GENNOVATE – Machine Learning (ML)

Pre-analysis plan

Gendered Dreams and Realities: A Machine Learning Analysis of Qualitative Data on Community Gender Climate and Youth Aspirations

Els Lecoutere, Caroline Muchiri, Niyati Singaraju, Gideon Kruseman

Date: 2 December 2025

1. Relevance and novelty

1.1. Relevance of the study topic

A rapidly expanding youth population is at the forefront of the profound demographic shifts shaping many low- and middle-income countries (LMICs), accompanied by rapid urbanization and evolving forms of internal and international migration (Tabutin et al. 2020; FAO 2022; UNDESA 2024). This large and growing youth cohort represents an unprecedented force for shaping the future of rural –and urban–economies in LMICs.

Yet, in 2023, 29 percent of youth in low-income countries are not in employment, education, or training (NEET) and two times more likely to be unemployed than adults; in lower-middle-income countries, 23 percent of youth are NEET and four times more likely to be unemployed. Young women are more likely to be NEET than young men (37 percent vs 20 percent in low-income countries; 35 percent vs 12 percent in lower-middle-income countries) (ILO 2024). Education, marriage, and parenthood influence NEET status differently by gender—education benefits young women but not young men in rural areas, while family responsibilities have the opposite pattern. Many young women classified as NEET may, in fact, be engaged in unpaid domestic and care work (Heckert et al. 2021).

At the same time, there is a continuing rural out-migration to urban areas by youth, which is accelerating in some regions (Nedumaran et al. 2025; Enahoro et al. 2025). Rates of internal migration vary widely by region and country, and by gender. Overall, female youth often internally migrate earlier, due to marriage, while young men tend to migrate later, primarily for employment, often after completing education (FAO, 2025). Following the upward trend of international migration, international youth migration has also grown. Youth make up about 15 percent of international migrants from LMICs, and there are approximately equal shares of men and women among young international migrants (FAO, 2025).

Understanding the gender dynamics that shape young women's and young men's aspirations, their local and distant opportunities, and their employment prospects in LMICs, as well as the factors driving these dynamics, is essential for informing policies and actions that help realize the demographic dividend by expanding opportunities for young people to pursue fulfilling jobs and livelihoods they value (Sumberg and Okali 2013; Elias et al. 2018; GENDER Impact Platform 2024; FAO 2025).

The research question guiding this study examines the relationship between community gender climate and gendered occupational aspirations and aspiration–attainment gaps among youth in LMICs. The study relies on a unique large N qualitative dataset—GENNOVATE—with rich qualitative data collected using standardized methods from a large set of communities across LMICs. To test the research hypotheses, the study applies Machine Learning Analysis.

1.2. Relevance of Machine Learning Methods

The complexity and scale of qualitative data—such as life stories and focus group discussions—from a large set of communities across LMICs present both a challenge and an opportunity. Traditional qualitative analysis, while rich in context, is often time-consuming, difficult to scale, and limited in its ability to detect subtle patterns across diverse cultural and socioeconomic settings. **Machine learning (ML) and natural language processing (NLP) methods** offer a transformative approach to systematically extract, quantify, and analyze themes from large qualitative datasets. By leveraging these methods, we can:

- **Uncover hidden patterns**: Identify nuanced relationships between community gender climates, aspirations, and aspiration-attainment gaps that may not be apparent through manual coding alone.
- **Scale analysis**: Process and compare data across a large set of communities efficiently, enabling cross-regional and cross-contextual insights.
- **Quantify qualitative data**: Transform narrative responses into structured variables (e.g., thematic frequencies, sentiment scores) that can be integrated into statistical models, such as regression or mediation analysis.

For example, **topic modelling** and **sentiment analysis** can help quantify the prevalence of restrictive or supportive gender norms across communities, while **named entity recognition (NER)** can extract and categorize mentions of occupations, resources, and barriers. These quantified variables can then be used to test hypotheses about how community gender climate shapes gendered youth aspirations and employment outcomes, addressing gaps in the existing literature (Petesch et al., 2018; Elias et al., 2018).

1.3. Novelty of the approach

This study introduces two innovations in the application of machine learning to qualitative data in the context of gender and youth aspirations:

1. **Context-Sensitive Quantification**: Unlike traditional qualitative analysis, which often relies on manual coding and small-scale interpretation, our approach uses **supervised and unsupervised ML techniques** to systematically quantify themes while preserving contextual nuances. For instance, we employ **qualitative comparative analysis (QCA)** alongside ML to identify combinations of gender climate factors (e.g., decision-making autonomy, mobility, governance) that correlate with specific aspiration outcomes. This hybrid approach bridges the gap between qualitative depth and quantitative rigor, offering a more robust understanding of how gender dynamics operate in diverse LMIC contexts.
2. **Mediation Analysis with Qualitative Data**: While mediation analysis is commonly used in quantitative research, its application to qualitative data remains underexplored. By extracting and quantifying mediators—such as **perceived gender norms** or **access to resources**—from text data, we can test theoretical pathways (e.g., how gender climate influences aspirations through perceived opportunities). This method allows us to move beyond descriptive analysis to **causal inference**, providing actionable insights for policy interventions (Hayes, 2018; Lopez et al., 2022).

A third innovation relates to the preparation of qualitative data for structured analysis using machine learning methods:

3. **Ethical and Scalable Automation**: The use of **large language models (LLMs)** for initial coding, combined with human validation, represents a novel balance between efficiency and accuracy. This hybrid approach ensures that the richness of qualitative data is preserved while reducing the burden of manual analysis. Moreover, by focusing on **aggregated, anonymized patterns**, we address ethical concerns about reidentification, making the method suitable for sensitive datasets in LMICs.

The integration of ML and NLP into qualitative analysis is not just a technical advancement—it is a **paradigm shift** in how we study gender dynamics and youth aspirations in LMICs. Traditional methods often struggle to capture the **contextual fluidity** of gender norms and opportunity spaces (Sumberg and Okali, 2013; Elias et al., 2018). By quantifying qualitative data, we can:

- **Identify leverage points** for interventions, such as targeting communities with restrictive gender climates or addressing specific barriers (e.g., access to land or education).
- **Inform gender-inclusive policies** by providing evidence on how norms and opportunities interact to shape gendered youth aspirations and employment.
- **Scale insights** across regions, enabling comparative analysis that can inform global and local development strategies.

2. Theory, empirical foundation, research questions, and hypotheses

2.1. Theoretical foundation and supporting prior evidence

A large body of scholarship shows that young people's aspirations emerge within an *opportunity space*, defined as the spatial and temporal distribution of more or less viable livelihood options available to them (Sumberg & Okali, 2013). This opportunity space is shaped by both socio-economic conditions (access to local and distant livelihood opportunities) and socio-relational factors, including gender norms and relations. Aspirations are also informed by dominant cultural ideas about what constitutes a worthy future, which are themselves gendered, as well as by an individual's personal history and circumstances (Zipin et al., 2015).

The local **normative climate**—defined as the formal and informal institutions that regulate social interactions within a community—and, in particular, the **gender climate**, understood as the contextually specific and fluid set of gender norms and their interaction with other local dynamics, shape women's and men's sense of agency and life opportunities, including those of young women and men (Petesch et al. 2018a, López et al. 2022). The gender climate and gender norms shape how young women and men imagine their futures by influencing self-concepts, perceptions of occupational opportunity space and status (Leavy and Smith 2010, p. 9); through internalized gender norms (Armstrong and Crombie 2000; Overå 2007) and through social expectations (Petesch et al. 2018a).

Gender norms are central to how youth imagine their futures. They influence self-concepts, perceived occupational possibilities, and perceived status attached to different livelihood options (Leavy & Smith, 2010). Youth who internalize restrictive norms may view certain occupations—especially agricultural or leadership roles—as inappropriate or unattainable (Armstrong & Crombie, 2000; Overå, 2007).

Young people often experience an **'aspiration–attainment' gap**, reflecting a disconnect between what they aspire to and what their socio-cultural and economic circumstances make possible (Leavy and Smith 2010, p. 9) Gender differences in "aspiration–attainment" gaps can arise when the gender normative climate, and its inherent gender norms, does not only affect what young women and men *aspire* to—also differentially affects their capacity to *achieve* those aspirations.

Importantly, it is not only community-level gender climate and norms that shape aspirations, but also how young women and men internalize those norms and respond to the associated social expectations and sanctions (Pearse and Connell, 2016; Boudet et al., 2013). The 'gender climate' framework (Petesch et al., 2018a; López et al., 2022) emphasizes that norms are contextual, fluid, and experienced differently across communities but also within communities. Therefore, the relationship between the gender climate and young women's and men's aspirations may be **mediated** by the way they perceive and internalize the prevailing gender norms. If they perceive these as restrictive, they may internalize expectations that limit their aspirations.

Prior evidence shows the relevance of gender climate and its diversity. Lopez et al. (2022) demonstrate that gender climate is conceptually distinct from, and not correlated with, non-gendered community demographic, economic, or infrastructural characteristics. They also show that the degree to which the community-level gender climate is restrictive or relaxed varies not only between but also within country cohorts of communities. Other literature emphasizes the pertinence of the gender climate and the embedded gendered opportunity space and norms for young women and men in agricultural LMIC societies. Leavy & Smith (2010) argue that young people's willingness to consider farming as a viable component of their livelihoods is positively related to their access to productive resources such as land, credit, labour and other resources required for more commercial forms of agriculture.

Elias et al. (2018) show that gender norms—as a component of the opportunity space— play a central role in shaping youths' aspirations and the extent to which they can pursue them. For instance, in many agricultural communities, restrictive gender norms not only limit young women's access to land and capital but also their exposure to information, training, networks, and new agricultural practices while prescribing early marriage and domestic roles as priorities. Together, these discourage young women from aspiring to modern agriculture-related or other occupations beyond traditional agriculture and homemaking. In contrast, young men may be expected to migrate or assume responsibility for family livelihoods, even when their own aspirations may lie elsewhere. Elias et al. (2018) demonstrate that gendered norms and opportunity structures also shape the extent to which young women and men are able to pursue and achieve their aspirations. Rietveld et al. (2020) argue that, in rural and agricultural contexts, aspirations are often shaped by local gender norms, household resources, exposure to innovation, and intergenerational attitudes toward farming and mobility. They illustrate that youth in Uganda commonly navigate multiple livelihood strategies over time—investing in or intensifying farming, engaging in off-farm income diversification, and migrating. The opportunities, interests, and capacities that shape youth's strategies differ sharply by gender, reflecting persistent inequalities in access to land and social relations structured by family responsibilities and financial dependence, which together contribute to the marginalization of young women. Alwab et al. (2023) illustrate that, although formal barriers to gender equality have eased over time with successive regimes in Ethiopia, cumulative gender- and age-related constraints—many grounded in norms, traditions, and social relations—continued to limit adolescents' aspirations and achievements, with girls consistently facing greater restrictions than boys.

Elias et al., (2018) point out that the conversations also revealed that young women and men often aspire to futures that challenge existing gender hierarchies. Aspirational sentiments, for example, young women expressing a desire for education, leadership, or entrepreneurship, and young men supporting women's empowerment or cooperative household roles, reflect emerging shifts toward greater gender equality. Even when such aspirations are not yet realized, they illuminate the direction of social and normative change. They suggest that when youth encounter relaxed gender climates and enabling institutions, their aspirations can translate into transformative action and innovation within agriculture and beyond (Elias et al., 2018). Understanding youth aspirations and local gender climates (the degree of restrictiveness or relaxation) offers critical insight into the interplay between social norms, opportunity structures, and agency. Restrictive gender norms constrain both young women's and men's ability to envision and pursue diverse livelihoods, while relaxed, egalitarian climates expand the space for innovation, equality, and youth-driven transformation in rural communities.

2.2. Research question(s) and theory-based research hypotheses

Based on these theoretical and empirical foundations, we derived a conceptual framework, visualised in Figure 1, and six research hypotheses about the relationships between community gender climate and gendered occupational aspirations and aspiration-attainment gaps among youth in LMICs.

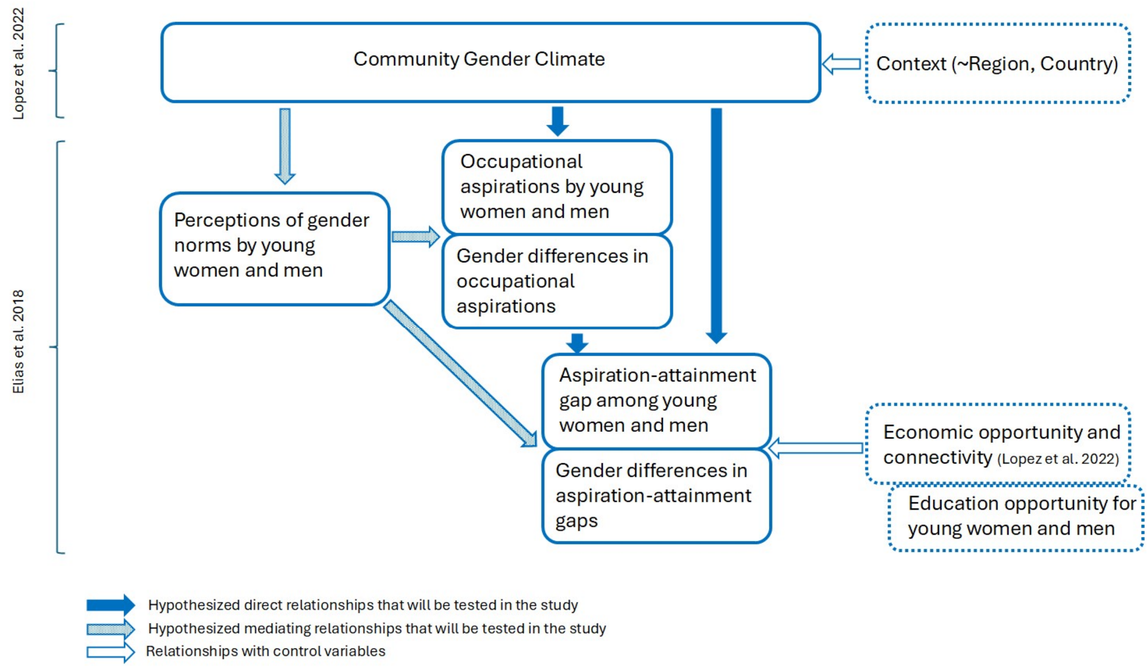


Figure 1. Conceptual framework

Research question: What are the relationships between community gender climate and gendered occupational aspirations and aspiration-attainment gaps among youth?

Research hypotheses about the direct relationship between gender climate and aspirations

i. A more relaxed (vs restrictive) community gender climate is correlated with young women and men occupational aspirations for jobs and careers beyond traditional agriculture and/or homemaker
ii. Young women's and men's occupational aspirations for jobs and careers beyond traditional agriculture and/or homemaker are more aligned in a more relaxed (vs restrictive) community gender climate
iii. Aspiration-attainment gaps among young women and men are less likely and less wide in a more relaxed (vs restrictive) community gender climate
iv. Gender differences in aspiration-attainment gaps are smaller in a more relaxed (vs restrictive) community gender climate

Research hypotheses about the mediating role of perceptions of gender norms

v. A more relaxed (vs restrictive) community gender climate is correlated with young women and men perceiving less restrictive/less discriminatory gender norms
vi. Young women's and men's perceptions of gender norms influence gendered occupational aspirations and aspiration-attainment gaps

Gender climate, opportunity space, norms, aspirations, and achievements–and how these relate–are all embedded within broader contextual conditions, including regional and country characteristics and community-level demographic, economic, and infrastructural features, that also shape educational opportunities.

3. Data

The analysis for this paper will draw on an **anonymised GENNOVATE dataset**, which comprises openly accessible data from 108 rural communities across 23 countries across Africa, Asia and Latin America.[1,2] This represents a substantial subset of the original GENNOVATE sample of 137 communities in 26 countries. The retrieved dataset has been carefully curated, consolidated, and anonymised in accordance with rigorous ethical standards to prevent any direct or indirect identification of individuals or communities, and has been prepared in alignment with the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles (Muchiri et al., 2024).

However, we will use a subset of 66 communities, of which we have the necessary contextual variables calculated by Lopez et al (2022) using original data, and of which we have access to the anonymised data.[3]

Gender climate

- Adopting the conceptualisation and operationalisation by Lopez et al. 2022
- ***Instruments***=Community Profile (Key Informant Interviews KII); Ladder of Life (Focus Group Discussions FGD); Innovation capacities (FGD); Aspirations of Youth (FGD) - (See Table 2 p 6, Lopez et al. 2022)

Youth aspirations, youth perceptions of gender norms

- **Youth in Gennovate:** the age of male participants ranged from 14-29 while women were 13-35 years of age
- ***Instrument***=Aspirations and perceptions of gender norms by youth captured in FGD with young women and FGD with young men; carried out per community (Elias et al. 2018).

  Research in GENNOVATE studies explored youth aspirations through focus group discussions (FGDs) within communities, specifically with older adolescents and young adults aged 16–24, typically divided by gender to capture differences in opportunity and perspective.

  Each group had 8–12 participants and a minimum of 6 directly engaged in agricultural or natural resource management livelihoods (production, collection, processing, or trade). Groups were composed to reflect diversity in education, socio-economic status, and marital experience, while excluding individuals who were markedly more educated or advantaged than their peers to avoid dominance in discussion. The discussion guide probed core topics to understand young people's views and solicit answers to the question *"What changes would you like to see in the village for the young men and women who live here?"*

  Key themes included agency over strategic life decisions; gender norms surrounding education and aspirations; enabling and constraining factors for agricultural and natural resource management innovation (both hardware and software innovations); gender norms shaping household bargaining over housework and care roles; women's economic roles, access to networks, control of productive assets, physical mobility; and family formation norms and practices (Petesch et al., 2018c).

Control variables

Lopez et al. (2022) used a variant of principal component analysis on non-gendered data of community characteristics (mainly drawn from community profiles developed through key informant interviews) and

---

[1] Bangladesh, India, Nepal, Pakistan, Ethiopia, Tanzania, Zimbabwe, Mexico, Rwanda, Burundi, DRC Congo, Kyrgyz Republic, Colombia, Uzbekistan, Morocco, Vietnam, Uganda, Malawi, Kenya, Nigeria, Burkina Faso, Niger, and Mali.

[2] The rich metadata of the anonymised GENNOVATE dataset has been developed and is already available. At the time of developing this pre-analysis plan, the anonymised GENNOVATE dataset is not yet openly accessible but is expected to become openly accessible and FAIR in the first quarter of 2026.

[3] Lopez et al. (2022) included 70 communities, including 4 from Afghanistan. The latter are not included in the open-access, anonymized GENNOVATE data for ethical reasons. Hence, we have 66 in our subset of communities.

identified two distinct components, unrelated to gender climate, which they labelled **'Opportunity'** and **'Connectivity'**. They then assigned values for each component per community. We will use these components as the control variables.

4. Method

4.1. Definition of key concepts and their indicators

4.1.1. Restrictive and relaxed (egalitarian) gender climate

The concept of **gender climate**, which is inspired by Petesch et al. (2018a) concept of 'local normative climate', refers to "the prevailing set of gender norms in a community, and how they interact with other dynamics in that same context to differentially shape women's and men's sense of agency and opportunities in their lives" (Lopez et al. 2022, p3).

The gender climate differs across communities depending on how supportive or restrictive gender norms are toward women's and men's participation in education, decision-making, mobility, leadership, and access to assets and opportunities. Petesch et al. (2018a) describe these as existing along a continuum, with some gender norms being more **restrictive** and others more **relaxed** or **egalitarian**.

A **restrictive gender climate** is characterized by formal and informal norms that limit women's mobility, decision-making power, educational opportunities, and access to assets and resources (Lopez et al. 2022). Such environments are often marked by women's economic dependence on men, their exclusion from leadership roles, and higher reported levels of gender-based violence. Typically, men control agricultural income and make key decisions over both commercial and subsistence crops or livestock, and permanent male migration may further reinforce women's constrained agency. **Restrictive** norms also manifest in family formation practices, domestic responsibilities, and community expectations that confine women's roles to the household.

Conversely, a **relaxed or egalitarian gender climate** features social norms that support women's education, leadership, and freedom of movement, and enable women to access and control productive assets and resources. Empirical analyses by Lopez et al. (2022) show that in such contexts, women are more likely to make or contribute to decisions regarding agriculture and income use, have the freedom to make strategic life decisions, participate in markets, work for pay in agricultural and non-agricultural jobs, and own and use technologies such as mobile phones. A **relaxed** gender climate also implies the absence of reported violence against women and recognizes their right to actively participate in governance and leadership and move freely in public spaces. Lopez et al., (2022) show that the dimensions most strongly linked with a relaxed gender climate are decision-making, mobility, and governance.

4.1.2. Gender norms

Social norms are rules of action shared by people in a society or group that specify what behaviors are regarded as normal and acceptable for the members of that group (Cislaghi and Heise, 2018a: in Cislaghi and Heise, 2020). **Gender norms** are a subset of the broader social norms that define what is considered appropriate behavior, responsibilities, and opportunities for men, women, girls, and boys within a given community at a particular point in time (Cislaghi and Heise, 2020). As social institutions, gender norms regulate people's daily interactions and shape their identity, freedom, voice, and access to resources. Unlike formal laws, they are maintained through internalized beliefs and collective expectations that individuals should act in gender-appropriate ways.

4.1.3. Youth aspirations and aspiration-attainment gap

The definition of **youth** is context-specific and debated. The United Nations (UN) defines youth as persons between the ages of 15 to 24 years. Yet youth is more than a demographic category (age group); it is a dynamic and transitional life stage characterized by identity formation, intergenerational linkages, social navigation, and the pursuit of future possibilities (GENDER Impact Platform, Youth Position Paper; CGIAR Youth Strategy, *Forthcoming*). During this period, young people navigate social expectations and opportunity structures while shaping their own visions of adulthood and success.

Aspirations are a widely used concept in youth studies. **Aspirations** are understood as the "ability to identify and set goals for the future, while being inspired in the present to work toward those goals" (Quaglia and Cobb, 1996, p. 130; in Elias et al., 2018). **Occupational aspirations** relate to employment and careers or entrepreneurship, while educational aspirations relate to education and training.

The **aspiration–attainment gap** refers to the distance between what youth hope to achieve and the social and economic realities they face. This gap is particularly evident in rural contexts where educational and employment opportunities are scarce, and where social norms constrain young people's choices, especially those of young women's. For example, a young woman may aspire to become a teacher or entrepreneur but face barriers such as limited schooling, restricted mobility, and gendered expectations around caregiving. The persistence of such gaps reflects both material and normative inequalities that hinder young people from realizing their full potential.

4.1.4. Indicators of key concepts

- **Gender climate** is operationalised through five dimensions: mobility, gender-based violence, decision-making, governance and leadership, and education; each measured through a set of variables (Lopez et al. 2022).
  We will follow the method by Lopez et al. 2022 to operationalise gender climate and **classify communities' gender climate by their degree of restriction or relaxation**. The variables per dimension and criteria for the degree of restriction or relaxation per variable are listed in Table 2 p 6 and p 7-9. Further details are available in the online supplementary resources (Lopez et al. 2022).

- **Perceptions of gender norms by young women and men**
  We will follow the method used by Elias et al. (2018) and the coding tree they applied (documented in the GENNOVATE Codebook)[4] to analyze the structural factors, including gender norms, that define men's and women's ability to capture opportunities in agriculture. Where relevant and possible, we will follow the same typology of norms as Elias et al. (2018): i) 'Men are (stronger and better) farmers'; ii) Norms restricting young women's mobility and the gender division of labour; iii) Norms limiting women's access to assets and credit; iv) Norms limiting women's access to information; v) Social stigmas against breaking gender norms

- **Occupational aspirations by young women and men**
  Following Elias et al. (2018), occupational aspirations will be operationalized by what young women and men verbally express as future job or livelihood goals they strive towards. We will follow the method described in

[4] GENNOVATE Codebook Annex: 5a_Gender norms: "Gender norms are the "typical" and "appropriate" roles/conducts for that gender in their daily life, i.e. women must keep a close watch on children, or should never talk to men in public. Assess norms around roles. capacities, conducts, and how they affect access/use/benefits from innovs or instns, Do not use norm nodes for individual practices/conducts unless associated w/ wider expectations/reports of how that gender does, should or ought to behave."

Elias et al. (2018) and the coding tree they applied (documented in the GENNOVATE Codebook)[5]. To examine occupational aspirations by young women and men, they analyzed how education came up in the respondents' narratives and contributed to shaping their aspirations in and beyond agriculture.

In line with Elias et al. (2018), occupational aspirations can be classified by aspirations in the domain of i) traditional agriculture and/or homemaker vs ii) modern agriculture, or jobs as an agri-preneur or agricultural professional, or iii) blue- or white-collar jobs.

- **Gender differences in occupational aspirations**
  Gender differences in occupational aspirations will be operationalised by the ***within-community*** **difference** in the extent to which young women and men express occupational aspirations for jobs and careers **beyond traditional agriculture and/or homemaker** (*class I of occupational aspirations)*

- **Aspiration-attainment gap among young women and men**
  Following Elias et al. (2018), we will analyze aspiration-attainment gaps among young women and men by looking into the overlap and discrepancies between their occupational aspirations and realities as expressed by the respondents. There are two obvious **classes**: i) one where there is no discrepancy between the occupational aspiration and what is/can be achieved; ii) another one where a discrepancy exists. In case the data is rich enough, we may consider refining a classification based on the width of aspiration-attainment gaps.

- **Gender differences in aspiration-attainment gap**
  Gender differences in aspiration-attainment gaps will be operationalised by the ***within-community*** **difference in the likelihood** of the existence aspiration-attainment gaps among young women versus young men. In case the data is rich enough, we may consider to also look at the difference in the width of aspiration-attainment gaps among young women versus young men within the community.

- **Control variables**
  - On the one hand, **context** will be captured by the region and country in which the communities are located.
  - On the other hand, contextual demographic, economic, and infrastructural characteristics, including educational opportunities, will be captured by the indicators (economic) **Opportunity** and **Connectivity** as defined and operationalised by Lopez et al. (2022) (Details in Table 2, p 6)

    Communities with a large population, a high share of households selling their agricultural produce in local markets, good (agricultural/non-agricultural) job or vocational training, and access to a secondary school score have a relatively high 'Opportunity' score. Communities with limited access to local markets, health clinics, secondary schools, and to job or vocational training, which also tend to be smaller, have a relatively low 'Opportunity' score (for details see Lopez et al. 2022 p12, Fig 3) (Scores per community in Annex 2 copied from Lopez et al. (2022) Table 3).

    Communities with good access to communication and infrastructure services such as internet access, the presence of a secondary school, electricity, and a bus line within a 30-minute walk, have a relatively high 'Connectivity' score. In contrast, communities with limited or no access to electricity, secondary school or bus line, and greater distance to the nearest town with government offices have a relatively low

[5] GENNOVATE Codebook Annex: Aspirations+: "All FGDs end with aspirations for youth, and youth are asked about their future goals and in E.10 and desired roles in E.12. Xcode as relevant the objects of the aspirations, i.e.aspiration to be a nurse, policeman, etc can be xcoded to the non-agri livelihoods. Statements of how things should be can be coded here -- ie. what age woman should end their education or should start having babies. Be CAREFUL with "should" statements, as they can also indicate normative conducts rather than aspirations."

'Connectivity' score (see Lopez et al. 2022 p12, Fig 3, and scores per community in Table 3) (Scores per community in Annex 2 copied from Lopez et al. (2022) Table 3).

4.2. Human learning

The next step involves a human-led learning phase, where researchers conduct qualitative analysis of text from a selection of relevant data instruments across 30 randomly selected communities used as learning cases. This analysis will identify examples of text, narratives, expressions, and sentiments that correspond to the key classification concepts: (i) gender climate; (ii) young women's and men's perceptions of gender norms; (iii) occupational aspirations of young women and men; (iv) within-community gender differences in occupational aspirations; (v) aspiration–attainment gaps for young women and men; and (vi) gender differences in aspiration–attainment gaps within communities. These example text excerpts will then be used to train the machine learning model through multiple iterations, enabling it to accurately recognize these concepts and classify respondents' experiences, narratives, expressions, and sentiments in the full dataset.

4.3. Analytical methods

Unit of analysis=Community

Descriptive methods

- Descriptives aggregated by country, organised by region, of the key indicators: Gender climate, Occupational aspirations by young women and men, gender differences in occupational aspirations, aspiration-attainment gap among young women and men, and gender differences in aspiration-attainment gap
- Illustrative quotes of the key indicators by country, organised by region

Analytical methods to test research hypotheses (RH)

<table>
<tr><th>Research hypotheses about the direct relationship between gender climate and aspirations</th><th>Indicators</th><th>Analytical method</th></tr>
<tr><td>i. A more relaxed (vs restrictive) community gender climate is correlated with young women and men occupational aspirations for jobs and careers beyond traditional agriculture and/or homemaker</td><td>x=Gender climate
y=Occupational aspirations by young women and men</td><td rowspan="4">A) Analysis of correlation between x and y across communities
B) Within-community analysis of correlation between x and y; while 'controlling' for context. Backed by illustrative quotes
C) Detection of patterns of correlation and/or cross-country comparison. Backed with illustrative quotes</td></tr>
<tr><td>ii. Young women's and men's occupational aspirations for jobs and careers beyond traditional agriculture and/or homemaker are more aligned in a more relaxed (vs restrictive) community gender climate</td><td>x=Gender climate
y=Gender differences in occupational aspirations</td></tr>
<tr><td>iii. Aspiration-attainment gaps among young women and men are less likely and less wide in a more relaxed (vs restrictive) community gender climate</td><td>x=Gender climate
y=Aspiration-attainment gap among young women and men</td></tr>
<tr><td>iv. Gender differences in aspiration-attainment gaps are smaller in a more relaxed (vs restrictive) community gender climate</td><td>x=Gender climate
y=Gender differences in aspiration-attainment gap</td></tr>
</table>

Research hypotheses about the mediating role of perceptions of gender norms		
v. A more relaxed (vs restrictive) community gender climate is correlated with young women and men perceiving less restrictive/less discriminatory gender norms	x=Gender climate y=Perceptions of gender norms by young women and men	
vi. Young women's and men's perceptions of gender norms influence gendered occupational aspirations and aspiration-attainment gaps	x=Gender climate y=Perceptions of gender norms by young women and men z_1=Occupational aspirations; and z_2=Aspiration-attainment gap among young women and men	A) Analysis of mediation by y in the relationship between x and z_1/z_2 across communities B) Within-community analysis of mediation by y in the relationship between x and z_1/z_2; while 'controlling' for context. Backed with illustrative quotes C) Detection of patterns of mediation and/or cross-country comparison. Backed with illustrative quotes

4.4. Machine learning

To systematically analyse the large-scale qualitative dataset—comprising life stories and focus group discussions from 66 communities—we employ **machine learning (ML) and natural language processing (NLP) techniques**. These methods enable us to **quantify qualitative themes**, **identify patterns**, and **test hypotheses** about the relationships between gender climate, aspirations, and aspiration-attainment gaps. Below, we outline the specific ML approaches, their application, and their integration with traditional qualitative and statistical analyses.

4.4.1 Data Preparation for Machine Learning

1. **Text Segmentation**:
    a. Life stories and focus group discussions are segmented into meaningful units (e.g., sentences or paragraphs) to facilitate coding and analysis.
    b. Metadata (e.g., gender, SES, community) are aligned with each text segment.
2. **Codebook Operationalization**:
    a. The codebook (Section 4.1) is operationalized for ML by defining **keywords, phrases, and rules** for each theme (e.g., "Restrictive Gender Norm," "Occupational Aspirations").
    b. Example: The theme *"Restrictive Gender Norm"* is associated with phrases like *"women cannot," "not allowed,"* or *"men decide."*
3. **Training Data**:
    a. A subset of **30 communities** (as described in Section 4.3) is manually coded by researchers to create a **labelled dataset** for supervised learning.
    b. This dataset includes examples of text segments corresponding to key concepts (e.g., gender climate, perceptions of norms, aspirations).

4.4.2 Machine Learning Techniques

1. **Supervised Learning for Thematic Coding**

- **Objective**: Automate the coding of text segments based on the predefined codebook.
- **Method**:

- **Text Classification**: Train a classifier (e.g., Logistic Regression, Random Forest, or BERT) to label text segments with relevant themes (e.g., "Restrictive Gender Norm," "Occupational Aspirations").
- **Validation**: Use **k-fold cross-validation** to assess model performance. Human coders review a random sample of ML-labelled segments to ensure accuracy.

- **Output**: Each text segment is assigned one or more thematic labels, along with a **confidence score** (e.g., 0–1).

2. Unsupervised Learning for Theme Discovery

- **Objective**: Identify latent themes or patterns not explicitly defined in the codebook.
- **Method**:
  - Topic Modelling (Latent Dirichlet Allocation or BERTopic): Discover recurring themes across the corpus. Topic modelling helps uncover latent themes in large qualitative datasets. It aligns with the Gennovate code book developed by Petesch et al. for validation. The approach requires tuning for each identified topic. It may miss nuanced or rare themes. The approach requires sufficient Natural Language Programming expertise for setup and validation.
  - Clustering (K-Means or Hierarchical Clustering): Group similar text segments to identify emergent patterns.
- **Output**: Themes and clusters are reviewed by researchers to validate their relevance to the research questions.

3. Sentiment Analysis

- **Objective**: Quantify the tone and attitude expressed in text segments (e.g., positive/negative sentiments toward gender norms or aspirations).
- **Method**:
  - Use pre-trained sentiment analysis models (e.g., VADER, TextBlob) to score text segments.
  - Customize sentiment lexicons to capture domain-specific language (e.g., "I feel trapped" → negative sentiment).
  - Sentiment analysis is easy to implement using pre-trained models and quantifies attitudes, for instance towards gender norms. The approach may oversimplify complex narratives. It is especially useful for strong sentiments and less useful for neutral or mixed sentiments. There may be biases in sentiment analysis as the models are trained on western sentiments (Jones-Garcia et al., 2021).
- **Output**: Sentiment scores (e.g., -1 to 1) for each text segment.

4. Named Entity Recognition (NER)

- **Objective**: Extract and categorize key entities (e.g., occupations, resources, institutions).
- **Method**:
  - Use pre-trained NER models (e.g., spaCy) to identify entities like *"teacher," "land,"* or *"microfinance group."*
  - Fine-tune the model on a labelled subset of the data to improve accuracy for domain-specific terms.
  - NER is useful for opportunity space analysis as it extracts key entities such as occupations and resources. However, it requires custom training for domain-specific terms and may be limited by the translation quality of the interview transcriptions. This Approach requires data labelling on top of pretrained models.
- **Output**: Counts and categories of entities per text segment or community.

4.4.3 Integration with Qualitative and Statistical Analysis

1. **Quantification of Qualitative Data**:
    a. Thematic labels, sentiment scores, and entity counts are aggregated by **individual** and **community** to create structured variables for statistical analysis.
    b. Example: *"Restrictive_Norm_Count"* or *"Aspiration_Sentiment_Score"* per community.
2. **Mediation and Regression Analysis**:
    a. Quantified variables are used as **independent, dependent, or mediator variables** in statistical models (see Section 4.3 for variable descriptions).
    b. Example: Testing whether *perceived gender norms* (mediator) explain the relationship between *gender climate* (independent variable) and gendered *aspiration-attainment gaps* (dependent variable).
3. **Qualitative Comparative Analysis (QCA)**:
    a. ML-extracted themes are used in QCA to identify **combinations of conditions** (e.g., gender climate + SES) that lead to specific outcomes (e.g., high aspiration-attainment gaps).

4.4.4 Ethical Considerations and Validation

- **Human-in-the-Loop**:
    - ML-labelled data is validated by human coders to ensure accuracy and contextual relevance.
    - Discrepancies are used to refine the ML models iteratively.
- **Anonymization**:
    - All text data is anonymized to prevent reidentification of communities or individuals.
    - Aggregated results are reported at the group level (e.g., by gender, SES, or region).
- **Transparency**:
    - The use of ML methods, including prompts, models, and validation steps, is documented to ensure reproducibility.

4.4.5 Expected Outputs

- **Structured Dataset**: A quantified dataset with variables for thematic frequencies, sentiment scores, and entity counts, linked to metadata (e.g., gender, SES).
- **Statistical Models**: Results from regression, mediation, and QCA analyses, including effect sizes, confidence intervals, and illustrative quotes.
- **Visualizations**: Thematic maps, bar charts, and network graphs to communicate patterns and relationships.

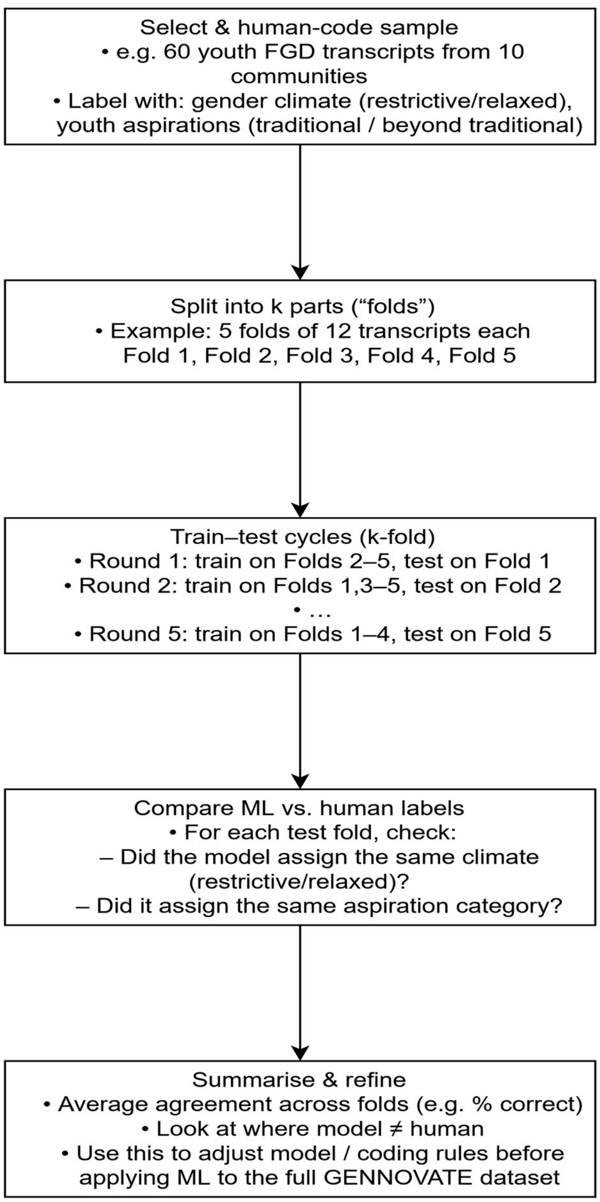
4.4.6 Limitations and Mitigations

- **Small Sample Size**:
    - While 66 communities provide rich qualitative data, the sample size limits the complexity of ML models. We mitigate this by using **simpler models** (e.g., logistic regression) and **bootstrapping** for robust estimates.
- **Contextual Nuances**:
    - ML models may miss cultural or contextual nuances. We address this through **human validation** and **iterative refinement** of the codebook and models.
- **Bias in ML Models**:
    - Pre-trained models may reflect biases. We customize models using **domain-specific training data** and audit results for fairness.

4.5. Benchmarking for internal validation

To ensure that the machine-learning model correctly identifies the concepts and relationships defined in this study (e.g., restrictive vs. relaxed gender climate, youth aspirations categories, perceptions of gender norms), we will benchmark the model against a "human-coded" subset of the data. As explained above, a small sample of transcripts will be coded manually by trained researchers using the definitions and indicators described in the Method section (following Lopez et al., 2022; Elias et al., 2018).

The machine-learning model will then not only be trained on part of this human-coded sample but also tested on another part, rotating these portions so that all coded data contribute to both training and testing. This step-by-step comparison allows us to check how closely the model's classifications match human judgement and refine the model where needed. This internal validation ensures that the ML model is learning the same meaning of the concepts that guide GENNOVATE analysis.



4.6. Robustness checks and external validation

After the model performs well on the human-coded sample, we will compare the patterns it produces with findings from published GENNOVATE studies that used similar concepts and indicators (e.g., Lopez et al., 2022; Elias et al., 2018; Petesch et al., 2018a; Rietveld et al., 2020). For example, we will examine whether communities the model identifies as having a more relaxed gender climate also display the characteristics described in these studies (such as wider mobility for women, more equitable decision-making, or more diverse youth aspirations).

This comparison helps confirm that the ML-derived patterns are consistent with established qualitative findings in similar contexts. Our approach follows the validation principles used in Jones-Garcia et al. (2021) and Lasdun et al. (2024), where ML insights were checked against known qualitative interpretations to ensure reliability and relevance.

5. References

References related to the study topic

Alwab, B. G., Lecoutere, E., & Jones, N. (2022). Adolescents’ capabilities and aspirations across gender and generations in Amhara, Ethiopia. *Journal of Eastern African Studies*, *16*(3), 472–494. https://doi.org/10.1080/17531055.2022.2162191

Armstrong, P. I., & Crombie, G. (2000). Compromises in adolescents' occupational aspirations and expectations from Grades 8 to 10. *Journal of Vocational Behavior*, *56*(1), 82–98. https://doi.org/10.1006/jvbe.1999.1709

Boudet, A. M., Petesch, P., Turk, C., & World Bank. (2013). *On norms and agency: Conversations about gender equality with women and men in 20 countries*. World Bank. https://doi.org/10.1596/978-0-8213-9862-3

Cislaghi, B., & Heise, L. (2018). Theory and practice of social norms interventions: eight common pitfalls. *Global Health*, *14*, Article 83. https://doi.org/10.1186/s12992-018-0398-x

Cislaghi, B., & Heise, L. (2020). Gender norms and social norms: Differences, similarities and why they matter in prevention science.[1] *Sociology of Health & Illness*, *42*(2), 407–422. https://doi.org/10.1111/1467-9566.13008

Elias, M., Mudege, N., Lopez, D. E., Najjar, D., Kandiwa, V., Luis, J., Yila, J., Tegbaru, A., Ibrahim, G., Badstue, L., Njuguna-Mungai, E., & Bentaibi, A. (2018). Gendered aspirations and occupations among rural youth, in agriculture and beyond: A cross-regional perspective. *Journal of Gender, Agriculture and Food Security*, *3*(1), 82–107. https://hdl.handle.net/10568/99065

Enahoro, D. K., Mensah, C., & Gbegbelegbe, S. (2025). What do we know about the future of food systems in West and Central Africa? In K. Wiebe & E. Gotor (Eds.), *Part two: What do we know about the future of food systems in selected regions?* (Chapter 17, pp. 98–102). International Food Policy Research Institute. https://hdl.handle.net/10568/175454

Food and Agriculture Organization (FAO). (2022). *The future of food and agriculture – Drivers and triggers for transformation* (The Future of Food and Agriculture, No. 3). FAO. https://doi.org/10.4060/cc0959en

Food and Agriculture Organization (FAO). (2025). *The status of youth in agrifood systems*. FAO. https://doi.org/10.4060/cd5886en

GENDER Impact Platform. (2024). *Contributing to collective global targets for youth through CGIAR research and innovation: Position paper*. CGIAR GENDER Impact Platform. https://hdl.handle.net/10568/168111

Heckert, J., Pereira, A., Doss, C., Myers, E., & Quisumbing, A. (2021). Structural transformation and gendered transitions to adulthood among rural youth: Cross-national evidence from low- and middle-income countries. *Journal of Development Studies*, *57*(4), 614–634. https://doi.org/10.1080/00220388.2020.1808196

International Labour Organization (ILO). (2024). *GET 2024 Report*. ILO. https://www.ilo.org/sites/default/files/2024-11/GET_2024_EN_web4.pdf

Leavy, J., & Smith, S. (2010). *Future farmers: Youth aspirations, expectations and life choices* (Discussion Paper 013).[2] Future Agricultures Consortium. https://www.academia.edu/download/33069740/Leavy_and_Smith_2010_future_farmers.pdf

López, D. E., Frelat, R., & Badstue, L. B. (2022). Towards gender-inclusive innovation: Assessing local conditions for agricultural targeting. *PLOS ONE*, *17*(3), Article e0263771. https://doi.org/10.1371/journal.pone.0263771

Nedumaran, S., Thomas, J., Nandi, R., Padmanabhan, J., & Afari-Sefa, V. (2025). What do we know about the future of food systems in South Asia? In K. D. Wiebe & E. Gotor (Eds.), *Part two: What do we know about the future of food systems in selected regions?* (Chapter 20, pp. 115–120). International Food Policy Research Institute. https://hdl.handle.net/10568/175457

Overå, Ragnhild. (2007). When Men do Women's Work : Structural Adjustment , Unemployment and Changing Gender Relations in the Informal Economy of Accra, Ghana. *The Journal of Modern African Studies. 45*. https://doi.org/10.1017/S0022278X0700287X

Pearse, R., & Connell, R. W. (2016). Gender norms and the economy: Insights from social research. *Feminist Economics*, *22*(1), 30–53. https://doi.org/10.1080/13545701.2015.1078485

Petesch, P., Bullock, R., Feldman, S., Badstue, L. B., Rietveld, A., Bauchspies, W., Tegbaru, A., Yila, J., & Kawarazuka, N. (2018a). Local normative climate shaping agency and agricultural livelihoods in Sub-Saharan Africa. *Journal of Gender, Agriculture and Food Security*, *3*(1), 108–130. https://doi.org/10.22004/ag.econ.293590

Petesch, P., Feldman, S., Elias, M., Badstue, L. B., Najjar, D., Rietveld, A., Bullock, R., Kawarazuka, N., & Luis, J. (2018b). Community typology framed by normative climate for agricultural innovation, empowerment, and poverty reduction.[3] *Journal of Gender, Agriculture and Food Security*, *3*(1), 131–157. https://hdl.handle.net/20.500.11766/8717

Rietveld, A. M., van der Burg, M., & Groot, J. C. J. (2020). Bridging youth and gender studies to analyse rural young women's and men's livelihood pathways in Central Uganda. *Journal of Rural Studies*, *74*, 25–35. https://doi.org/10.1016/j.jrurstud.2020.01.020

Sumberg, J., & Okali, C. (2013). Young people, agriculture, and transformation in rural Africa: An 'opportunity space' approach. *Innovations: Technology, Governance, Globalization*, *8*(1–2), 259–269. https://doi.org/10.1162/INOV_a_00178

Tabutin, D., Schoumaker, B., Coleman, H., Dutreuilh, C., Reeve, P., Tovey, J., & van Hoorn Alkema, B. (2020). The demography of Sub-Saharan Africa in the 21st century: Transformations since 2000, outlook to 2050. *Population*, *75*(2/3), 165–286. https://doi.org/10.3917/popu.2002.0169

United Nations, Department of Economic and Social Affairs (UNDESA). (2024). *World population prospects 2024: Summary of results*. UN DESA/Population Division. https://desapublications.un.org/file/20847/download

Zipin, L., Sellar, S., Brennan, M., & Gale, T. (2015). Educating for futures in marginalized regions: A sociological framework for rethinking and researching aspirations. *Educational Philosophy and Theory*, *47*(3), 227–246. https://doi.org/10.1080/00131857.2013.839376

References related to the methodology

GENNOVATE Metadata (*Forthcoming*)

Jones-García, E., Kruseman, G., & Brown, B. (2021). *Emotion classification and sentiment analysis for sustainable agricultural development: Exploring available tools for analyzing African farmer interviews.* (Integrated Development Program Discussion Paper No. 5). CIMMYT. https://hdl.handle.net/10883/21789

Hayes, A.F., 2018. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* Guilford publications.

Lasdun, V., Güereña, D., Ortiz-Crespo, B., Mutuvi, S., Selvaraj, M., & Assefa, T. (2024). Participatory AI for inclusive crop improvement. *Agricultural Systems*. https://doi.org/10.1016/j.agsy.2024.104054

Lopez, D. E., & Rietveld, A. (2024). *Initial research protocol for re-using and re-analysing GENNOVATE data*. https://hdl.handle.net/10568/173264

Lopez, D.E and Romain Frelat, R. (2022). Towards Gender-Inclusive Innovation: Assessing Local Conditions for Agricultural Targeting. GitHub Pages, last update – 25th March 2022. Retrieved November 20, 2025, from https://rfrelat.github.io/GenderClimate.html

Lopez D.E, Frelat R, Badstue L.B. (2020). Introduction to multivariate analysis for gender and agriculture. *Zenodo*. Retrieved November 20, 2025, from https://zenodo.org/records/4395535

Muchiri, C., Lopez, D. E., & Kruseman, G. (2024). *Making qualitative data open access: Guidance document for making qualitative data FAIR—Findable, accessible, interoperable, and reusable—Using the GENNOVATE case study* (CGIAR GENDER Impact Platform Report). ILRI. https://hdl.handle.net/10568/168114

Petesch, P., Badstue, L., & Prain, G. (2018c). *Gender norms, agency, and innovation in agriculture and natural resource management: The GENNOVATE methodology*. CIMMYT. https://gennovate.org/wp-content/uploads/2018/10/GENNOVATE-Methodology_Feb2018_FINAL.pdf

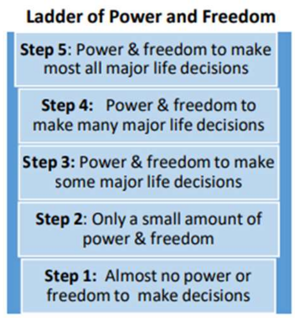
Petesch, P. 2015 (version Sept.). *GENNOVATE Codebook: Enabling Gender Equality in Agricultural and Environmental Innovation*. GENNOVATE (*Internal document*)

6. Annexes

6.1. Annex 1

The GENNOVATE Methodology

Module 1: Agency, education, and innovation – data from the ladder of power & freedom with steps 1 to 5, moving from a more restrictive to more relaxed rating. Consists of gender differences in education opportunities and constraints for innovation (unequal opportunities to learn about and try out new farming practices or agricultural innovations),



Module 2: Gender norms, livelihoods, and family formation

Includes data on the different traditions or customs that surround agricultural activities (10 years ago & current). It also explores how gender norms influence women's control over productive resources and income, their mobility in public spaces, and how these norms constrain or enable their agency and participation in innovation and economic life.

How restrictive or relaxed norms are:

How easy or difficult will it be for Diana to go ahead and purchase the plot of land in the absence of David's support?

Very easy 1
Easy...2
Neither easy nor difficult...... 3
Difficult 4
Very difficult..........................5

For Diana to go ahead and spend her inheritance without David's support?	____

Flipchart 6

How easy or difficult will it be for David to go ahead and purchase the motorbike in the absence of Diana's support?

Very easy 1
Easy...2
Neither easy nor difficult...... 3
Difficult 4
Very difficult..........................5

For David to go ahead and spend his inheritance without Diana's support?	____

Freedom of movement:

Practically no women move freely on their own in the village	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*Practically all women move freely on their own in the village*
Responses											

6.2. Annex 2 'Opportunity' and 'Connectivity' scores per community copied from Lopez et al. (2022) Table 3

Country	Region*	Village Code**	Gender Climate	Opportunity	Connectivity
Afghanistan (AF) n = 4	Kabul	AF1	-3.42	2.90	-0.19
	Nangarhar	AF2	-1.82	3.48	-2.55
	Kabul	AF3	-2.97	-3.77	-0.98
	Nangarhar	AF4	-3.58	1.68	-1.50
Bangladesh (BD) n = 6	Mymensingh	BD1	-2.08	0.38	1.11
	Dhaka	BD2	-1.88	-0.19	-0.33
	Rangpur	BD3	-2.19	1.13	1.59
	Rangpur	BD4	-1.60	1.36	1.68
	Khulna	BD5	-2.15	0.75	0.02
	Rajshahi	BD6	-1.80	0.44	-1.35
Ethiopia (ET) n = 8	Oromia	ET1	0.22	0.74	-2.20
	Oromia	ET2	-0.48	1.66	-1.93
	Amhara	ET3	-0.84	1.51	-1.22
	Amhara	ET4	-0.32	-0.25	-1.51
	Oromia	ET5	-0.24	1.04	-0.14
	Oromia	ET6	0.51	1.61	-1.29
	SNNPR[a]	ET7	0.01	1.39	-2.68
	SNNPR[a]	ET8	-0.81	-0.31	-3.10
India (IN) n = 12	Haryana	IN1	-0.46	0.26	1.77
	Bihar	IN2	0.38	1.66	1.51
	Uttar Pradesh	IN3	0.77	-2.50	1.08
	Madhya Pradesh	IN4	-1.61	-2.53	-0.03
	Bihar	IN5	-1.77	-0.98	0.60
	Madhya Pradesh	IN6	0.58	-2.30	-0.06
	Madhya Pradesh	IN7	0.25	-1.68	-0.27
	Bihar	IN8	-0.94	-1.93	0.40
	Punjab	IN9	-0.94	0.88	2.79
	Uttar Pradesh	IN10	0.15	-1.17	1.33
	Bihar	IN11	-2.25	-3.55	0.02
	Uttar Pradesh	IN12	-0.36	-1.49	0.86
Malawi (MW) n = 2	Central Region	MW1	0.55	-1.45	-1.20
	Central Region	MW2	1.00	-0.74	-3.18
Mexico (MX) n = 6	Oaxaca	MX1	3.11	-0.01	0.24
	Chiapas	MX2	0.79	0.14	0.39
	Oaxaca	MX3	1.03	-0.78	-3.26
	Chiapas	MX4	2.67	0.66	1.72
	Chiapas	MX5	2.99	0.90	0.61
	Oaxaca	MX6	0.88	-2.07	-0.15
Morocco (MO) n = 3	Fes-Meknes	MO1	-1.73	-0.96	1.14
	Fes-Meknes	MO2	-1.13	-1.95	0.96
	Fes-Meknes	MO3	-1.13	-2.32	-0.07
Nepal (NP) n = 6	Bagmati Pradesh	NP1	4.48	-0.11	0.35
	Province No. 5	NP2	2.20	0.77	1.44
	Gandaki Pradesh	NP3	-0.10	0.36	0.26
	Karnali	NP4	1.39	-0.04	0.54
	Province No. 5	NP5	0.87	0.95	1.26
	Gandaki Pradesh	NP6	3.18	1.58	1.62
Nigeria (NI) n = 4	Plateau State	NI1	1.81	-1.05	-1.48
	Oyo State	NI2	2.14	3.62	0.32
	Kaduna State	NI3	0.88	1.76	1.36
	Oyo State	NI4	2.31	1.55	-1.56
Pakistan (PK) n = 7	KPK[b]	PK1	-1.26	-1.77	1.23
	KPK[b]	PK2	-2.89	-0.07	-0.62
	KPK[b]	PK3	-1.70	0.10	0.76
	KPK[b]	PK4	-2.61	2.23	1.29
	Balochistan	PK5	-2.16	-0.56	2.33
	Balochistan	PK6	-1.35	-0.20	0.68
	Sindh	PK7	-2.16	-0.42	-0.54
Tanzania (TZ) n = 4	Morogoro Region	TZ1	1.37	-1.87	1.04
	Arusha region	TZ2	1.83	-0.95	-0.42
	Morogoro Region	TZ3	1.37	-3.29	0.15
	Tanga Region	TZ4	1.31	0.52	-0.75
Uzbekistan (UZ) n = 4	Bukhara	UZ1	-1.41	1.11	1.96
	Samarqand	UZ2	2.36	-0.47	2.60
	Andijon	UZ3	1.58	2.26	2.85
	Kashkadaryo	UZ4	0.86	0.71	1.32
Zimbabwe (ZW) n = 4	Masvingo	ZW1	3.45	0.81	-1.64
	Midlands	ZW2	1.37	0.59	-0.97
	Mashonaland Central	ZW3	1.56	-0.07	-3.20
	Mashonaland Central	ZW4	1.92	0.32	-2.78

*Or equivalent to Province or State.

**Codes are used to facilitate analysis as well as to protect the anonymity of the villages surveyed.

[a] Southern Nations, Nationalities, and Peoples' Region.

[b] Khyber Pakhtunkhwa.

https://doi.org/10.1371/journal.pone.0263771.t003