with number of training sessions missed	54
2.1	Log hourly wages regressed on cognitive and non-cognitive skills, OLS	72
2.2	Probability of being hired through social networks: marginal effects after probit	77
2.3	Returns to skills using an endogenous switching model	78
2.4	Returns to skills using an endogenous switching model - Starting and current wage	82
2.5	Determinants of the within-firm hiring channel wage gap	86
A2.1	Correlations between the skills measures	91
A2.2	Characteristics of firms in samples	91
A2.3	Characteristics of employees	92
A2.4	Returns to non-cognitive skills with skills derived from exploratory factor analysis	93

A2.5	Variance inflation factor for different OLS models	94
A2.6	OLS and endogenous switching models without education as a covariate	94
A2.7	Log earnings regressed on cognitive and non-cognitive skills and determinants of hours worked per week – OLS	95
A2.8	Probability of being hired through different hiring channels: marginal effects after multinomial logit	96
A2.9	Endogenous switching model with interactions for being a white collar worker	97
A2.10	Returns to skills using an endogenous switching model - Multiple hiring channels	98
A2.11	Characteristics of firms used in wage gap analysis sample	99
3.1	Distribution of trustworthy agents and cheaters	108
3.2	Change of trust due to demonetization	120
3.3	IV estimates of the determinants of trust – Network size	127
3.4	IV estimates of the determinants of trust – Network density	128
3.5	Estimates of determinants of trust by caste membership for both measures of social networks	130
3.6	Estimates of determinants of trust by gender for both measures of social networks	133
3.7	Distribution of employment status by caste membership	135
A3.1	Descriptive statistics of the individual level sample	142
A3.2	Balance checks - unmatched samples	143
A3.3	Balance checks - matched samples	145
A3.4	OLS estimates of the determinants of trust	146
A3.5	Estimates of determinants of trust by employment status for both measures of social networks – Non-salaried occupations	147
A3.6	Estimates of determinants of trust by employment status for both measures of social networks – Salaried occupations	148
A3.7	IV estimates of the determinants of trust, using days passed since demonetization as instrument	149
A3.8	IV estimates of the determinants of trust, without controlling for agreeableness	150
A3.9	IV estimates with migration dummies	151
A3.10	IV estimates of determinants of trust across the poverty distribution	151
A3.11	IV estimates of determinants of trust across the emotional stability distribution	152

List of Figures

1.1	Sequence of <i>MUVA</i> Attitude program and data collection	20
1.2	Acquiescence scores over the age range and years of schooling	30
1.3	Difference between ranking by self-assessment and observational exercise	43
A1.1	Share of people neither agreeing nor disagreeing with a statement .	49
A1.2	Scree plots for naive self-assessment	49
A1.3	Scree plots for acquiescence corrected self-assessment	49
2.1	Conceptual framework illustrated	61
A2.1	Current wages by hiring channel and industry	92
A2.2	Within firm wage gap	99
3.1	Map of the study region	105
A3.1	Common support for matching	144
A3.2	Relationship between days passed since demonetization and networks	149