How do cognitive proximity and knowledge space position affect firms' innovation? Evidence from micro and small manufacturing firms in South Africa

Title	How do cognitive proximity and knowledge space position affect firms' innovation? Evidence from micro and small manufacturing firms in South Africa
Author(s)	Galetti, Jefferson R. B.;Tregenna, Fiona
Publication Date	2025-01-04
Publisher	Wiley
Repository DOI	https://doi.org/10.1111/grow.70021

Growth and Change

growth and change

ORIGINAL ARTICLE OPEN ACCESS

How Do Cognitive Proximity and Knowledge Space Position Affect Firms' Innovation? Evidence From Micro and Small Manufacturing Firms in South Africa

Jefferson R. B. Galetti[1,2] | Fiona Tregenna[2]

[1]J.E. Cairnes School of Business and Economics, University of Galway, Galway, Ireland | [2]DSI/NRF South African Research Chair in Industrial Development, College of Business and Economics, University of Johannesburg, Johannesburg, South Africa

Correspondence: Jefferson R. B. Galetti (jefferson.galetti@universityofgalway.ie)

Received: 12 January 2023 | **Revised:** 26 October 2024 | **Accepted:** 25 December 2024

Funding: This work was supported by the National Research Foundation of South Africa under Grant No. 98627 and 110691.

Keywords: cognitive proximity | innovation | knowledge space | micro and small enterprises (MSEs) | South Africa

ABSTRACT

Cognitive proximity between firms, and their position in relation to other firms in a knowledge space, can influence firms' innovation performance. While some studies have begun analysing this in developing countries, there is a dearth of evidence in Africa, and more generally for small and especially micro enterprises as well as for informal enterprises. To fill this gap, we utilise rich new survey data covering 711 micro and small manufacturing firms in Johannesburg, South Africa. We develop and apply a novel multidimensional measure of cognitive proximity, using information on the skills most needed in firms' activities, firms' proximity in an industrial space, and the types of external co-operation in which they engage. We find that cognitive proximity is positively associated with innovation outcomes. Although this relationship is increasing for most of the sample, there is evidence of an inverted-U relationship for firms located at higher percentiles, suggesting that too much proximity may lead to a cognitive lock-in. Firms' knowledge space position is also associated with differential innovation outcomes. In extensions of the model, we find interesting differences in how both cognitive proximity and knowledge space positions are associated with the degree of innovation novelty and also vary according to whether the firm is formal or informal.

JEL Classification: O33, R11, D85, L26

1 | Introduction

The determinants of firm-level innovation have been extensively investigated (see, e.g. Crepon, Duguet, and Mairessec 1998; Dosi 1988; Rosenberg 1982). There is growing interest in the role of cognitive proximity and knowledge space in influencing firms' innovation activities and performance (Alberti, Belfanti, and Giusti 2021; Escolar et al. 2023; Whittle and Kogler 2020). However, the evidence on this remains limited in the context of developing countries, and more so in the African context; this may be in part because the available data in these contexts typically does not enable the empirical analysis of these topics. This paper analyses the effects of cognitive proximity and firms' position in the knowledge space on the innovation performance of micro and small manufacturing enterprises, including informal enterprises, in Johannesburg, South Africa.

Cognitive proximity has been identified as facilitating interaction between firms and fostering favourable conditions for the exchange of effective information and mutual learning (Boschma 2005; Torre and Rallet 2005; Whittle and Kogler 2020). Cognitive proximity refers to similarities in levels of knowledge and expertise between firms and enables a better understanding of local information, its absorption, and its application to commercial ends (Cohen and Levinthal 1990; Fitjar, Huber, and Rodríguez-Pose 2016). However, if two firms have the same knowledge sets, they would not have novel elements to learn from each other, leading to a potential cognitive lock-in. Therefore, firms that are cognitively close but not approaching identical knowledge sets, will have a mix of knowledge overlap and variety that can enhance their capacities to recombine knowledge that is complementary yet slightly dissimilar (Nooteboom 2000; Nooteboom et al. 2007).

Two firms sharing similar knowledge and expertise will in practice not necessarily engage in interactions that result in knowledge exchange and learning. However, knowledge spillovers will generally be more likely and fruitful between firms that are cognitively close than distant (Frenken, Oort, and Verburg 2007; Grillitsch, Asheim, and Trippl 2018; Nooteboom 2000). Based on relational sets of skills, technology, and exchanged information (Hidalgo et al. 2007; Whittle and Kogler 2020), firms with distinct degrees of cognitive proximity can be represented in an abstract knowledge space where firms with similar knowledge sets are positioned in the densest areas (Hidalgo et al. 2007; Kogler, Rigby, and Tucker 2013).

However, relatively little is known about how cognitive proximity and knowledge space influence innovation in developing countries, in particular where both formal and informal MSEs operate in less diversified productive and technological structures (Eder and Trippl 2019; Imbs and Wacziarg 2003). On the one hand, these firms typically rely heavily on localised networks and informal communication (Acs and Audretsch 1988; Kraemer-Mbula et al. 2019), taking advantage of the tacit knowledge exchanges that take place locally. On the other hand, developing countries' characteristics may result in relatively low levels of local knowledge exchange, which may hinder the innovation process (Grillitsch and Nilsson 2015).

This study investigates whether cognitive proximity and firms' position in the abstract knowledge space (measured as both degree centrality and betweenness centrality) matter for innovation in manufacturing MSEs in Johannesburg, South Africa. The study contributes to the literature in four main ways. The first key contribution is to provide a more detailed understanding of the relationship between cognitively close MSEs and their probabilities of introducing innovation in a developing country context, specifically in an African context. This contrasts with most empirical innovation studies which tend to focus on developed countries (Eder and Trippl 2019; Hervás-Oliver et al. 2021). We use a rich novel dataset focused on innovation and covering 711 manufacturing MSEs in Johannesburg. This helps to address the paucity of literature on this topic in Africa, as well as in developing countries more broadly. The context-specificity of knowledge spaces based on cognitive proximity makes empirical evidence from different contexts especially important for this topic.

Second, to compute the firms' degree of cognitive proximity, we consider three relevant dimensions and develop and apply a novel composite variable using information on (1) skill space based on the proximity between skills that MSEs most need in their activities (Fitjar and Timmermans 2017; Galetti, Tessarin, and Morceiro 2021), (2) firms' proximity in an industry space represented by the sectors in which they operate (Neffke, Henning, and Boschma 2011), and (3) the types of information exchanged in external co-operation they engage in. Our approach departs from previous studies that compute measures of proximity based only on one relational dimension. It also incorporates recent insights from the skill-proximity approach (Content and Frenken 2016; Whittle and Kogler 2020) and takes into account the exchange of different types of knowledge, such as managerial, institutional, industry-specific and market knowledge (Alberti, Belfanti, and Giusti 2021). Together, these three relevant dimensions result in a more nuanced, comprehensive and multidimensional measurement of cognitive proximity between firms.

Third, our study includes specifically micro (i.e. not only small) enterprises, and informal (as well as formal) enterprises. The related empirical literature generally does not cover micro and informal firms, as these categories are excluded from most firm-level surveys. Both these types of firms are especially important in a developing country context (Sookram and Watson 2008), and furthermore are likely to have distinctive characteristics and patterns in terms of cognitive proximity and innovation (Andersson and Lööf 2012; Booyens, Hart, and Ramoroka 2018). Besides an empirical comparison, this analysis also enriches our understanding of the innovation behaviour of micro and informal firms.

The fourth contribution extends the literature on how cognitive proximity and firms' knowledge space position are associated with innovation outcomes, to also consider the relevance of the degree of novelty of innovation. This sheds light on how cognitive proximity is differentially associated with innovations with different degrees of novelty, specifically comparing 'less novel' innovation that is only new to the firm or community with 'more novel' innovation that is new to markets or to the world.

The results suggest that the degree of cognitive proximity between MSEs and their position in an abstract knowledge space increases the likelihood of innovations being introduced, especially for informal MSEs and those related to less novel innovations. Although there is a positive relationship between cognitive proximity and innovation outcomes for most firms, we also find evidence of an inverted-U relationship for firms located at higher percentiles, suggesting that excessive proximity may lead to cognitive lock-in. There are interesting mixed results for the variables associated with knowledge space positions and we observe interesting differences by the type of innovation (overall, product, and process/service), the degree of innovation novelty, and according to whether firms are formal or informal.

The article proceeds as follows. Section 2 provides an overview of the relevant theoretical framework and literature. Section 3 discusses the methodology and data used in the analysis. The results are presented and discussed in Section 4, and Section 5 concludes.

2 | Theoretical Framework and Literature

2.1 | Cognitive Proximity and Innovation

Innovation can be understood as a process in which social actors create, absorb and recombine different knowledge sets to generate new products and processes, which may be proximate to the extant knowledge (Nelson and Winter 1982; Weitzman 1998). As not all types of knowledge may readily be recombined, the concept of proximity is useful for understanding how agents deal with problems of coordination and uncertainty and increase their capacity for learning from others (Boschma 2005; Torre and Rallet 2005). Cognitive proximity refers to the closeness of firms with similar levels of knowledge and expertise, which facilitates information transmission and knowledge exchanges (Boschma 2005; Nooteboom 2000). That proximity can be generated both by the similarity of knowledge within the same industry (Marshallian externalities) and by a related variety of knowledge existing in urban structures (Jacobs' externalities) (Caragliu, de Dominicis, and de Groot 2016; Frenken, Oort, and Verburg 2007; Van der Panne 2004). Existing evidence suggests that cognitive proximity is related to increased levels of innovation (Castaldi, Frenken, and Los 2015; Fitjar, Huber, and Rodríguez-Pose 2016; Nooteboom et al. 2007).

Cognitive proximity between firms is dynamic and is modified through continuous social interactions (whether organised or unplanned), which shorten cognitive distance and increase firms' learning capacity, as well as the stock of knowledge that may potentially be exchanged (Bathelt, Malmberg, and Maskell 2004; Malmberg and Maskell 2006; Nooteboom 2000). It is important to note that, while cognitive proximity facilitates effective communication and knowledge absorption, too much proximity might not improve the innovation capacity of firms, and may even diminish it (Nooteboom 2000). 'Excessive' knowledge similarity between firms may undermine the novelty of the information they exchange and lead to a 'cognitive lock-in', as firms' might not be able to recognise new opportunities to innovate.

In summary, cognitive proximity between firms potentially enhances their likelihood of exchanging knowledge, but the effective introduction of innovation requires a certain level of dissimilarity between the types of knowledge to be recombined. Boschma and Frenken (2010) refer to this as the 'proximity paradox', and its influence on the innovation capacity will depend on the optimal level of cognitive proximity between firms. We thus expect cognitive proximity to increase the probability of fostering learning and knowledge transmission. However, excessive cognitive proximity can reduce this, particularly by reducing the degree of novelty in the information that the firms exchange. This suggests that the relationship between innovation and cognitive proximity follows an inverted U-shaped pattern. Based on the above discussion, we hypothesise that:

Hypothesis 1. *Firms that are close to other firms in terms of cognitive proximity are more likely to introduce innovations, yet too much cognitive proximity is negatively associated with the probability of innovating.*

2.2 | The Knowledge Space Structure and Innovation

An economy can be represented in abstraction as a relational framework of space in which firms that rely on similar knowledge bases are positioned closer than firms with different knowledge sets (Hidalgo et al. 2007; Whittle and Kogler 2020). Although proximity in knowledge space does not mean that firms will necessarily engage in effective relationships to exchange information, their cognitive similarities would improve the likelihood of transmitting both formal and tacit knowledge, enabling mutual learning and promoting innovation (Castaldi, Frenken, and Los 2015; Fitjar, Huber, and Rodríguez-Pose 2016). Frenken, Oort, and Verburg (2007) emphasised that spillovers would be more effective when the knowledge of agents is related. From the perspectives of knowledge space and relatedness, firms that are closely related in cognitive terms are more likely to learn from each other and are centrally positioned in the knowledge space compared to less related firms positioned in the peripheral areas (Neffke, Henning, and Boschma 2011; Xiao, Boschma, and Andersson 2018). As a result of differentiated cognitive distances between firms, potential access to and absorption of relevant knowledge flows is uneven. Indeed, several studies have shown that diversification and growth are more likely to occur in regions with relatively high levels of relatedness between both industries (Galetti, Tessarin, and Morceiro 2021; Neffke, Henning, and Boschma 2011) and occupations (Farinha et al. 2019; Galetti, Tessarin, and Morceiro 2024).

Therefore, based on their cognitive proximity, related firms would be located at the densest areas of the knowledge space where their access to relevant knowledge created by the local 'buzz' would be potentially higher. In contrast, cognitively distant firms would have a lower capacity to access and absorb relevant information and would be located in the peripheral regions of the knowledge space with limited access to knowledge flows (Stuck, Broekel, and Revilla Diez 2016; Whittle and Kogler 2020). As a result, firms occupying central positions in the knowledge space (i.e., cognitively similar firms), potentially have more opportunities to access knowledge to innovate and increase their propensity for introducing innovations (Gilsing et al. 2008; Zaheer and Bell 2005).

Based on these considerations, the 'degree' and 'betweenness' of firms' centrality in a knowledge space are important factors in our framework. Degree centrality can be defined as the number of direct links a firm has with other firms. Since we do not have detailed information in our dataset on actual links between firms, we interpret degree centrality as the number of firms significantly close[1] to a specific firm in cognitive terms. This approach is motivated by Hidalgo et al. (2007), who elaborate a product space with a 'core-periphery' structure. Nodes at the core, that is, highly connected similar products, are more likely to be reached than disconnected nodes composed of less similar products. This relational space framework was extended to represent technologies, industries, occupations and firms (Farinha et al. 2019; Kogler, Rigby, and Tucker 2013; Neffke and Henning 2013). Thus, in our context, the densest parts of an abstract space are populated by firms that are the most similar in cognitive terms. Firms' positionality in the knowledge space varies according to their cognitive distance. The higher the number of cognitively related firms, the higher the potential to access additional sets of knowledge that can be useful for firms' innovation.

Betweenness centrality considers both direct and indirect links. It is based on the 'shortest path' concept, which refers to the minimum number of 'steps' along a network to reach another node. Relevant information tends to flow along the shortest paths within a network, and firms with large betweenness centrality are in a better position to access it (Stuck, Broekel, and Revilla Diez 2016). In our framework, a firm operating in an economy with many closely cognitively related firms would be embedded in the knowledge space's densest areas. That is, it would be located 'between' similar firms, where the access to relevant knowledge flows would potentially be higher. The higher the 'betweenness centrality' of a firm, the greater the benefits from knowledge spillovers and the greater the amount of knowledge it can potentially access (Stuck, Broekel, and Revilla Diez 2016).

This leads to the next hypothesis:

Hypothesis 2. *The greater the degree and betweenness centrality of firms in a knowledge space, the higher their probability of introducing innovations.*

2.3 | Innovation in Developing Countries

Developing countries have socioeconomic characteristics that negatively affect their innovation capacity (Cirera and Muzi 2020; Zanello et al. 2016). Relatively low levels of skills in their workforce, low investments in R&D, and poor educational and technological infrastructure depress opportunities for potential recombination of related knowledge (Eder and Trippl 2019; Grillitsch and Nilsson 2015). Furthermore, amongst firms in developing economies, innovation is likely to be particularly incremental for MSEs, which typically lack the resource base and economies of scale to undertake R&D-based innovations with a high degree of novelty (Coccia 2017; Paus, Robinson, and Tregenna 2022; Zanello et al. 2016).

MSEs are responsible for generating a significant share of jobs in Africa, although they often pay lower wages and experience higher employment turnovers than large firms (Page and Söderbom 2015). Indeed, MSEs contribute over a third of South Africa's GDP and employ 47% of the country's labour force (International Finance Corporation (IFC), 2018; Trade & Industrial Policy Strategies (TIPS), 2023). South African MSEs implement complementary strategies to innovate (Kamutando and Tregenna 2023), and their use of communication technologies has positive effects on innovation (Gaglio, Kraemer-Mbula, and Lorenz 2022). In addition, their workforce has shown the ability to learn on the job and use their own ideas to solve problems (Kraemer-Mbula et al. 2019).

The informal economy accounts for a significant share of the South African economy. It accounts for 18% of total employment in South Africa and 19% in the City of Johannesburg (Stats SA 2020). Its contribution to the country's GDP is estimated at 5.2% (Musara and Nieuwenhuizen 2020). Although the informal sector is essential in providing employment opportunities and income for a great part of the population, its precarious conditions may negatively affect knowledge spillovers and harm the innovation process. Furthermore, informal firms generally do not benefit from government support measures and incentives, or at least not to the same extent as formal firms, including interventions to encourage firm-level innovation.

Considering these contextual characteristics of South Africa as a developing country, we extend our analysis to two additional dimensions. In the first extension, we further analyse how the main variables of interest are associated with the degree of novelty of the innovations. Innovation that is only new to the firm or community (in geographical terms) is less novel, meaning that firms are just implementing or using an existing innovation for the first time. This type of innovation, with a relatively low threshold of novelty, also represents an example of diffusion (OECD 2018). In contrast, an innovation that is new to the market or the world, defined as an innovation that has not been available in the markets in which a firm operates, represents a more novel innovation with a higher threshold of novelty (OECD 2018). Given the typically incremental nature of MSEs' innovation, it is expected that their cognitive proximity and knowledge space position may be more strongly associated with the introduction of less novel innovations than more novel innovations. In the second extension, we assess whether MSEs' formal or informal status may change the relationship between the main variables and innovation. Given the poor conditions for skills development and knowledge creation in informal contexts, it is reasonable to expect a negative effect of informality on the relationship between the variables of interest. Therefore, we formulate the following two additional hypotheses:

Hypothesis 3. *The higher the firms' cognitive proximity and the more central their position in the knowledge space, the higher the probability of introducing less novel innovation relative to more novel innovation.*

Hypothesis 4. *The higher the firms' cognitive proximity and the more central their position in the knowledge space, the higher the probability of introducing innovation in formal firms compared to informal ones.*

3 | Methodology and Data

3.1 | Data

Innovation in MSEs derives in part from the fact that MSEs are well adapted to local 'buzz' and have a more flexible organisational structure to benefit from accessing tacit and sticky knowledge that circulates locally (Booyens 2011; Kamutando and Tregenna 2023; Nooteboom 1994). Indeed, Kraemer-Mbula et al. (2019) have shown that informal communication and knowledge exchange are crucial for developing innovation capacity in African countries.

The City of Johannesburg offers a relevant context for providing evidence on local knowledge flows, interactive learning and innovation: its economic relevance promotes agglomeration economies through localisation externalities that attract workers with higher levels of education, increase local productivity and intensify the flows of goods and knowledge (Krugell and Rankin 2012). The City of Johannesburg is the largest and most economically developed city in the country, accounting for 10% of South Africa's population and 16% of GDP (City of Johannesburg 2021), and thus has significance and relevance as a geographical focus area for this study. At the same time, in common with the rest of South Africa, Johannesburg faces profound social and economic challenges. These include high rates of unemployment and poverty; deficiencies in both 'hard' and 'soft' infrastructure; weak social cohesion; poor spatial integration; weak public transport systems; pronounced unevenness in educational quality and attainment; a high degree of informality in economic activities; weak industrialisation and a lack of structural transformation; and limitations in the provision of basic services, such as water, electricity and sanitation (Andreoni et al. 2021; Andreoni and Tregenna 2021; Tregenna et al. 2022). While there are specificities to the case of South Africa based on the country's apartheid history, many of these challenges are in common with other middle-income countries and developing countries more widely. As discussed, developing country characteristics have relevance for innovation and specifically for the topic of this study. This includes the importance of innovation for growth and development; innovation generally being more incremental, more resource-constrained, and less R&D intensive than in advanced economies; and the relevance of considering innovation in the informal economy.

We use data from a rich new survey focused on innovation, covering manufacturing MSEs in Johannesburg.[2] This survey is among the most comprehensive conducted in African countries to understand the extent and impact of firm-level innovation, specifically among MSEs, with few similar surveys in other developing countries and especially in Africa.

The survey covers MSEs, classified as those with 50 or fewer workers (this threshold is based on official definitions[3] of MSEs). Firms undertaking any manufacturing activities, even if manufacturing was not their primary activity, were included in the survey.

Unlike most other firm-level surveys, especially other innovation surveys, this survey includes micro enterprises, as well as informal enterprises.[4] Given the importance of both types of enterprises in developing countries, especially in Africa, this dataset enables a valuable extension of the empirical literature on cognitive proximity and knowledge space positionality to include all MSEs (including the smallest firms with fewer than five employees) and both formal and informal enterprises.

The survey comprised 74 questions, spanning various aspects of firm characteristics and behaviour, with a focus on innovation (see Table A1 in the Appendix for the questions used in this study). Firms were asked about their demographic and business profiles, innovation processes, training and skills development, linkages, and financial dimensions. The survey questions on innovation followed the Oslo Manual (OECD 2018).

The survey included all enumerator areas (EAs) classified for commercial or industrial use in the urban development zone (UDZ) of the City of Johannesburg. Also included were any additional EAs classified as industrial in the most recent (2011) Census, and contiguous residential EAs where manufacturing activities extended across EA boundaries. The survey thus covered 200 EAs, that is all 142 industrial EAs in the City of Johannesburg and all 58 commercial EAs in the UDZ. The target population for the survey was thus all formal or informal firms with 50 or fewer workers, which undertake any manufacturing activities, located in the designated geographical areas.

In all the EAs, the first step in the data collection was a census of businesses within the designated geographic area, which numbered 1226 firms. Efforts were made to comprehensively count and identify the firms in the census, including through using an existing database and through fieldworkers walking and driving up and down the streets in the designated areas in order to identify firms. However, the census does not necessarily represent the entire universe of qualifying firms, as some firms might have been inadvertently missed or where there was no access to a firm. The census exercise collected data on whether or not manufacturing was the primary activity of the firm, the type of manufacturing involved, the age and size of the firm, and whether or not the business was formally registered. From those firms surveyed in the census, non-manufacturing enterprises and those with more than 50 employees were excluded. Attempts were made to interview all qualifying firms in this census population, without any sampling within it. From June to August 2019 face-to-face interviews were successfully conducted with 724 firms[5] by trained fieldworkers from a professional survey company. After cleaning and processes, the data of 711 of these 724 firms were useable as, for example, not all firms completed the full survey. These 711 firms constitute the final database of the survey, and all were used in the empirical analysis.

Next, Section 3.2 presents the estimation model. Section 3.3 explains the dependent variables measuring innovation, Section 3.4 sets out how we measured cognitive proximity, Section 3.5 explains both measures of firms' knowledge space position (degree centrality and betweenness centrality), and Section 3.6 discusses the control variables.

3.2 | Model

To assess how cognitive proximity between firms and their position in the knowledge space are associated with the probability of introducing various types of innovations, we estimate a linear probability model (LPM) and binomial logistic model, in which innovation is the dependent variable. We followed other studies in the field (Boschma, Balland, and Kogler 2015; Winters 2013) and selected the LPM estimated by OLS as the preferred specification because of its practical advantages.[6]

Thus, the model takes the following form:

$$
\begin{aligned}
P(\text{Innovation} = 1) = \alpha_i + \beta_1 & \cdot \text{Cognitive Proximity}_i + \beta_2 \\
& \cdot \text{Cognitive Proximity}_i{}^2 + \beta_3 \\
& \cdot \text{Degree Centrality}_i + \beta_4 \\
& \cdot \text{Betweenness Centrality}_i + \beta_i \\
& \cdot \text{Controls} + \theta_s + \varepsilon_i,
\end{aligned}
\tag{1}
$$

As errors for firms belonging to the same industry may be correlated, the regression results were adjusted using heteroscedasticity-robust standard errors clustered at the industry level (Cameron, Gelbach, and Miller 2011; Wooldridge 2003).

Based on our first two hypotheses, we expect cognitive proximity, degree centrality and betweenness centrality to show positive and statistically significant coefficients, and that the squared term of cognitive proximity would show a negative and statistically significant coefficient for the likelihood of introducing innovations.

3.3 | Dependent Variables: Innovation

Innovation involves introducing a new or improved product or process that differs significantly from previous products or processes respectively (OECD 2018). Consistent with this, in our survey firms were asked if they had introduced entirely new or significantly improved products, processes and/or services (see Table A1 in the Appendix for further details). The innovation variable is binary and takes a value of one if a firm has introduced any of the three innovation types, and zero otherwise. Our outcome variables thus measure the actual introduction of innovations by the sampled firms (as reported by respondents).

The baseline innovation outcome measure (variable 'Innovation') is the introduction of an entirely or significantly improved innovation, regardless of the type of innovation (product, process or service). In one set of extensions, we separately analyse different types of innovation as dependent variables. First, we distinguish between product innovation, and process and service innovation (with the latter two types of innovation combined for sufficient coverage for the validity of the analysis). Second, we differentiated based on the novelty of innovation: less novel innovations that are only new to the business or the community versus more novel innovations that are new to the market or the world.[7]

3.4 | Measuring Cognitive Proximity

To measure cognitive proximity between MSEs, we introduced a new variable based on three different dimensions. The recombination of dispersed knowledge creates coordination costs that could be alleviated if an agent possesses a set of knowledge that allows it to recognise, assimilate and apply new information (Cohen and Levinthal 1990). This process is facilitated if agents share some degree of proximity in their knowledge and expertise sets, allowing them to communicate and better understand each other (Boschma 2005; Torre and Rallet 2005). Computing the co-occurrence of three dimensions across firms helps to identify similar features shared by local MSEs that facilitate their ability to interact and learn from other firms.

First, we used industrial classification to compute technological proximity, as is common in the relatedness literature (Castaldi, Frenken, and Los 2015; Frenken, Oort, and Verburg 2007). This first dimension used to compute cognitive proximity aims to capture the traditional view that spillovers may arise from Marshallian economies when firms operating in the same industry can benefit from sharing production factors, inputs and knowledge that is locally available (Caragliu, de Dominicis, and de Groot 2016; Van der Panne 2004). For this reason, MSEs operating in the same industry share technological expertise that is associated with common organisational knowledge and similar ways of solving operational problems and day-to-day issues. Technologically related firms are typically better able to benefit from local knowledge spillovers (Frenken, Oort, and Verburg 2007; Neffke, Henning, and Boschma 2011), enhancing their innovation output (Castaldi, Frenken, and Los 2015). All firms in the dataset were classified into 22 manufacturing sectors,[8] with their sectoral identification entering into the computation of the technological proximity of the firms.

Second, we considered the types of skills most needed in the business, in line with the most recent skill-relatedness approach (Content and Frenken 2016; Whittle and Kogler 2020). According to Jacobs' externalities, firms may benefit from the knowledge variety in diversified urban centres (Caragliu, de Dominicis, and de Groot 2016; Van der Panne 2004). With this knowledge variety, workers specialise in related skills that complement each other when utilised together (Neffke 2019; Neffke and Henning 2013). Thus, MSEs are better placed to learn from others when employing workers with related skills. Some studies show that firms' proximity in the skills space has influenced regional diversification and development (Fitjar and Timmermans 2017; Galetti, Tessarin, and Morceiro 2021; Neffke and Henning 2013). Firms were asked to indicate the three most needed skills in their business from a list of 11 different skills. Based on the responses, we were able to 'link' each firm to others in terms of skills similarity.

Third, we included the extent of co-operation with other MSEs in the same industry based on different types of knowledge. This knowledge heterogeneity is essential to the recombination process that enables the creation of new products, processes, and services (Martynovich and Taalbi 2022; Weitzman 1998). Besides industry- and skill-specific knowledge already captured by the previous dimensions, this dimension aims to capture the firm's expertise that may arise from information diversity. This is in terms of market knowledge (information on products and processes, and how to protect ideas), managerial knowledge, labour market knowledge (collective negotiations) and external support knowledge (government support) (Alberti, Belfanti, and Giusti 2021; Brenner 2007; Sammarra and Biggiero 2008). MSEs are closer in a cognitive dimension if they engage in external co-operation that involves exchanging similar types of knowledge. Interactions that are aimed to share information of general interest tend to strengthen cognitive proximity and enhance the learning processes between MSEs (Bathelt, Malmberg, and Maskell 2004; Malmberg and Maskell 2006). Based on the survey responses, we were able to classify MSEs as similar where they exchange similar types of knowledge.

Therefore, relying on these three dimensions, we aimed to capture the elements related to knowledge generated from specialisation (or Marshallian externalities), such as industry-specific knowledge, as well as knowledge variety in urban areas (from Jacobs' externalities), such as the heterogeneous types of knowledge and skills locally available.

Having identified the three dimensions to be included in the new variable measuring cognitive proximity, the next step was to arrange this information in a matrix with 43 knowledge entries × 711 firms. Each cell of the matrix takes a value of one if a firm in the row indicated the variable in the column as one of the most needed skills, the sector of activity or type of knowledge exchanged, and zero otherwise. For each firm, a cell with the value one in a specific variable represents an occurrence, and if two or more firms have a value of 1 for the same variable, we can compute the co-occurrence, or in other words, how many times that variable co-occurred among the firms in the sample. To obtain the cognitive proximity φ between firms, we followed Steijn (2021) to compute the minimum of the pairwise conditional probability of firms i and j having a similar set of knowledge based on skills, technological proximity and the type of co-operation. We thus calculated the number of normalised co-occurrences between firms from the matrix and compared them to the random probability of co-occurrence, according to:

$$\varphi_{i,j} = \frac{C_{ij}}{m\left(\frac{S_i}{T}\frac{S_j}{T-S_i} + \frac{S_j}{T}\frac{S_i}{T-S_j}\right)} \tag{2}$$

where C_{ij} is the number of co-occurrences of firms i and j, S_i and S_j are the number of occurrences of firms i and j, respectively. T is the total of all firms' knowledge specialisations, and m is the total number of co-occurrences. The cognitive proximity $\varphi_{i,j}$ indicates the probability of two firms having a similar set of skills, technological proximity and types of co-operation. Two firms are closer or related in the cognitive dimension if $\varphi > 1$, and are distant or unrelated if $\varphi < 1$. We analysed pairs of cognitively close MSEs ($\varphi > 1$) in order to focus on the relevant areas of the relational space where the likelihood of knowledge spillovers is greater. Thereafter, we averaged φ values across MSEs and rescaled cognitive proximity between 1 and 100. We also added a quadratic term of this variable to capture potential non-linearity between the cognitive proximity of MSEs and innovation outcomes.

3.5 | Measuring the Knowledge Space Position of Firms

To identify a firm's position in the abstract knowledge space, we computed two variables: degree centrality and betweenness centrality. These are each included in our regressions as possible determinants of innovation outcomes.

In our framework, since cognitive distance between firms varies with their similarities in terms of skills, technological expertise, and informational co-operation, we conceptualise degree centrality (Cd_i) as the number of firms significantly close to a specific firm in cognitive terms, that is with $\varphi_{i,j} > 1$. Thus, degree centrality is given by the sum of cognitively close firms, according to:

$$Cd_i = \sum_{i \neq j} \varphi_{ij} \tag{3}$$

Betweenness centrality (Cb_i) is the second variable measuring firms' positionality in the abstract knowledge space. This refers to the extent to which a firm is located between other cognitively close firms, that is, with $\varphi_{i,j} > 1$ in the knowledge space. These firms would be embedded in the densest areas of the knowledge space, with paths between cognitively related firms in which knowledge flows and can potentially be accessed. Betweenness centrality is calculated as the fraction of the shortest paths between other firms that pass through the firm i, where i is neither an origin nor a destination (Stuck, Broekel, and Revilla Diez 2016), according to:

$$Cb_i = \sum_{s \neq t \neq v \in V} \frac{\sigma(s,t|i)}{\sigma(s,t)}, \tag{4}$$

where $\sigma(s,t|i)$ is the total number of shortest paths between s and t passing through i, and $\sigma(s,t)$ is the total number of shortest paths between s and t, independent of whether or not they pass through i (Coscia 2021). Both variables were rescaled to between 1 and 100.

3.6 | Control Variables

We include a range of control variables that may also influence firms' innovation outcomes. The names, types and definitions of all variables are listed in Table A1 in the Appendix. The first set of controls is related to the MSEs' internal resources and absorption capacity (Cohen and Levinthal 1990). We included the educational level and years of experience of managers, the age of the business, a binary variable related to R&D, and the firm size. We also controlled for several characteristics of the firm's labour force, such as the availability of the skills most needed, a variable indicating the extent to which skills are needed from workers that change over time, the capacity of MSEs to retain workers, and the types of workers' training. Taking account of the developing country context, we also controlled for the formal status of a firm with two alternative measures capturing different aspects of in/formality: first, whether an MSE is registered with the country's tax authority (South African Revenue Service); and second, whether or not all employees have formal contracts with the firm.[9]

The second set of variables added as controls is related to the establishment of external connections with different actors located within and outside the region. These connections may compensate for firms' lack of internal expertise and limited resources (Jong and Vermeulen 2006; Nooteboom 1994). In addition, as there is no information from the dataset on cognitive proximity between MSEs and other types of firms, the inclusion of external connections helps to alleviate this gap. We controlled for the type of customers from which a firm gets most of its sales, and for customer location in terms of whether most firms' customers are located in South Africa (baseline) or abroad. We also included measures of the number of suppliers and control for firms' participation in business associations and inter-firm co-operation. In addition, we included dummy variables for the 22 manufacturing sectors, to control for heterogeneity across industries.

4 | Results

4.1 | Descriptive Statistics

Descriptive statistics for all firms are shown in Table 1. Of all firms in the sample, 53.7% introduced an innovation as per the aforementioned definition. When considering only product innovation, 48.8% of the MSEs are innovators, while the innovation rate for processes and services combined was 14.6%. Almost a fifth of the 711 MSEs introduced 'more novel' innovations that are new to the market or new to the world, while 34.6% of the MSEs introduced 'less novel' innovations that are only new to the firm or the community.

South Africa has a significant informal economy: only 27.1% of MSEs in the sample are registered with the South African Revenue Service and only a third of the MSEs have formal contracts with all their employees. There are significant differences in the innovation behaviour of the formal and informal firms in our sample (see the comparative descriptive statistics in

TABLE 1 | Descriptive statistics.

	All firms			Yes		No	
	N	Min	Max	*N*	%	*N*	%
Innovation	711	0	1	382	53.73	329	46.27
Product innovation	711	0	1	347	48.80	364	51.20
Process and service innovation	711	0	1	104	14.63	607	85.37
Innovation—less novel	711	0	1	246	34.60	465	65.40
Innovation—more novel	711	0	1	136	19.13	575	80.87
New to the business	711	0	1	154	21.66	557	78.34
New to the community	711	0	1	92	12.94	619	87.06
New to the market	711	0	1	111	15.61	600	84.39
New to the world	711	0	1	25	3.52	686	96.48

	All firms					Innovative firms		Non-innovative firms	
	N	Min	Max	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Cognitive proximity	711	1	100	21.66	12.18	21.77	11.56	21.52	12.88
Cognitive proximity^2	711	1	10,000	617.2	803.78	607.38	699.36	628.54	911.11
Degree centrality	711	1	100	43.05	9.7	43.12	9.65	42.97	9.77
Betweenness centrality	711	1	100	5.08	5.88	5.64	6.95	4.42	4.22
Education	711	1	10	5.78	2.01	6.34	1.9	5.13	1.95
R&D	711	0	1	0.15	0.35	0.21	0.41	0.07	0.26
Size (log)	711	0	3.91	1.28	1.02	1.45	1.03	1.08	0.97
Skills—availability	711	1	4	2.04	1.14	1.9	1.07	2.2	1.2
Skills—retaining	711	1	4	2.37	1.11	2.34	1.1	2.41	1.12
Skills—training	711	0	1	0.79	0.41	0.81	0.39	0.75	0.43
Skills—stability	711	1	4	2.21	1.09	2.09	1.03	2.34	1.14
Experience	711	1	53	13.27	10.56	13	10.67	13.6	10.44
Age of business	711	2	119	14.85	13.99	14.16	12.84	15.64	15.2
Customer—individuals	711	0	1	0.73	0.44	0.71	0.46	0.77	0.42
Customer—business	711	0	1	0.26	0.44	0.28	0.45	0.23	0.42
Customer—government	711	0	1	0.01	0.08	0.02	0.12	0.00	0.00
Customer—local	711	0	1	0.95	0.21	0.93	0.25	0.97	0.16
Customer—abroad	711	0	1	0.05	0.18	0.07	0.22	0.03	0.12
# Suppliers	711	1	6	2.07	0.83	2.2	0.83	1.92	0.8
Member—Bus. assoc.	711	0	1	0.09	0.29	0.1	0.3	0.09	0.28
Co-operation	711	0	1	0.48	0.5	0.55	0.5	0.39	0.49
Registration	711	0	1	0.27	0.44	0.32	0.47	0.21	0.41
Formal workers	711	0	1	0.33	0.47	0.45	0.50	0.20	0.40

Table A2 in the Appendix). Formal firms are significantly more likely to undertake innovation, as well as to undertake 'more novel' innovation. These differences are apparent for both of our measures of firms' formality. Formal enterprises also have a slightly higher level of cognitive proximity than informal enterprises, although their figures for the position in the knowledge space are very close to each other.

The relationships between the variables of interest are analysed in the econometric results that follow. Our main results are discussed in Section 4.2, while Section 4.3 extends this analysis with the dimension of innovation novelty, and Section 4.4 presents the results for formal and informal MSEs.

4.2 | Main Results

Table 2 reports the econometric results for overall innovation (Column 1), product innovation (Column 2), and process and service innovation (referred to as non-product innovation,

TABLE 2 | Linear probability model—overall innovation, product innovation, and non-product innovation.

	Overall innovation (1)	Product innovation (2)	Non-product innovation (3)	Innovation - CEM (4)	Product innovation - CEM (5)	Non-product innovation - CEM (6)
Cognitive proximity	0.011*** (0.002)	0.010*** (0.002)	0.004** (0.002)	0.032*** (0.010)	0.031*** (0.011)	0.037** (0.018)
Cognitive proximity^2	−0.0001*** (0.00004)	−0.0001*** (0.00004)	−0.0001*** (0.00002)	−0.001*** (0.0002)	−0.0005** (0.0002)	−0.001*** (0.0003)
Degree centrality	0.001 (0.001)	0.003 (0.002)	−0.004*** (0.002)	−0.001 (0.004)	−0.001 (0.004)	−0.017** (0.007)
Betweenness centrality	0.008** (0.003)	0.008*** (0.003)	−0.0003 (0.001)	0.017*** (0.006)	0.018*** (0.006)	0.013 (0.015)
Education	0.054*** (0.010)	0.048*** (0.011)	0.022*** (0.005)	0.014 (0.019)	0.013 (0.019)	−0.008 (0.039)
R&D	0.120** (0.054)	0.112 (0.070)	0.115** (0.055)	0.162* (0.086)	0.104 (0.094)	0.144 (0.134)
Size (log)	0.013 (0.013)	−0.019 (0.013)	0.048*** (0.010)	0.022 (0.042)	0.011 (0.042)	0.086 (0.084)
Skills—availability	−0.068*** (0.017)	−0.054*** (0.021)	−0.029** (0.014)	−0.037 (0.028)	−0.030 (0.028)	0.078 (0.058)
Skills—retaining	0.022 (0.022)	0.032* (0.017)	0.003 (0.014)	0.037 (0.031)	0.038 (0.030)	−0.068 (0.046)
Skills—training	0.044 (0.033)	0.049 (0.043)	0.011 (0.017)	−0.026 (0.069)	0.014 (0.067)	−0.120 (0.137)
Skills—stability	−0.035** (0.017)	−0.039** (0.017)	−0.0002 (0.008)	−0.023 (0.029)	−0.019 (0.029)	−0.046 (0.053)
Customer—business	0.001 (0.042)	−0.016 (0.046)	0.013 (0.037)	0.063 (0.075)	0.029 (0.074)	0.232* (0.133)
Customer—government	0.322*** (0.096)	0.081 (0.223)	0.244 (0.195)	−0.050 (0.163)	−0.243 (0.220)	−0.363 (0.293)
Customer—abroad	0.216*** (0.059)	0.228*** (0.049)	−0.136*** (0.043)	0.303** (0.122)	0.265** (0.130)	−0.506** (0.248)
# Suppliers	0.030 (0.020)	0.034 (0.024)	0.004 (0.020)	0.032 (0.048)	0.036 (0.053)	−0.078 (0.083)
Member—Bus. assoc.	−0.090** (0.045)	−0.084* (0.045)	−0.040 (0.031)	−0.309*** (0.119)	−0.322*** (0.108)	−0.057 (0.157)
Co-operation	0.085** (0.033)	0.075** (0.038)	−0.007 (0.031)	0.078 (0.068)	0.039 (0.069)	−0.017 (0.118)
Experience	0.0004 (0.001)	−0.001 (0.002)	0.003* (0.002)	0.0001 (0.003)	−0.001 (0.003)	0.008 (0.006)
Age of business	−0.002 (0.001)	−0.001 (0.001)	−0.001 (0.001)	0.003 (0.004)	0.003 (0.004)	−0.008 (0.010)
Registration	0.016 (0.029)	0.020 (0.037)	−0.010 (0.049)	0.027 (0.071)	0.022 (0.070)	0.095 (0.115)
Formal workers	0.174*** (0.032)	0.175*** (0.033)	0.122*** (0.026)	0.205*** (0.067)	0.206*** (0.065)	0.361*** (0.110)
Constant	−0.106 (0.116)	−0.158 (0.141)	0.076 (0.142)	−0.512 (0.349)	−0.477 (0.346)	0.395 (0.615)
Observations	711	711	711	348	354	118
R^2	0.238	0.195	0.184	0.191	0.177	0.390
Adjusted R^2	0.190	0.144	0.133	0.094	0.083	0.119
Industry dummies	YES	YES	YES	YES	YES	YES
F statistic	4.961***	3.847***	3.598***	1.975***	1.890***	1.437*

Note: Heteroskedasticity-robust standard errors are shown in parentheses.
Coefficients are statistically significant at *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Column 3). We focus on the LPM results in our discussion for the reasons set out in Section 3.2 and since both methods (LPM and Logit) produce similar results; the Logit model results are shown in Tables A4 and A5 in the Appendix. In each case, we are interested in the association between cognitive proximity and firms' knowledge space positions (degree centrality and betweenness centrality), and innovation outcomes.

A key finding in all estimations is that the higher the cognitive proximity of an MSE with other firms, the higher the probability of introducing innovation. Thus, firms with similar sets of workers' skills, technological expertise and related external co-operation seem better able to absorb knowledge to apply in their innovation processes. The range between MSEs in the 25th and 75th percentile of cognitive proximity is 13.2 points. Focusing on the overall innovation, an MSE located at the 75th percentile in terms of cognitive proximity has a probability of introducing innovation that is 14.6 percentage points higher than a 25th percentile MSE. While the differential probability of introducing product innovation for MSEs in the same percentile range is quite similar (13 percentage points), the differential probability for introducing non-product innovations is around 5.3 percentage points.

We also find evidence for the proximity paradox: coefficients of the quadratic terms of cognitive proximity have the hypothesised negative signs and are significant for all types of innovation. However, the peak[10] of the effects on innovation occurs at a relatively high level of cognitive proximity at 38.5 for overall innovation (around the 90th percentile), 45.2 (95th percentile) for product innovation, and 22.6 (60th percentile) for non-product innovation. Since novel combinations depend on a certain degree of dissimilarity to be recombined effectively, the positive association between cognitive proximity and innovation are increasing for most of the MSEs in our sample. Nevertheless, the relationship between innovation and cognitive proximity follows an inverted U-shaped curve. Together, these results support Hypothesis 1.

Regarding the association between innovation and MSEs' position in the knowledge space, there is an interesting mix of results. Degree centrality—the number of firms significantly related to a specific firm in cognitive terms—is positive but not statistically significant for overall and product innovation in Columns 1 and 2 respectively. It also presented an unexpected negative signal for process and service innovation. A possible explanation for the latter result could be that, in our sample of firms, those with similar knowledge sets to many others become more accustomed to doing things in the same way, and hence, less innovative in their ways of doing things in terms of production processes and services; this could be explored further in future research.

When looking at the measure of betweenness centrality, we found more consistent results for overall innovation and for product innovation. Both estimations showed a positive and statistically significant coefficient of similar magnitude on betweenness centrality, indicating that being more centrally positioned in the middle of the knowledge space where knowledge is likely to circulate intensively, increases the likelihood of introducing innovations. MSEs at the 75th percentile of the distribution have a slightly higher probability (3.75 percentage points) of innovating than MSEs at the 25th percentile. However, we did not find any association between betweenness centrality on process and service innovation. We thus found partial evidence of our Hypothesis 2, where only betweenness centrality seems to play a significant role in product innovation and overall innovation.

Despite having a firm-specific dependent variable while the explanatory variables of interest are constructed based on the similarities between multiple firms, we nevertheless cannot rule out the possibility that the main explanatory variables could be correlated with the error term. As we lack an ideal instrument to identify a causal relationship between both cognitive proximity and knowledge space positions and innovation, we adopted some additional measures to alleviate possible endogeneity concerns to the extent possible.

It is worth noting that all econometric specifications include industry fixed effects that are directly estimated by including dummy variables. This accounts for the effects of unobservable characteristics related to a specific industry that could influence innovation outcomes.

As our key strategy to address potential endogeneity, we re-estimate our core regressions using a different sample obtained after applying a k-to-k coarse exact matching (CEM) approach to identify appropriate matches for innovative and non-innovative firms. Our identification strategy is to match innovative MSEs with the control MSEs that have similar characteristics. We use relatively coarse bins for the educational level of managers, years of working experience that managers have in the sector, the age of the business, number of suppliers, size of the firm, and some characteristics of the workers' skills. Table A3 in the Appendix provides summary statistics and technical details for the matching process. These results are presented in Table 2, Columns 4 to 6. The coefficients of interest are consistent in sign and significance with our main results (Columns 1–3), with a larger magnitude. For example, an MSE at the 75th percentile has a probability of introducing an overall innovation that is 40 percentage points higher than an MSE at the 25th percentile.

While the results are reassuring and help alleviate to some extent the concerns about endogeneity, as well as confirm the robustness of our results, we do not have all the elements to interpret our results as causal. The cross-sectional nature of our dataset also limits the ability to draw strong conclusions about causality. However, our estimates are relevant to improve our understanding of the relationship between innovation and cognitive proximity and knowledge space positions of manufacturing firms.

4.3 | Extension: Novelty of Innovation

We then extend the analysis to investigate if the relationship between cognitive proximity and the knowledge space position of MSEs and their innovation outcomes varies for different degrees of innovation novelty. This extension is motivated in part by a recognition of the different roles typically played by small and large firms in innovation systems, with our focus being on MSEs. Larger firms generally tend to play a major role in scientific and technological breakthroughs by employing specialised teams focused on science and R&D-based innovations. In contrast, both smaller and informal firms typically follow more incremental improvements in existing technologies, their implementation, application, differentiation and adaptation, with lower innovation novelty (Nooteboom 1994). Furthermore, consideration of the degree of novelty brings nuances and deeper insights to our understanding of the relationship between firms' cognitive proximity and abstract knowledge space positionality, and their innovation performance. Bringing in the novelty of innovation is particularly relevant in the developing country context where, as discussed earlier, innovation is predominantly incremental with limited novelty.

'Less novel' innovations refer to the lower threshold for innovation novelty (OECD 2018), referring to innovations that are only new to either the business or the community. 'More novel' innovations refer to those that are new to either the market (i.e. implemented for the first time in the market or sector) or to the world (i.e. the introduction of innovation for all markets and industries, domestic and international) (OECD 2018). Our survey data indicates that 40% of the innovations introduced by our sample firms are new to the business and 24% are new to the community (classified collectively here as 'less novel' innovation). Meanwhile, 29% are new to the market and only 7% are new to the world, which are classified collectively here as 'more novel' innovation. Taking into account the nature of firms in the sample, we expect that cognitive proximity between MSEs and their embeddedness in the abstract knowledge space might be associated more strongly with less novel innovations.

Table 3 provides the estimates for the extended model, with an increasing order of novelty in Columns 1 and 2. The inverted-U relationship between cognitive proximity and innovation (as per Hypothesis 1) is observed only for less novel innovation. Firms at the 75th percentile of the cognitive proximity distribution experience a probability of innovating that is 22.9 percentage points higher than counterparts at the 25th percentile. In a similar pattern as with overall innovation, the likelihood of introducing innovations with lower novelty is increasing with cognitive proximity for most of the MSEs in our sample. Only for MSEs located around the 90th percentile onwards, does the association between those variables start falling.

Interestingly, when we consider the higher level of innovation novelty ('more novel' innovation), we did not find a role for cognitive proximity in the innovation process of MSEs. A possible explanation for these results might be that the introduction of innovations with greater degrees of novelty relies more on the combination of unrelated knowledge sets than related ones that circulate locally (Castaldi, Frenken, and Los 2015), including possibly through cognitive proximity with larger firms and firms outside of the local area. That is, highly novel innovations require more than cognitive proximity with other MSEs in the surrounding area; they typically involve more intensive innovation efforts including through in-house R&D, as well as inputs of knowledge and ideas from beyond the local area, for example through international investors or knowledge exchanges. In contrast, cognitive proximity is conducive to exchanging simpler types of knowledge and ideas, which contribute to less novel innovations.

TABLE 3 | Linear probability model—degree of the innovation novelty.

	Innovation—Less novel (1)	Innovation—More novel (2)
Cognitive proximity	0.017*** (0.004)	−0.006 (0.004)
Cognitive proximity^2	−0.0002*** (0.0001)	0.0001 (0.0001)
Degree centrality	0.002 (0.002)	−0.001 (0.001)
Betweenness centrality	0.007** (0.003)	0.001 (0.001)
Education	0.034*** (0.009)	0.020* (0.011)
R&D	0.059 (0.055)	0.061** (0.026)
Size (log)	−0.004 (0.015)	0.017 (0.011)
Skills—availability	−0.048*** (0.010)	−0.020 (0.019)
Skills—retaining	0.011 (0.019)	0.011 (0.009)
Skills - training	0.076** (0.034)	−0.033 (0.032)
Skills—stability	−0.029*** (0.011)	−0.006 (0.015)
Customer—business	−0.0002 (0.045)	0.001 (0.024)
Customer—government	−0.063 (0.219)	0.385 (0.263)
Customer—abroad	0.019 (0.136)	0.197* (0.113)
# Suppliers	0.043** (0.021)	−0.013 (0.013)
Member - bus. assoc.	−0.060 (0.055)	−0.030 (0.068)
Co-operation	0.089*** (0.033)	−0.004 (0.030)
Experience	−0.001 (0.002)	0.002 (0.001)
Age of business	0.001 (0.001)	−0.002** (0.001)
Registration	−0.048 (0.049)	0.064* (0.033)
Formal workers	0.079* (0.041)	0.094*** (0.034)
Constant	−0.095 (0.146)	−0.010 (0.103)
Observations	711	711
R^2	0.134	0.120
Adjusted R^2	0.079	0.064
Industry dummies	YES	YES
F statistic	2.452***	2.159***

Note: Heteroskedasticity-robust standard errors are shown in parentheses. Coefficients are statistically significant at $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

In terms of the knowledge space position of MSEs, we find a positive and significant relationship between betweenness centrality and the introduction of less novel innovations. The differential in the probability of introducing innovations between

75th percentile firms and 25th percentile firms is about nine percentage points. The coefficients of betweenness centrality for more novel innovations, and degree centrality for both less and more novel innovations, are not statistically different from zero.

The results suggest that firms' cognitive proximity and betweenness position in the knowledge space may facilitate access, understanding, absorption, and application of local knowledge to implement less novel innovation, according to Hypothesis 3.

On the one hand, our results highlight how MSEs are important in the context of developing economies. MSEs are able to meaningfully observe new products in the market, learn how to adapt them to their necessities and incorporate them into their portfolios. On the other hand, MSEs in developing countries face challenging conditions that impose enormous disadvantages, such as low levels of clusters and agglomeration externalities and weak organisational and institutional support structures (Isaksen and Trippl 2016). As relatively little knowledge exchange takes place in peripheral conditions (Eder and Trippl 2019), cognitive similarities between firms are not as important for more novel innovation (relative to less novel innovation). Thus firms need to rely on internal capacities to compensate for weak local knowledge spillovers (Eder and Trippl 2019; Grillitsch and Nilsson 2015).

4.4 | Extension: Formality of the Firms

As discussed, a high degree of informality is observed among our sample firms; this is of particular interest as a distinguishing characteristic of developing economies more broadly. From our descriptive statistics (Table 1), innovative MSEs were observed to be more formal than non-innovative MSEs, for both measures of firm formality. Our two alternative measures of firms' formality were already included as control variables in our regressions discussed above. From our main results, Table 2 showed that only workforce formality (the variable 'formal workers') is positively and significantly associated with the probability of introducing all types of innovations across all econometric specifications. The formal registration of a firm for tax (the variable 'Registration') is not found to be statistically significant (Table 2).

We now extend our analysis concerning firms' formality. A distinction is made between firms that are classified as formal based on their tax registration status (variable 'Registration'), and those whose workforce is formal by having formal contracts with the firm (variable 'Formal workers'). We interact each of these two measures of formality with each of our four explanatory variables of interest (cognitive proximity and its square, degree centrality, and betweenness centrality). The extension focused on firm formality has an interesting mix of results (see Table 4), disaggregated for the three types of innovation, namely overall, product, and non-product (i.e., combined process and service innovation).

The coefficients of the four explanatory variables of interest remain broadly as expected and consistent with the main results for both formally registered firms and those with a formal workforce (i.e. both measures of formality), and for both overall innovation and product innovation. In addition, for regressions with formally registered firms, the positive coefficient of degree centrality is statistically significant for overall and product innovation (Columns 1 and 3 of Table 4). However, the coefficient of cognitive proximity is no longer statistically significant for non-product innovation.

The results on the interaction terms between the two dimensions of formality and the four explanatory variables of interest are unexpected and do not support Hypothesis 4. In terms of cognitive proximity, we do not find any difference between MSEs with a formally employed workforce and those with informal workers. This variable remains positive and statistically associated with overall and product innovation for all MSEs, regardless of the formality of the workforce (Columns 2 and 4 of Table 4). However, MSEs with informal workers seem to benefit more from the betweenness centrality than those that have all formal employees (as evidenced by the negative and significant coefficients on the interaction term betweenness centrality and formal workers in Columns 2 and 4 of Table 4). In contrast, coming to the formality of MSEs, the key finding is that formal registration of firms negates the positive association between cognitive proximity and firms' degree centrality in the knowledge space and their overall and product innovation performance. This is evident from Columns 1 and 3 of Table 4, showing the magnitude of the interaction terms to exceed in magnitude and with the opposite sign to our main explanatory variables.

Together, these results indicate that, although relevant for all MSEs, at least the access to local knowledge provided by cognitive proximity is specific to the informal MSEs not officially registered. One possible interpretation for these results is that the types of knowledge that could be exchanged between MSEs may generally not be complex and not require high absorptive capacity, but rather relate primarily to incremental adaptations, thus facilitating the learning processes of informal MSEs and enabling them to adopt mature technologies. Thus, while innovation is lower among informal firms relative to formal firms, the positive role of cognitive proximity and knowledge space position on innovation is stronger for informal firms.

5 | Conclusion

This study investigates whether cognitive proximity and knowledge space position are associated with MSEs' innovation patterns and performance. We used new data from a survey of 711 manufacturing MSEs located in Johannesburg, South Africa. Drawing on theoretical insights from the proximity literature (Boschma 2005; Boschma and Frenken 2010; Torre and Rallet 2005), we introduced a multidimensional measure of cognitive proximity between firms and applied it to map an abstract knowledge space in which we could position each firm from our database. We tested the association of cognitive proximity and knowledge space embeddedness with firms' introduction of innovations.

One way in which this article contributes to the literature is the development and application of a new three-dimensional measure of cognitive proximity. Based on co-occurrences

TABLE 4 | Linear Probability Model—overall innovation, product innovation, and non-product innovation for formal and informal MSEs.

	Overall innovation		**Product innovation**		**Non-product innovation**	
	Registration (1)	**Formal workers (2)**	**Registration (3)**	**Formal workers (4)**	**Registration (5)**	**Formal workers (6)**
Cognitive proximity	0.017*** (0.005)	0.015*** (0.005)	0.015*** (0.005)	0.015*** (0.005)	0.003 (0.002)	0.001 (0.004)
Cognitive proximity^2	−0.0002*** (0.0001)	−0.0002* (0.0001)	−0.0001* (0.0001)	−0.0001 (0.0001)	−0.0001** (0.00004)	−0.00004 (0.00004)
Degree centrality	0.004** (0.002)	0.0004 (0.003)	0.007*** (0.002)	0.002 (0.003)	−0.003** (0.001)	−0.005** (0.002)
Betweenness centrality	0.009*** (0.004)	0.016*** (0.005)	0.009*** (0.003)	0.017*** (0.005)	0.0004 (0.002)	0.002 (0.003)
Registration	0.772*** (0.182)	0.019 (0.027)	0.837*** (0.244)	0.024 (0.035)	0.168 (0.226)	−0.010 (0.047)
Formal workers	0.179*** (0.031)	0.301 (0.200)	0.180*** (0.030)	0.365 (0.228)	0.120*** (0.026)	−0.084 (0.241)
Registration × cognitive proximity	−0.020** (0.008)		−0.019** (0.009)		0.002 (0.008)	
Registration × cognitive proximity^2	0.0002 (0.0001)		0.0001 (0.0001)		−0.00001 (0.0001)	
Registration × degree centrality	−0.009*** (0.002)		−0.011*** (0.003)		−0.004 (0.003)	
Registration × betweenness centrality	−0.007 (0.008)		−0.0001 (0.007)		−0.006 (0.005)	
Formal workers × cognitive proximity		−0.009 (0.007)		−0.012 (0.008)		0.007 (0.009)
Formal workers × cognitive proximity^2		0.00004 (0.0001)		0.0001 (0.0001)		−0.0001 (0.0001)
Formal workers × degree centrality		0.002 (0.004)		0.002 (0.003)		0.003 (0.003)
Formal workers × betweenness centrality		−0.013*** (0.005)		−0.015** (0.006)		−0.003 (0.004)
Constant	−0.292** (0.128)	−0.190 (0.156)	−0.367** (0.144)	−0.269 (0.202)	0.037 (0.095)	0.131 (0.197)
Observations	711	711	711	711	711	711
R^2	0.250	0.245	0.211	0.204	0.189	0.189
Adjusted R^2	0.198	0.193	0.156	0.149	0.133	0.133
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes
F statistic	4.818***	4.690***	3.863***	3.709***	3.358***	3.368***

Note: Heteroskedasticity-robust standard errors are shown in parentheses. All regressions control for all variables listed in Table 1.
Coefficients are statistically significant at $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

analysis, we computed similarities that facilitate firms' ability to potentially learn from each other by using information on different skills that MSEs need most in their activities, their technological proximity represented by the sectors in which they operate, and the types of external co-operation they engage in.

Our application of this multi-dimensional measure of cognitive proximity further contributes to the literature by providing empirical evidence on how cognitive proximity and localised knowledge space are associated with innovation outcomes in a developing country, where existing empirical evidence is sparse, especially in the African context. Our focus on MSEs is especially germane given the lack of evidence for these firms, in particular micro firms and informal firms that are typically excluded from firm-level surveys and hence from the related literature, yet that are very important in developing countries.

In addition to testing how cognitive proximity and knowledge space centrality are associated with overall innovation, product innovation and process innovation, we also add to the literature by considering differential outcomes in terms of the novelty of innovation and the formal status of the firms.

Our findings suggest that the relationship between cognitive proximity and innovation is positive and increasing for most MSEs. However, we also found evidence of an inverted U-shaped curve associated with the 'proximity paradox' for MSEs with high levels of cognitive proximity. In summary, proximity potentially enables efficient knowledge exchange that enhances the probability of innovating, but too much proximity undermines the novelty of the knowledge exchanged. This may lead to a cognitive lock-in at very high levels of proximity, thus potentially diminishing the MSEs' innovation capacity and performance. We found some evidence that more central MSEs located 'between' other cognitively close firms in the knowledge space are better positioned to potentially access new knowledge sets and increase their likelihood of introducing innovation.

Our extended results reveal interesting differences by the novelty of innovation and by firms' formal status. The inverted-U relationship found for cognitive proximity in our main results only holds for less novel innovations (that are only new to the firm or community). It appears that more novel innovations, which are new to the market or world, depend less on locally circulating knowledge sets. In terms of firms' knowledge space position, we find betweenness centrality to be important only to less novel innovations.

In addition, we find that cognitive proximity and knowledge space centrality are relevant for informal (non-registered) MSEs. These firms tend to rely more on potential localised knowledge flows, and their cognitive similarities help them to access, absorb and recombine related knowledge in the form of innovation diffusion. This is associated with less novel innovation that is only new to the firm or community. While rates of innovation are lower for informal firms relative to formal firms, the role of cognitive proximity and knowledge space position on innovation is stronger for informal firms. This is likely related to informal firms undertaking primarily incremental innovations that depend more on locally circulating knowledge, as well as informal firms having weaker innovation capacity, such as R&D capabilities and hence benefitting more from locally circulating knowledge.

Our results may have some important policy implications, especially for developing countries. First, as cognitive proximity matters for innovation, policy intervention should focus on developing and supporting the local stock of related knowledge that can be more easily recombined. However, the proximity paradox suggests that the local knowledge space cannot be too narrow because of the risk of cognitive lock-in. Thus, policies oriented to increasing the stock of local knowledge also need to pay attention to ways of diversifying it, with a special focus on more complex knowledge that can be recombined to create novel innovations. A greater variety of related knowledge at the local level is associated with better economic performance (Frenken, Oort, and Verburg 2007), the emergence of new industries (Galetti, Tessarin, and Morceiro 2021), and innovation (Castaldi, Frenken, and Los 2015). Second, policies oriented to strengthen the connections among firms' managers and workers, and between them and other social actors, should also be considered. Trade missions, sectoral industry fairs and workshops organised by government and business associations could result in more frequent interactions and more access to novel information. Third, policies could encourage new financial mechanisms to strengthen the internal resources and the R&D capacity of MSEs, given that these resources are relevant to compensate for the lack of MSEs' external support. This is particularly pertinent for informal MSEs, which generally do not benefit much from government innovation incentives and other support measures. Fourth, it is important that policies promote investments in physical infrastructure and digital technologies, which are essential for establishing external pipelines with universities, research institutions, suppliers and customers within and outside the region. Finally, while we find interesting results in terms of the importance of cognitive proximity and knowledge space position for the innovation performance of informal MSEs, their overall innovation performance remains inferior to formal MSEs. While recognising the innovative capacity of informal firms, this finding nevertheless points to the importance of regularising and formalising these firms, allowing them to recombine more complex knowledge and increase the share of novel information.

This study is not without its limitations. While our dataset is rich and novel, it provides a static snapshot, whereas the mechanisms underlying learning and innovations are dynamic and can change over time. We expect that follow-up surveys can add a temporal dimension to increase our understanding of the dynamic evolution of local knowledge. Our analysis also focuses on a relational knowledge space framework (Hidalgo et al. 2007) rather than assessing a formal network made especially for collaborative innovations (Gilsing et al. 2008). Finally, we focus on cognitive proximity, with geographical proximity already established because firms are within the geographical limits of Johannesburg. In addition, studies might investigate the interactions between cognitive and spatial proximities, while new surveys can add new types of proximity, considering the role of organisational, institutional, and social proximities in innovation processes.

Acknowledgments

Open access funding provided by IReL.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Endnotes

[1] Pairs of firms significantly close in a cognitive dimension are those with a conditional probability of having similar knowledge sets that exceeds the probability of a random occurrence. See Section 3.3 for details.

[2] The survey was undertaken under the auspices of the DSI/NRF South African Research Chair in Industrial Development, through the project 'Community of Practice in Innovation and Inclusive Industrialisation'.

[3] For the definition of MSEs in South Africa, see https://www.gov.za/sites/default/files/gcis_document/201903/423041gon399.pdf and for an international definition, see OECD%20Glossary%20of%20Statistical%20Terms%20-%20Small%20and%20medium-sized%20enterprises%20(SMEs)%20Definition

[4] For example, the World Bank Enterprise Surveys, which are the most widely used sources of firm-level data internationally and especially in developing countries, exclude micro enterprises (with fewer than five employees) and informal enterprises.

[5] The reasons for attrition to 724 firms included, for example: refusal to be interviewed (103 firms); the business had closed since being surveyed in the census (4 firms); no response in attempts to set up an interview (67 firms); not doing any manufacturing activities (20 firms); wrong number (15 firms).

[6] Specifically, there are some theoretical issues with using a logit model with fixed effects, such as the sector fixed effects used in our regressions. The linear model also facilitates the intuitive interpretation of the results since the coefficients represent the marginal effects. The main results using LPM are similar to those for logit regressions in Tables A4 and A5 in the Appendix.

[7] Although novelty may be difficult to ascertain because of subjective characteristics, the Oslo Manual provides some recommendations that aim to minimise potential bias. The introduction of questions related to factual behaviours significantly diminishes the bias of subjective assessments. In addition, the novelty should be considered in the context of geographical area, sector, market and firm size (OECD 2018). In the data collection for our survey, 'help notes' were provided for interviewers to explain the meanings of each of the categories of innovation novelty to respondents, in accessible non-technical language while retaining consistency and accuracy regarding these categories. Nonetheless, it is possible that there could be some inaccuracies in these responses, for instance in a case where a respondent indicates that an innovation was new to the world without being aware that the same innovation had already been introduced in another country.

[8] Q18 in the panel 'Independent variables' in Table A1 in the Appendix lists the 22 manufacturing sectors.

[9] These two measures of informality are only weakly correlated, confirming that they capture different aspects of firm's formality.

[10] The value at which the effect of the cognitive proximity on innovation is maximised is found at the point $\frac{-\beta_{cog}}{2\beta_{quad}}$. For the overall innovation in Column 1, the coefficient for cognitive proximity is 0.0113 and the coefficient for the quadratic term is −0.00015.

References

Acs, Z. J., and D. B. Audretsch. 1988. "Innovation in Large and Small Firms: An Empirical Analysis." *American Economic Review* 78, no. 4: 678–690.

Alberti, F. G., F. Belfanti, and J. D. Giusti. 2021. "Knowledge Exchange and Innovation in Clusters: A Dynamic Social Network Analysis." *Industry & Innovation* 28, no. 7: 1–22. https://doi.org/10.1080/13662716.2021.1904840.

Andersson, M., and H. Lööf. 2012. "Small Business Innovation: Firm Level Evidence From Sweden." *Journal of Technology Transfer* 37, no. 5: 732–754. https://doi.org/10.1007/s10961-011-9216-9.

Andreoni, A., P. Mondliwa, S. Roberts, and F. Tregenna. 2021. "Framing Structural Transformation in South Africa and beyond." In *Structural Transformation in South Africa: The Challenges of Inclusive Industrial Development in a Middle-Income Country*, edited by A. Andreoni, P. Mondliwa, S. Roberts, and F. Tregenna. Oxford, United Kingdom: Oxford University Press.

Andreoni, A., and F. Tregenna. 2021. "The Middle-Income Technology Trap and Premature Deindustrialization in South Africa." In *Structural Transformation in South Africa: The Challenges of Inclusive Industrial Development in a Middle-Income Country*, edited by A. Andreoni, P. Mondliwa, S. Roberts, and F. Tregenna. Oxford, United Kingdom: Oxford University Press.

Bathelt, H., A. Malmberg, and P. Maskell. 2004. "Clusters and Knowledge: Local Buzz, Global Pipelines and the Process of Knowledge Creation." *Progress in Human Geography* 28, no. 1: 31–56. https://doi.org/10.1191/0309132504ph469oa.

Booyens, I. 2011. "Are Small, Medium- and Micro-Sized Enterprises Engines of Innovation? The Reality in South Africa." *Science and Public Policy* 38, no. 1: 67–78. https://doi.org/10.1093/scipol/38.1.67.

Booyens, I., T. G. B. Hart, and K. H. Ramoroka. 2018. "Local Innovation Networking Dynamics: Evidence From South Africa." *European Journal of Development Research* 30, no. 4: 749–767. https://doi.org/10.1057/s41287-017-0123-2.

Boschma, R., and K. Frenken. 2010. "The Spatial Evolution of Innovation Networks: A Proximity Perspective." In *The Handbook of Evolutionary Economic Geography*, edited by R. Boschma and R. Martin. Cheltenham, United Kingdom: Edward Elgar Publishing.

Boschma, R. 2005. "Proximity and Innovation: A Critical Assessment." *Regional Studies* 39, no. 1: 61–74. https://doi.org/10.1080/0034340052000320887.

Boschma, R., P.-A. Balland, and D. F. Kogler. 2015. "Relatedness and Technological Change in Cities: The Rise and Fall of Technological Knowledge in US Metropolitan Areas From 1981 to 2010." *Industrial and Corporate Change* 24, no. 1: 223–250. https://doi.org/10.1093/icc/dtu012.

Brenner, T. 2007. "Local Knowledge Resources and Knowledge Flows." *Industry & Innovation* 14, no. 2: 121–128. https://doi.org/10.1080/13662710701252310.

Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2011. "Robust Inference With Multiway Clustering." *Journal of Business & Economic Statistics* 29, no. 2: 238–249. https://doi.org/10.1198/jbes.2010.07136.

Caragliu, A., L. de Dominicis, and H. L. F. de Groot. 2016. "Both Marshall and Jacobs Were Right." *Economic Geography* 92, no. 1: 87–111. https://doi.org/10.1080/00130095.2015.1094371.

Castaldi, C., K. Frenken, and B. Los. 2015. "Related Variety, Unrelated Variety and Technological Breakthroughs: An Analysis of US State-Level Patenting." *Regional Studies* 49, no. 5: 767–781. https://doi.org/10.1080/00343404.2014.940305.

Cirera, X., and S. Muzi. 2020. "Measuring Innovation Using Firm-Level Surveys: Evidence From Developing Countries." *Research Policy* 49, no. 3: 103912. https://doi.org/10.1016/j.respol.2019.103912.

City of Johannesburg. 2021. City of Johannesburg Integrated Development Plan 2021–26. https://www.joburg.org.za/documents_/Documents/2021-2016%20Final%20IDP/2021-26%20FINAL%20IDP%2021May%202021.pdf.

Coccia, M. 2017. "Sources of Technological Innovation: Radical and Incremental Innovation Problem-Driven to Support Competitive Advantage of Firms." *Technology Analysis & Strategic Management* 29, no. 9: 1048–1061. https://doi.org/10.1080/09537325.2016.1268682.

Cohen, W. M., and D. A. Levinthal. 1990. "Absorptive Capacity: A New Perspective on Learning and Innovation." *Administrative Science Quarterly* 35, no. 1: 128–152. https://doi.org/10.2307/2393553.

Content, J., and K. Frenken. 2016. "Related Variety and Economic Development: A Literature Review." *European Planning Studies* 24, no. 12: 2097–2112. https://doi.org/10.1080/09654313.2016.1246517.

Coscia, M. 2021. *The Atlas for the Aspiring Network Scientist*. Copenhagen, Denmark: IT University of Copenhagen.

Crepon, B., E. Duguet, and J. Mairessec. 1998. "Research, Innovation and Productivity: An Econometric Analysis at the Firm Level." *Economics of Innovation and New Technology* 7, no. 2: 115–158. https://doi.org/10.1080/10438599800000031.

Dosi, G. 1988. "Sources, Procedures, and Microeconomic Effects of Innovation." *Journal of Economic Literature* 26, no. 3: 1120–1171.

Eder, J., and M. Trippl. 2019. "Innovation in the Periphery: Compensation and Exploitation Strategies." *Growth and Change* 50, no. 4: 1511–1531. https://doi.org/10.1111/grow.12328.

Escolar, E. G., Y. Hiraoka, M. Igami, and Y. Ozcan. 2023. "Mapping Firms' Locations in Technological Space: A Topological Analysis of Patent Statistics." *Research Policy* 52, no. 8: 104821. https://doi.org/10.1016/j.respol.2023.104821.

Farinha, T., P.-A. Balland, A. Morrison, and R. Boschma. 2019. "What Drives the Geography of Jobs in the US? Unpacking Relatedness." *Industry & Innovation* 26, no. 9: 988–1022. https://doi.org/10.1080/13662716.2019.1591940.

Fitjar, R. D., F. Huber, and A. Rodríguez-Pose. 2016. "Not Too Close, Not Too Far: Testing the Goldilocks Principle of 'optimal' Distance in Innovation Networks." *Industry & Innovation* 23, no. 6: 465–487. https://doi.org/10.1080/13662716.2016.1184562.

Fitjar, R. D., and B. Timmermans. 2017. "Regional Skill Relatedness: Towards a New Measure of Regional Related Diversification." *European Planning Studies* 25, no. 3: 516–538. https://doi.org/10.1080/09654313.2016.1244515.

Frenken, K., F. V. Oort, and T. Verburg. 2007. "Related Variety, Unrelated Variety and Regional Economic Growth." *Regional Studies* 41, no. 5: 685–697. https://doi.org/10.1080/00343400601120296.

Gaglio, C., E. Kraemer-Mbula, and E. Lorenz. 2022. "The Effects of Digital Transformation on Innovation and Productivity: Firm-Level Evidence of South African Manufacturing Micro and Small Enterprises." *Technological Forecasting and Social Change* 182: 121785. https://doi.org/10.1016/j.techfore.2022.121785.

Galetti, J. R. B., M. S. Tessarin, and P. C. Morceiro. 2021. "Skill Relatedness, Structural Change and Heterogeneous Regions: Evidence From a Developing Country." *Papers in Regional Science* 100, no. 6: 1355–1376. https://doi.org/10.1111/pirs.12629.

Galetti, J. R. B., M. S. Tessarin, and P. C. Morceiro. 2024. "Types of Occupational Relatedness and Branching Processes Across Brazilian Regions." *Area Development and Policy* 9, no. 1: 1–23. https://doi.org/10.1080/23792949.2022.2117217.

Gilsing, V., B. Nooteboom, W. Vanhaverbeke, G. Duysters, and A. van den Oord. 2008. "Network Embeddedness and the Exploration of Novel Technologies: Technological Distance, Betweenness Centrality and Density." *Research Policy* 37, no. 10: 1717–1731. https://doi.org/10.1016/j.respol.2008.08.010.

Grillitsch, M., B. Asheim, and M. Trippl. 2018. "Unrelated Knowledge Combinations: The Unexplored Potential for Regional Industrial Path Development." *Cambridge Journal of Regions, Economy and Society* 11, no. 2: 257–274. https://doi.org/10.1093/cjres/rsy012.

Grillitsch, M., and M. Nilsson. 2015. "Innovation in Peripheral Regions: Do Collaborations Compensate for a Lack of Local Knowledge Spillovers?" *Annals of Regional Science* 54, no. 1: 299–321. https://doi.org/10.1007/s00168-014-0655-8.

Hervás-Oliver, J.-L., M. D. Parrilli, A. Rodríguez-Pose, and F. Sempere-Ripoll. 2021. "The Drivers of SME Innovation in the Regions of the EU." *Research Policy* 50, no. 9: 104316. https://doi.org/10.1016/j.respol.2021.104316.

Hidalgo, C. A., B. Klinger, A.-L. Barabási, and R. Hausmann. 2007. "The Product Space Conditions the Development of Nations." *Science* 317, no. 5837: 482–487. https://doi.org/10.1126/science.1144581.

Imbs, J., and R. Wacziarg. 2003. "Stages of Diversification." *American Economic Review* 93, no. 1: 63–86. https://doi.org/10.1257/000282803321455160.

International Finance Corporation (IFC). 2018. *The Unseen Sector: A Report on the MSME Opportunity in South Africa*. Washington, D.C., United States: IFC. https://www.ifc.org/en/insights-reports/2019/msme-opportunity-south-africa.

Isaksen, A., and M. Trippl. 2016. "Path Development in Different Regional Innovation Systems: A Conceptual Analysis." In *Innovation Drivers and Regional Innovation Strategies*, edited by M. D. Parrilli, R. D. Fitjar, and A. Rodríguez-Pose. New York, NY, United States: Routledge.

Jong, J. P. J. de, and P. A. M. Vermeulen. 2006. "Determinants of Product Innovation in Small Firms: A Comparison Across Industries." *International Small Business Journal* 24, no. 6: 587–609. https://doi.org/10.1177/0266242606069268.

Kamutando, G., and F. Tregenna. 2023. "Complementarity Between Product and Process Innovation in Small and Micro-Enterprises in Johannesburg, South Africa." *Economics of Innovation and New Technology* 33, no. 6: 1–27. https://doi.org/10.1080/10438599.2023.2254241.

Kogler, D. F., D. L. Rigby, and I. Tucker. 2013. "Mapping Knowledge Space and Technological Relatedness in US Cities." *European Planning Studies* 21, no. 9: 1374–1391. https://doi.org/10.1080/09654313.2012.755832.

Kraemer-Mbula, E., E. Lorenz, L. T. Greenish, et al. 2019. "Are African Micro- and Small Enterprises Misunderstood? Unpacking the Relationship Between Work Organisation, Capability Development and Innovation." *International Journal of Technological Learning, Innovation and Development* 11, no. 1: 1. https://doi.org/10.1504/IJTLID.2019.097411.

Krugell, W., and N. Rankin. 2012. "Agglomeration and Firm-Level Efficiency in South Africa." *Urban Forum* 23, no. 3: 299–318. https://doi.org/10.1007/s12132-012-9144-2.

Malmberg, A., and P. Maskell. 2006. "Localized Learning Revisited." *Growth and Change* 37, no. 1: 1–18. https://doi.org/10.1111/j.1468-2257.2006.00302.x.

Martynovich, M., and J. Taalbi. 2022. "Dynamic Recombinant Relatedness and its Role for Regional Innovation." *European Planning Studies* 31, no. 5: 1–25. https://doi.org/10.1080/09654313.2022.2121154.

Musara, M., and C. Nieuwenhuizen. 2020. "Informal Sector Entrepreneurship, Individual Entrepreneurial Orientation and the Emergence of Entrepreneurial Leadership." *Africa Journal of Management* 6, no. 3: 194–213. https://doi.org/10.1080/23322373.2020.1777817.

Neffke, F. 2019. "The Value of Complementary Co-Workers." *Science Advances* 5, no. 12: Article 12. https://doi.org/10.1126/sciadv.aax3370.

Neffke, F., and M. Henning. 2013. "Skill Relatedness and Firm Diversification." *Strategic Management Journal* 34, no. 3: 297–316: Article 3. https://doi.org/10.1002/smj.2014.

Neffke, F., M. Henning, and R. Boschma. 2011. "How Do Regions Diversify Over Time? Industry Relatedness and the Development of New Growth Paths in Regions." *Economic Geography* 87, no. 3: 237–265. https://doi.org/10.1111/j.1944-8287.2011.01121.x.

Nelson, R. R., and S. G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, MA, United States: Belknap Press of Harvard University Press.

Nooteboom, B. 1994. "Innovation and Diffusion in Small Firms: Theory and Evidence." *Small Business Economics* 6, no. 5: 327–347. https://doi.org/10.1007/BF01065137.

Nooteboom, B. 2000. "Learning by Interaction: Absorptive Capacity, Cognitive Distance and Governance." *Journal of Management & Governance* 4, no. 1: 69–92. https://doi.org/10.1023/A:1009941416749.

Nooteboom, B., W. Van Haverbeke, G. Duysters, V. Gilsing, and A. van den Oord. 2007. "Optimal Cognitive Distance and Absorptive Capacity." *Research Policy* 36, no. 7: 1016–1034. https://doi.org/10.1016/j.respol.2007.04.003.

OECD. 2018. *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation*. 4th ed. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/science-and-technology/oslo-manual-2018_9789264304604-en.

Page, J., and M. Söderbom. 2015. "Is Small Beautiful? Small Enterprise, Aid and Employment in Africa." *African Development Review* 27, no. S1: 44–55. https://doi.org/10.1111/1467-8268.12138.

Paus, E., M. Robinson, and F. Tregenna. 2022. "Firm Innovation in Africa and Latin America: Heterogeneity and Country Context." *Industrial and Corporate Change* 31, no. 2: 338–357. https://doi.org/10.1093/icc/dtac006.

Rosenberg, N. 1982. *Inside the Black Box: Technology and Economics*. Cambridge, United Kingdom: Cambridge University Press.

Sammarra, A., and L. Biggiero. 2008. "Heterogeneity and Specificity of Inter-Firm Knowledge Flows in Innovation Networks." *Journal of Management Studies* 45, no. 4: 800–829. https://doi.org/10.1111/j.1467-6486.2008.00770.x.

Sookram, S., and P. K. Watson. 2008. "Small-Business Participation in the Informal Sector of an Emerging Economy." *Journal of Development Studies* 44, no. 10: 1531–1553. https://doi.org/10.1080/00220380802265520.

Stats, S. A. 2020. Quarterly Labour Force Survey, Quarter 1: 2020. [Online]. http://www.statssa.gov.za/publications/P0211/P02111stQuarter2020.pdf.

Steijn, M. P. A. 2021. "Improvement on the Association Strength: Implementing a Probabilistic Measure Based on Combinations Without Repetition." *Quantitative Science Studies* 2, no. 2: 778–794. https://doi.org/10.1162/qss_a_00122.

Stuck, J., T. Broekel, and J. Revilla Diez. 2016. "Network Structures in Regional Innovation Systems." *European Planning Studies* 24, no. 3: 423–442: Article 3. https://doi.org/10.1080/09654313.2015.1074984.

Torre, A., and A. Rallet. 2005. "Proximity and Localization." *Regional Studies* 39, no. 1: 47–59. https://doi.org/10.1080/0034340052000320842.

Trade & Industrial Policy Strategies (TIPS). 2023. *The State of Small Business in South Africa*. Pretoria, South Africa: The Real Economy Bulletin - Special Edition. https://www.tips.org.za/manufacturing-data/the-real-economy-bulletin/the-state-of-small-business-in-south-africa.

Tregenna, F., A. K. Ewinyu, A. Oqubay, and I. Valodia. 2022. "The South African Economy: Challenges and Complexities." In *The Oxford Handbook of the South African Economy*, edited by A. Oqubay, F. Tregenna, and I. Valodia. Oxford, United Kingdom: Oxford University Press.

Van der Panne, G. 2004. "Agglomeration Externalities: Marshall Versus Jacobs." *Journal of Evolutionary Economics* 14, no. 5: 593–604: Article 5. https://doi.org/10.1007/s00191-004-0232-x.

Weitzman, M. L. 1998. "Recombinant Growth." *Quarterly Journal of Economics* 113, no. 2: 331–360. https://doi.org/10.1162/003355398555595.

Whittle, A., and D. F. Kogler. 2020. "Related to What? Reviewing the Literature on Technological Relatedness: Where We Are Now and where Can We Go?." *Papers in Regional Science* 99, no. 1: 97–113. https://doi.org/10.1111/pirs.12481.

Winters, J. V. 2013. "Human Capital Externalities and Employment Differences Across Metropolitan Areas of the USA." *Journal of Economic Geography* 13, no. 5: 799–822. https://doi.org/10.1093/jeg/lbs046.

Wooldridge, J. M. 2003. "Cluster-Sample Methods in Applied Econometrics." *American Economic Review* 93, no. 2: 133–138. https://doi.org/10.1257/000282803321946930.

Xiao, J., R. Boschma, and M. Andersson. 2018. "Industrial Diversification in Europe: The Differentiated Role of Relatedness." *Economic Geography* 94, no. 5: 514–549. https://doi.org/10.1080/00130095.2018.1444989.

Zaheer, A., and G. G. Bell. 2005. "Benefiting From Network Position: Firm Capabilities, Structural Holes, and Performance." *Strategic Management Journal* 26, no. 9: 809–825. https://doi.org/10.1002/smj.482.

Zanello, G., X. Fu, P. Mohnen, and M. Ventresca. 2016. "The Creation and Diffusion of Innovation in Developing Countries: A Systematic Literature Review." *Journal of Economic Surveys* 30, no. 5: 884–912. https://doi.org/10.1111/joes.12126.

Appendix A

TABLE A1 | Variables.

Variables	Type	Definition	Survey questions
Dependent variables			
Innovation	Binary	Takes value equal to 1 when a MSE introduces an entirely new or significantly improved innovation, regardless of the type, and zero otherwise.	Q29. During the last financial year (FY2019), has your establishment done any of the following types of innovation? (i) Introduced entirely new products; (ii) significantly improved products; (iii) entirely new services; (iv) significantly improved services; (v) entirely new processes; (vi) significantly improved processes; (vii) none.
Product innovation	Binary	Takes value equal to 1 when a MSE introduces an entirely new or significantly improved product innovation, and zero otherwise.	Q29. During the last financial year (FY2019), has your establishment done any of the following types of innovation? (i) Introduced entirely new products; (ii) significantly improved products; (iii) none.

(Continues)

TABLE A1 | (Continued)

Variables	Type	Definition	Survey questions
Process and service innovation	Binary	Takes value equal to 1 when a MSE introduces an entirely new or significantly improved process or service innovation, and zero otherwise.	Q29. During the last financial year (FY2019), has your establishment done any of the following types of innovation? (i) Entirely new services; (ii) significantly improved services; (iii) entirely new processes; (iv) significantly improved processes; (v) none.
Innovation novelty	Binary	Takes value equal to 1 when a MSE introduces an innovation that has the indicated degree of novelty and zero otherwise.	Q38. In relation to the main innovation described above (Q33), was your main innovation (i) only new to your business unit; (ii) new to your community; (iii) new to your market; (iv) new to the world; (v) none.
Independent variables			
Cognitive proximity	Continuous	Indicates the levels of similar knowledge and expertise that a MSE shares with all others. It ranges from 1 to 100.	Q18. What is the main manufacturing activities you conduct in this establishment? Food products; beverages; tobacco products; textiles; wearing apparel; leather and related products; wood and of products of wood and cork, except furniture; articles of straw and plaiting materials; paper and paper products; printing and reproduction of recorded media; coke and refined petroleum products; chemicals and chemical products; pharmaceuticals, medicinal chemical and botanical products; Rubber and plastics products; other non-metallic mineral products; basic metals; fabricated metal products, except machinery and equipment; electrical equipment; machinery and equipment; motor vehicles, trailers and semi-trailers; furniture; other manufacturing; repair and installation of machinery and equipment—except motor vehicles. Q52. What three types of skills would you say are most needed in your business? From this list please select the three most needed skills in your business. Communication skills (e.g. internal, clients, suppliers etc.); marketing skills; negotiation skills; financial knowledge & budgetary skills; business planning skills; computer use; coding and programing; technical skills; complex problem-solving skills; 'community skills' (knowing of the community, personal networks, etc); creative skills (i.e. come up with creative ideas and solutions). Q56. What kind of ideas do you find most useful to exchange with other firms in your industry/trade? From this list please select the most useful things to exchange with other firms in your industry/trade. Training; ideas for new products or services; information about business management; information about government support; collective negotiations; protecting your new ideas; None/Don't know.
Degree centrality	Continuous	Indicates the number of firms that a MSE is directly connected with in the knowledge network. Ranges from 1 to 100.	
Betweenness centrality	Continuous	Indicates the extent to which a MSE is located between other MSEs in the knowledge network. Ranges from 1 to 100.	
Control variables			
Education	Continuous	Index indicating the educational level of the manager or owner of the MSE. Ranges from level 1 to 10.	Q4. Which of the following best describes the highest level of education completed by the owner/manager of the business? (1) No formal school; (2) informal schooling only; (3) some primary school; (4) primary school completed; (5) some secondary school/high school; (6) secondary/High school completed; (7) post-secondary qualifications, other than a university (e.g. diploma or degree from a university of technology or college); (8) some university; (9) university completed; (10) post-graduate degree.
R&D	Binary	Takes value equal to 1 when a MSE engages in R&D activities for innovation, and zero otherwise.	Q32. During the last financial year (FY2019) has your establishment engaged in R&D activities for innovation? (R&D refers to creative work undertaken on a systematic basis to increase the stock of knowledge and using it for innovation (including software) development).

(Continues)

TABLE A1 | (Continued)

Variables	Type	Definition	Survey questions
Size (log)	Continuous	Natural logarithm of the number of employees (full-, part-time and occasional)	Q44. At the end of fiscal year 2019 (28th February 2019), how many people worked in this establishment? Please include all employees and managers.
Skills—availability	Continuous	Indicates the extent of the skills constraint facing the MSE. Ranges from level 1 to 4.	Q46. How difficult is it for this establishment to find employees with the required skills? (1) Very difficult; (2) fairly difficult; (3) not very difficult; (4) not at all difficult.
Skills—retaining	Continuous	Indicates the capacity of MSE to retain workers. Ranges from level 1 to 4.	Q47. How difficult is it for this establishment to retain employees? (1) Very difficult; (2) fairly difficult; (3) not very difficult; (4) not at all difficult.
Skills—training	Binary	Takes value equal to 1 if employees participate in training activities (group 1), and zero otherwise (group 2).	Q48. What are the most important ways through which employees in this establishment become more skilled at their jobs? Please tick the most important. Group 1: Participating in training; learning from more experienced colleagues and supervisors, for example, by asking for help, through conversations, or by observing; structured apprenticeships. Group 2: Learning by doing (trying out different ways to carry out their tasks; trial and error); self-driven online training/Internet/Youtube.
Skills—stability	Continuous	Indicates the degree to which skills needed from workers in an MSE change over time. Ranges from level 1 to 4.	Q51. How quickly do the skills needed from the employees in this establishment change? (1) Very quickly; (2) fairly quickly; (3) not very quickly; (4) No change at all.
Experience	Continuous	Managers/owners' years of experience in the sector.	Q12. How many years of experience does the manager/owner have working in this field or sector?
Age of business	Continuous	Age of the business in years.	Q15. In what year did the business in this establishment begin operations?
Customer—type	Categorical	Indicates the type of customers from which a MSE gets most of its sales: Individuals (baseline), business or government.	Q20. From which type of customers do you get most of your sales? Tick the most frequent type of customer for your products. Group 1: Individuals. Group 2: Small businesses; medium businesses; large businesses. Group 3: Government or public institutions.
Customer—local	Binary	Takes value equal to 1 if a MSE has most of its customers located in South Africa, and zero otherwise.	Q21. Where are most of your customers located?
# Suppliers	Continuous	Number of suppliers a MSE has.	Q24. How many suppliers does this establishment have? (1) None/Don't know; (2) 1—10; (3) 11–20; (4) 21–50; (5) 51–100; (6) more than 100.
Member—Bus. assoc.	Binary	Takes a value equal to 1 if a MSE belongs to a business association, and zero otherwise.	Q54. Do you belong to any business association?
Co-operation	Binary	Takes a value equal to 1 if a MSE co-operates with other firms in the same industry/sector, and zero otherwise.	Q55. Do you co-operate with other firms in the same industry/trade as you?
Registration	Binary	Takes a value equal 1 if a MSE is formally registered with the South African revenue service (SARS), and zero otherwise.	Q66. Is this business registered with any organisation? Q66a. Which of the following organisations is the business registered with? (1) SARS; (2) local municipality; (3) business association; (4) other.
Formal workers	Binary	Takes a value of 1 if a MSE has all full-time and part-time workers with contracts, and zero otherwise.	Q44a. Of this total of employees, how many had contracts? (1) full-time; (2) part-time; (3) occasional.

TABLE A2 | Descriptive statistics—registered (formal) and non-registered (informal) mses and mses with registered (formal) and non-registered (informal) workers.

	Formal - registration			Informal—registration			Formal—workers			Informal—workers		
	N	Innovative MSEs	%	*N*	Innovative MSEs	%	*N*	MSEs	%	*N*	Innovative MSEs	%
Innovation	193	124	64.25	518	258	49.81	236	171	72.46	475	211	44.42
Product innovation	193	110	56.99	518	237	45.75	236	155	65.68	475	192	40.42
Process and service innovation	193	42	21.76	518	62	11.97	236	64	27.12	475	40	8.42
Innovation—less novel	193	73	37.82	518	173	33.4	236	105	44.49	475	141	29.68
Innovation—more novel	193	51	26.42	518	85	16.41	236	66	27.97	475	70	14.74
New to the business	193	56	29.02	518	98	18.92	236	71	30.08	475	83	17.47
New to the community	193	17	8.8	518	75	14.48	236	34	14.41	475	58	12.21
New to the market	193	42	21.76	518	69	13.32	236	55	23.31	475	56	11.79
New to the world	193	9	4.66	518	16	3.09	236	14	5.93	475	11	2.32
	N	**Mean**	**St. Dev.**	**N**	**Mean**	**St. Dev.**	**N**	**Mean**	**St. Dev.**	**N**	**Mean**	**St. Dev.**
Cognitive proximity	193	23.25	13.15	518	21.06	11.6	236	23.05	13.3	475	20.96	11.54
Cognitive proximity^2	193	722.4	1013.10	518	578	707.4	236	707.5	955	475	572.3	713.78
Degree centrality	193	43.09	10.11	518	43.03	9.55	236	42.82	10.31	475	43.16	9.39
Betweenness centrality	193	4.9	4.27	518	5.15	6.38	236	5.62	7.99	475	4.81	4.46
Education	193	6.7	2.03	518	5.44	1.9	236	6.52	2.01	475	5.41	1.91
R&D	193	0.27	0.44	518	0.1	0.3	236	0.25	0.43	475	0.09	0.29
Size (log)	193	1.86	1.07	518	1.06	0.91	236	1.7	1.15	475	1.07	0.87
Skills—availability	193	2.22	1.15	518	1.97	1.13	236	2.11	1.12	475	1.99	1.14
Skills—retaining	193	2.56	1.05	518	2.31	1.13	236	2.42	1.13	475	2.35	1.1
Skills—training	193	0.83	0.38	518	0.77	0.42	236	0.85	0.36	475	0.75	0.43
Skills—stability	193	2.37	1.1	518	2.14	1.08	236	2.14	1.11	475	2.23	1.08
Experience	193	13.85	11.02	518	13.06	10.39	236	13.7	11.68	475	13.06	9.97
Age of business	193	17.44	16.52	518	13.88	12.81	236	16.45	15.87	475	14.04	12.91
# Suppliers	193	2.36	1.1	518	1.97	0.67	236	2.33	0.95	475	1.94	0.73
Member—Bus. assoc.	193	0.18	0.38	518	0.06	0.24	236	0.15	0.36	475	0.06	0.25
Co-operation	193	0.52	0.5	518	0.46	0.5	236	0.54	0.5	475	0.44	0.5
Registration	193	1	0	518	0	0	236	0.42	0.49	475	0.2	0.4
Formal workers	193	0.51	0.5	518	0.26	0.44	236	1	0	475	0	0

TABLE A3 | Summary statistics: control and treated MSEs at baseline (k-to-k matched).

	Non-innovative (control)	Treated (control)	Diff. In mean	*p*-value
Overall innovation				
Education	5.39	5.52	−0.13	0.40
Skills—availability	1.94	1.94	0	1.00
Skills—retaining	2.17	2.25	−0.08	0.50
Skills—stability	2.01	1.94	0.07	0.53
Experience	11.95	11.91	0.04	0.97
Age of business	10.86	10.66	0.2	0.80
# Suppliers	1.84	1.92	−0.08	0.22
Size (log)	3.54	3.84	−0.3	0.50
Product innovation				
Education	5.49	5.65	−0.16	0.31

(Continues)

TABLE A3 | (Continued)

	Non-innovative (control)	Treated (control)	Diff. In mean	*p*-value
Skills—availability	1.93	1.96	−0.03	0.78
Skills—retaining	2.18	2.30	−0.12	0.29
Skills—stability	2.02	1.95	0.07	0.53
Experience	11.64	11.60	0.03	0.97
Age of business	10.66	10.70	−0.04	0.96
# Suppliers	1.85	1.90	−0.06	0.32
Size (log)	3.83	3.54	0.29	0.50
Process and service innovation				
Education	6.15	6.32	−0.17	0.59
Skills—availability	1.92	1.97	−0.05	0.80
Skills—retaining	2.27	2.20	0.07	0.76
Skills—stability	2.05	1.97	0.08	0.64
Experience	11.85	11.81	0.03	0.99
Age of business	10.25	9.75	0.51	0.71
# Suppliers	2.07	2.12	−0.05	0.70
Size (log)	5.60	6.83	−1.23	0.32

Note: We report *t*-test for mean differences between treated and control groups, that is, for MSEs that introduced innovation and those that did not innovate, respectively. For Education, there are 5 bins (No formal school; Until primary school; Post-Secondary qualifications, other than a university (e.g. diploma or degree from a University of Technology or College); Some university; Post-graduate degree). Age of business was classified as less than 10 years; between 10 and less than 30 years; 30–50 years; and 50–120 years. Experience of manager in business: less than 10 years; 10–30 years; and 30–55 years. Size of MSEs: less than 10 employees; 10 to 30 employees, and 30 to 50 employees. Skill variables were classified according to the categories described in Table A1.

TABLE A4 | Logit model—overall innovation, product innovation, and non-product innovation.

	Overall innovation (1)	Product innovation (2)	Non-product innovation (3)	Innovation—Less novel (4)	Innovation—More novel (5)
Cognitive proximity	0.012*** (0.003)	0.010*** (0.002)	0.007*** (0.002)	0.024*** (0.006)	−0.005 (0.004)
Cognitive proximity^2	−0.0002*** (0.0001)	−0.0001** (0.0001)	−0.0002*** (0.0001)	−0.0004*** (0.0001)	0.0001* (0.00004)
Degree centrality	0.001 (0.001)	0.003 (0.002)	−0.004** (0.002)	0.001 (0.002)	−0.001 (0.001)
Betweenness centrality	0.010* (0.005)	0.010** (0.005)	0.0003 (0.002)	0.009** (0.004)	0.001 (0.001)
Education	0.053*** (0.014)	0.048*** (0.015)	0.025*** (0.005)	0.035*** (0.010)	0.019* (0.012)
R&D	0.124** (0.055)	0.112 (0.072)	0.078** (0.039)	0.052 (0.049)	0.050** (0.026)
Size (log)	0.015 (0.013)	−0.018 (0.012)	0.036*** (0.010)	−0.006 (0.017)	0.014 (0.010)
Skills—availability	−0.069*** (0.020)	−0.055** (0.025)	−0.033*** (0.012)	−0.050*** (0.010)	−0.019 (0.020)
Skills—retaining	0.021 (0.024)	0.031* (0.019)	0.001 (0.013)	0.009 (0.021)	0.011 (0.009)
Skills—training	0.047 (0.034)	0.050 (0.042)	0.010 (0.023)	0.076** (0.037)	−0.030 (0.031)
Skills—stability	−0.035** (0.018)	−0.040** (0.017)	−0.010 (0.012)	−0.030*** (0.011)	−0.005 (0.016)
Customer—business	−0.005 (0.042)	−0.018 (0.045)	0.014 (0.033)	0.001 (0.041)	0.003 (0.022)
Customer—govt.	0.462*** (0.001)	0.082 (0.251)	0.211 (0.181)	−0.075 (0.194)	0.369 (0.358)
Customer—abroad	0.226*** (0.073)	0.235*** (0.058)	−0.109*** (0.042)	0.018 (0.136)	0.168 (0.114)
# Suppliers	0.031 (0.022)	0.034 (0.025)	0.004 (0.016)	0.043** (0.019)	−0.010 (0.011)
Member—Bus. assoc.	−0.084** (0.040)	−0.073*(0.043)	−0.031 (0.021)	−0.059 (0.048)	−0.019 (0.064)
Co-operation	0.076** (0.034)	0.068* (0.038)	−0.001 (0.028)	0.077** (0.034)	0.005 (0.028)
Experience	0.001 (0.001)	−0.001 (0.002)	0.003 (0.002)	−0.001 (0.002)	0.002 (0.001)
Age of business	−0.002 (0.001)	−0.001 (0.001)	−0.001 (0.001)	0.0003 (0.001)	−0.003** (0.001)
Registration	0.020 (0.032)	0.019 (0.038)	0.0005 (0.043)	−0.043 (0.046)	0.057** (0.030)
Formal workers	0.176*** (0.033)	0.174*** (0.035)	0.111*** (0.032)	0.079** (0.038)	0.088*** (0.033)
Observations	711	711	711	711	711

(Continues)

TABLE A4 | (Continued)

	Overall innovation (1)	Product innovation (2)	Non-product innovation (3)	Innovation—Less novel (4)	Innovation—More novel (5)
Industry dummies	YES	YES	YES	YES	YES
Nagelkerke R^2	0.325	0.263	0.301	0.192	0.184
Chi-square	198.43***	156.59***	132.59***	106.72***	86.928***

Note: Heteroskedasticity-robust standard errors are shown in parentheses.
Coefficients represent average marginal effects (AME) from logit regressions and are statistically significant at *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

TABLE A5 | Logit model—overall innovation, product innovation, and non-product innovation for formal and informal MSEs.

	Overall innovation		Product innovation		Non-product innovation	
	Registration (1)	Formal workers (2)	Registration (3)	Formal workers (4)	Registration (5)	Formal workers (6)
Cognitive proximity	0.016*** (0.004)	0.012*** (0.004)	0.015*** (0.004)	0.013*** (0.004)	0.008 (0.008)	0.005 (0.005)
Cognitive proximity^2	−0.0002*** (0.0001)	−0.0001 (0.0001)	−0.0001** (0.0001)	−0.0001 (0.0001)	−0.0002 (0.0002)	−0.0001 (0.0001)
Degree centrality	0.004*** (0.002)	0.001 (0.002)	0.007*** (0.002)	0.002 (0.002)	−0.003 (0.002)	−0.007* (0.004)
Betweenness centrality	0.012** (0.006)	0.013*** (0.004)	0.011** (0.006)	0.015*** (0.005)	0.001 (0.002)	0.005 (0.004)
Education	0.051*** (0.014)	0.053*** (0.014)	0.045*** (0.015)	0.048*** (0.015)	0.027*** (0.007)	0.024*** (0.006)
R&D	0.121** (0.054)	0.117** (0.052)	0.107 (0.069)	0.102 (0.070)	0.077** (0.040)	0.080** (0.036)
Size (log)	0.009 (0.013)	0.016 (0.014)	−0.025** (0.012)	−0.017 (0.013)	0.037*** (0.009)	0.035*** (0.011)
Skills—availability	−0.064*** (0.019)	−0.067*** (0.020)	−0.050** (0.023)	−0.052** (0.024)	−0.033*** (0.011)	−0.032** (0.013)
Skills—retaining	0.020 (0.023)	0.022 (0.024)	0.028 (0.018)	0.032* (0.020)	0.001 (0.013)	0.006 (0.013)
Skills—training	0.046 (0.032)	0.050* (0.031)	0.046 (0.040)	0.052 (0.039)	0.011 (0.021)	0.013 (0.024)
Skills—stability	−0.038** (0.018)	−0.036** (0.017)	−0.044** (0.018)	−0.041** (0.017)	−0.011 (0.011)	−0.010 (0.011)
Customer—business	0.005 (0.040)	−0.009 (0.041)	−0008 (0.043)	−0.024 (0.045)	0.014 (0.031)	0.015 (0.035)
Customer—govt.	0.462*** (0.001)	0.462*** (0.001)	0.090 (0.217)	0.085 (0.250)	0.195 (0.190)	0.242 (0.198)
Customer—abroad	0.227** (0.081)	0.209*** (0.075)	0.240*** (0.065)	0.211*** (0.055)	−0.110** (0.043)	−0.113*** (0.040)
# Suppliers	0.030 (0.022)	0.032 (0.022)	0.031 (0.025)	0.034 (0.024)	0.001 (0.017)	0.007 (0.018)
Member—bus. assoc.	−0.074* (0.044)	−0.080* (0.044)	−0.066 (0.047)	−0.070 (0.045)	−0.034 (0.021)	−0.025 (0.023)
Co-operation	0.083** (0.033)	0.075** (0.032)	0.077** (0.038)	0.066* (0.035)	−0.003 (0.028)	−0.002 (0.037)
Experience	0.001 (0.001)	0.001 (0.001)	−0.001 (0.002)	−0.001 (0.002)	0.003 (0.002)	0.003 (0.002)
Age of business	−0.002 (0.001)	−0.002 (0.001)	−0.001 (0.001)	−0.001 (0.001)	−0.001 (0.001)	−0.002 (0.001)
Registration	0.494*** (0.029)	0.026 (0.030)	0.517*** (0.030)	0.026 (0.035)	0.211 (0.275)	−0.003 (0.041)
Formal workers	0.183*** (0.033)	0.242 (0.197)	0.182*** (0.033)	0.330 (0.198)	0.111*** (0.032)	−0.139 (0.199)
Registration × cognitive proximity	−0.018** (0.009)		−0.019** (0.009)		−0.001 (0.014)	
Registration × cognitive proximity^2	0.0001 (0.0001)		0.0001 (0.0001)		0.00002 (0.0002)	
Registration × degree centrality	−0.009*** (0.003)		−0.012*** (0.004)		−0.003 (0.003)	
Registration × betweenness centrality	−0.009 (0.009)		−0.001 (0.010)		−0.007 (0.006)	
Formal workers × cognitive proximity		−0.002 (0.012)		−0.010 (0.009)		0.006 (0.010)

(Continues)

TABLE A5 | (Continued)

	Overall innovation		Product innovation		Non-product innovation	
	Registration (1)	Formal workers (2)	Registration (3)	Formal workers (4)	Registration (5)	Formal workers (6)
Formal workers × cognitive proximity^2		−0.0001 (0.0002)		0.00004 (0.0002)		−0.0001 (0.0002)
Formal workers × degree centrality		0.002 (0.003)		0.002 (0.003)		0.006 (0.004)
Formal workers × betweenness centrality		−0.010 (0.006)		−0.013* (0.007)		−0.007 (0.007)
Observations	711	711	711	711	711	711
Industries dummies	YES	YES	YES	YES	YES	YES
Nagelkerke R^2	0.341	0.332	0.286	0.273	0.308	0.317
Chi-square	209.97***	203.74***	171.78***	163.31***	136.05***	140.47***

Note: Heteroskedasticity-robust standard errors are shown in parentheses.
Coefficients represent average marginal effects (AME) from logit regressions and are statistically significant at *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.