

Transforming small and medium enterprise integration into global value chains

Oliver Harman
Rahul Shukla



DIRECTED BY



FUNDED BY



**UK International
Development**

Partnership | Progress | Prosperity

Transforming Small and Medium Enterprise Integration into Global Value Chains*

Oliver Harman^{†‡} Rahul Shukla[§]

September 2026

IGC Working Paper GBR-26101

Abstract

Programmes that connect small firms to multinational buyers are expected to produce a link to the buyer, and the usual measure of whether they did is a count of the firms that came to trade with it. We ask what that count measures, using the internal record of TRANSFORM, a decade-long partnership between Unilever, the UK’s Foreign, Commonwealth and Development Office and EY in the form of an impact accelerator. We use a regression across 90 enterprises, configurational analyses coded from the enterprises’ own applications, and nineteen interviews with founders and with Unilever and programme staff. Sixteen per cent of enterprises ever integrated, and half of those relationships had lapsed by March 2026. Whether the buyer’s side saw a use for the proposed project at entry predicts integration, and ratings of the link during the project add nothing. In the programme team’s record, none of 24 enterprises whose project offered Unilever nothing it could use integrated, against 14 of 66 whose project did. A prior relationship with Unilever raises the rate above that threshold and is not necessary for it. Among enterprises the buyer had matched to a use at entry, no combination of what the enterprise itself brings, as recorded at entry, is sufficient for integration. Enterprises referred by a Unilever local market team, every one of them already known to the business, integrate at four times the rate of any other route. Those funded under the environmental pillars integrate no more often than those under the social ones. Staff interviews place the reason on the buyer’s side of the transaction. A contract needs several Unilever functions with unrelated objectives to agree at once, and an impact credential counts with the teams that fund pilots and carries no weight with those that sign contracts. Enterprises that never integrated report gains in capability and access to other buyers more often than those that did. An integration count therefore reflects the buyer’s purchasing decisions as much as the programme’s work, and such programmes should also be judged on what their participants keep.

Keywords: global value chains; supplier development; SME integration; buyer–supplier relationships; qualitative comparative analysis; corporate sustainability

JEL codes: F23 (multinational firms; international business); L14 (transactional relationships, contracts and reputation, networks); O19 (international linkages to development); Q56 (environment and development; sustainability); M14 (corporate culture; social responsibility)

*We thank Unilever staff, particularly Lisa Hawkes and James Holmes, for facilitating introductions and data sharing, and Rania Nasir for earlier engagement on funding acquisition, conceptualisation and research design. This research forms part of the project GBR-26101/ Integration of SMEs in Global Value Chains through MNEs, funded by FCDO Phase III (IGC) for £45,705. The research is conducted in collaboration with TRANSFORM, a joint initiative led by Unilever, the UK Government’s Foreign, Commonwealth & Development Office (FCDO), and EY (2015–2026), who provided contacts and access to entrepreneurs and project staff interviewed in the work. The views expressed are those of the authors and do not necessarily reflect those of TRANSFORM. Any errors are our own.

[†]Research Officer, Department of Geography and Environment, London School of Economics and Political Science, U.K.

[‡]Lecturer, Africa Urban Lab, African School of Economics, Zanzibar

[§]Research Coordinator, Firms, Trade and Productivity Initiative, International Growth Centre, London School of Economics and Political Science, U.K.

1 Introduction

Programmes that link small and medium enterprises in low- and middle-income countries to the supply chains of multinational buyers have become a common instrument of development policy. Funders judge such programmes on several outcomes, among them the reach of the supported enterprises to the households they serve. What sets the instrument apart from a grant is the buyer, and the outcome specific to it is how many of the enterprises the programme supports come to trade with that buyer. A relationship with a multinational can transmit finance, standards, management practice and market access to a small firm, and the evidence for those gains is strong. What that count measures is less clear. A buyer can extend a relationship only to firms it had already found useful, in which case the count records a choice the buyer made before the programme did anything. We do not know how much of an integration count is the programme's doing and how much is the buyer's selection. The answer is not obvious in advance. A programme that attaches the buyer's own staff to a small firm for two years could build a link that did not exist, or it could only confirm one the buyer had already seen. The question has stayed open because the evidence needed to answer it sits inside the buyer. Researchers seldom observe how a small firm came to a multinational's attention, whether the buyer saw a use for it before the programme began, whether the buyer's local business already knew it, or what the buyer's own organisation did while the programme ran.

We answer the question with the internal record of TRANSFORM, a partnership between Unilever, the UK's Foreign, Commonwealth and Development Office (FCDO) and EY. The programme ran as an impact accelerator from 2015 until its close in 2026, supported 90 enterprises across sub-Saharan Africa and South Asia, and attached a Unilever project lead to each of them. Its confidential portfolio tracker rates, for every enterprise it funded, how well the proposed project matched a use inside Unilever when the enterprise entered, which we call the match at entry. It also records whether the business already knew the enterprise or was working with it, who referred the enterprise, how far Unilever's local team engaged with it during the project, and whether it ever traded with Unilever. The programme team compiled the tracker after the event, so we read it beside two records written at the time and outside the buyer. The first is the enterprises' own applications, from which we code the same entry-stage conditions independently. The second is nineteen interviews with founders and with Unilever and programme staff, the only evidence we have on decisions inside the buyer.

We use these records in two ways. Across the 90 enterprises we compare integration rates by what was recorded at entry. Integration is rare, fourteen events in ninety, so we use exact tests, and penalised logits that correct for the small number of events. On the 56 enterprises with an application document, and again on the twelve whose founders we interviewed, we run a qualitative comparative analysis, which asks whether any combination of entry-stage characteristics is sufficient on its own for integration. We frame the question as three hypotheses, on selection at entry, on the buyer's decision after it, and on what participation produces. The design behind all three separates what was true of an enterprise before its project from what happened during it.

Sixteen per cent of the enterprises ever established a formal commercial relationship with Unilever, and half of those relationships had ended by March 2026. The first thing that sorts the enterprises is whether the buyer's side saw a use for the project at entry. None of the 24 enterprises whose project the programme team rated as offering Unilever nothing it could use went on to integrate, and 14 of the 66 rated as a match did. Above that threshold a stronger match adds little. The team's ratings of the link during the project add nothing to the match at entry as a predictor, and time in the programme is unassociated with integration. A prior relationship with the buyer is a separate matter. Of the 25 enterprises the business already knew, 9 integrated, against 4 of the 47 it did not know, so a relationship raises the rate without being a condition of integration. We read the two results together as evidence that an integration count records a judgement the buyer made at entry, about the project and about the enterprise, and that two years of programme support did not move enterprises across that threshold. What the count measures is therefore the buyer's selection as much as the programme's work. The threshold rests on a two-way table with an empty cell, so it does not depend on any modelling choice. The relationship result holds on the enterprises' own applications, with which the buyer's record agrees on two thirds of cases, and it holds with the twenty donor-selected enterprises removed. Because the tracker was compiled after the event, part of the threshold may be hindsight. We cannot exclude that reading, and we bound it throughout with the enterprises' contemporaneous applications.

The buyer’s country businesses made the matches that held. Enterprises that a Unilever market team put forward to the programme’s global selection panel integrated at 6 of 14, against 8 of 76 for every other route, and the programme’s rating of how supportive the local team was adds nothing once the route is known. The buyer’s record of prior relationships gives that result its content. Every one of the fourteen was already known to the country business or working with it before the programme, and 6 of the 9 integrators with a prior relationship arrived that way. Selection by head office, and the open calls run by regional panels, brought in enterprises the business did not know, and those integrated at the portfolio’s base rate. The referral result is therefore evidence about where a multinational’s knowledge of small local firms sits, in its country operations, and about how far a central programme can substitute for it. It does not imply that routing more enterprises through market teams would create the demand that made the route work.

Past that point the decision sat with the buyer. Among enterprises the buyer had matched at entry, no combination of the characteristics we can code from their own documents is sufficient for integration, in either coded sample. Those characteristics are the enterprise’s position in the chain, its credentials, its fit with the buyer’s stated priorities, its scale and its cohort. The combinations that do pick out integrators each describe one enterprise or two, and no set of conditions we try reaches more than three of the ten integrators in the larger sample. Founders and staff describe why. A supply or distribution contract needed procurement, brand and country teams inside Unilever to agree, and the project lead the programme provided could sponsor an enterprise without being able to commit any of them. The credentials the programme selected on did not help past that point. Enterprises funded under the environmental pillars integrated no more often than those under the social ones, and green credentials are associated with no higher rate on any basis we test. The reason staff give is that the teams that valued a credential were not the ones deciding what to buy.

Much of what the programme produced does not appear in the count at all. Among the relationships we can observe in detail, what the tracker calls integration is seldom first-tier supply, and the one enterprise supplying Unilever commercially among the interviewed cases had that relationship before it applied. Founders whose enterprises never integrated report a capability gained, and access to markets beyond Unilever, more often than those whose enterprises did, and they report credibility with other buyers less often. An integration count therefore understates what the programme did for the enterprises it supported, and we argue in the conclusion that a programme of this kind should be judged on what its participants gained whether or not they integrated.

We contribute, first, to the literature on how small firms in developing countries connect to multinational buyers. That work has established what a connection does for a supplier once it exists, raising employment and productivity through active transmission (e.g., Alfaro-Ureña et al., 2022), and that large-buyer entry sorts strongly on what suppliers already had (Iacovone et al., 2015). We observe the matching process from inside the buyer. The buyer’s record holds its judgement of the project and its knowledge of the enterprise separately, across a decade of one multinational’s supplier-development portfolio, and we show that the judgement is the threshold and the knowledge raises the rate above it. For the governance tradition, which treats the lead firm as one actor setting the terms of a chain (Gereffi et al., 2005), we add evidence that a lead firm’s decisions about a small supplier are divided among its functions and its country businesses. We show that integration most often fails between them, and that the matches which held were made by the country business that already knew the enterprise. That connects the procurement-side account of the buying firm (Villena, 2018; Xiao et al., 2018) and the economic-geography account of how multinationals engage the firms around their subsidiaries (Crescenzi et al., 2022; Crescenzi and Harman, 2023) to the value-chain outcome.

Second, we contribute to the evaluation of programmes that build such links. The corporate-accelerator literature finds that programmes selecting ventures for strategic fit identify firms already positioned for the relationship as much as they build capability (Shankar and Shepherd, 2019; Seitz et al., 2023). We study a public-private partnership in accelerator form whose partners came to hope its enterprises would enter the parent’s chain, and we separate what the programme produced for its participants from what the parent decided to buy. For the environmental-upgrading literature (Ponte, 2022a; De Marchi and Di Maria, 2019; Krishnan et al., 2023) we measure the integration gap between green and social suppliers inside a single buyer’s portfolio. On method, we read the buyer’s record beside the enterprises’ contemporaneous applications and the founders’ own histories, and we join set-theoretic and regression evidence in the manner of Schneider and Wagemann (2012). We treat the interview coding as data, and we code the configurational conditions from a corpus of applications and reports with a machine-assisted pipeline whose codes a second reader verified. Re-running that pipeline on application-stage evidence alone shows how much hindsight

enters when the whole record is used (Appendix B). The caution about integration counts as a performance metric applies to any programme that uses them as a measure of performance.

The rest of the paper is structured as follows. Section 2 sets out the literature and the three hypotheses it yields. Section 3 describes the programme, what it gives an enterprise, its records and the three samples. Section 4 sets out the methods and the rule by which the strands are read together. Section 5 reports the results by claim, and closes with the definition of integration, its association with reported impact, and the benefits of participation for enterprises that did not integrate. Sections 6 to 9 report robustness, discuss the findings against the literature, set out the limits of the design, and conclude. The appendix carries in full every result the text reports in a paragraph.

2 Literature and hypotheses

The evidence is strong on some questions but much weaker on others, particularly on how firms enter these relationships. We organise it around the question the paper asks, in five strands. They are the returns to a buyer link (Section 2.1), selection and treatment (Section 2.2), what the supplier brings and what follows (Section 2.3), what the buyer decides (Section 2.4) and impact credentials (Section 2.5). Each strand closes on the hypothesis or condition it became, and Supplementary Appendix S1 gives the audit trail from evidence to interview question.¹ We give more attention to the buyer’s side of the relationship than the literature does, because that is where we locate the binding constraint.

2.1 The returns to a multinational buyer link

The strongest evidence in this field concerns what happens to a small firm after it starts supplying a large one. Using Costa Rican data on the universe of firm-to-firm transactions, Alfaro-Ureña et al. (2022) find that first-time suppliers to multinationals gain four to nine per cent in productivity and twenty-six per cent in employment within four years, through active transmission. Masso and Vahter (2023) recover comparable magnitudes in Belgium and Estonia. Experimental studies establish the effect as causal. Randomly allocated export orders raised the quality and profits of Egyptian rug producers (Atkin et al., 2017), and management consulting raised the productivity of Indian textile firms (Bloom et al., 2013). The benefits extend to exporting, to sales to other buyers, to skills and wages, and, where certification shifts a firm’s position, to revenue (Carballo et al., 2021; Brambilla et al., 2012; Heyman et al., 2007; Bolwig et al., 2009). If integration delivers returns of this magnitude, the count of integrated firms becomes an obvious performance metric for a supplier-development programme.

The gains run through direct, first-tier supply, and second-tier suppliers in the Costa Rican data show none. Direct supply is also a narrower thing than a direct commercial relationship. A distributor that buys the multinational’s products and sells them on is linked to the buyer directly, but as its customer, on the other side of the transaction from the suppliers this evidence studies. The spillover literature finds its gains among suppliers, through backward linkages (Javorcik, 2004). Among the enterprises whose relationship with Unilever we can observe in detail, the only first-tier supply relationship pre-dates the programme, and the rest are intermediated flows and platforms that distribute Unilever’s products (Section 5.8). For the one test of returns the data allow, we therefore expect a null, and Section 5.9 reports it as such.

The strand also motivates H3. If what a buyer relationship transmits is capability, then what a programme built to form such relationships produces for its participants may be observable whether or not the relationship formed.

2.2 Made or matched? Selection and treatment in supplier development

That evidence is drawn overwhelmingly from firms that became suppliers and is largely silent on how they came to do so. Where entry is observed, the evidence points to sorting. The firms that benefited from Walmart’s entry into Mexico were those already equipped to respond (Iacovone et al., 2015), and the organising concept is absorptive capacity (Cohen and Levinthal, 1990). The evidence for it runs from low-capacity firms seeing limited gains from participation (Pleticha, 2021) to capability and international-business factors predicting SME performance while access to resources does not (Abe, 2015). The second result matters for programmes whose principal instrument is finance. Part of the sorting is on the entrepreneur. Founder-level international background raises export orientation and performance (Filatotchev et al., 2009), and in our own setting Colonnelli et al. (2026) find that an African founder’s prior exposure to a country raises the probability of being funded from it several-fold, with the effect running through the prior connection.

The experimental results point the other way. If advice can improve management practice, then a firm’s initial capability partly reflects the support it has already received. The practical question for any programme is then which margin it operates on, creating capability and connections or identifying firms already on the path (Paus and Gallagher, 2008).

¹The thematic priorities underpinning the interview instrument originated with the programme’s funders (FCDO and Unilever). We treat them as a starting point and re-ground them in the wider SME-GVC literature; the resulting organisation drives both the interview schedule and the QCA conditions of Section 4.2.2.

TRANSFORM is a public-private partnership in the form of what the programme itself calls an impact accelerator, with a corporate parent attaching its own staff to the ventures it funds. The corporate-accelerator literature asks the same question. Corporations select ventures for strategic fit with the parent or for venture emergence (Shankar and Shepherd, 2019), and programmes selecting on fit produce higher growth for the ventures they accelerate and worse follow-on financing (Seitz et al., 2023). The cleanest estimates of what participation adds compare accelerated ventures with those selected but not accelerated, or with rejected applicants (Cohen et al., 2019; Assenova and Amit, 2024; Yu, 2020). This study could not obtain that counterfactual.

The same literature separates two things a programme can be judged on. One is whether its ventures reach the commercial relationship the parent wants. The other is what they gain from the programme whether or not they do. This distinction gives us H1 (selection at the gate) and H3 (what participation produces), and it is the reason our data must separate what was true of a firm before the project from what happened during it. It also motivates two conditions. `UL_LINK_START` records whether a Unilever relationship pre-existed the intervention, and `TIE_PRIOR` whether the founder arrived with the international reach that governs access to capital in this region.

2.3 What the supplier brings, and what happens afterwards

The barriers literature is where the field has invested most, and it locates the constraint on the supplier’s side. Financial, technological and skills gaps limit entry (Rahman et al., 2020; Koleča, 2018; Horváth and Szabó, 2019). Access to finance shows the largest gap between demand and supply of any participation factor. External credit, though, benefits firms already in value chains while burdening the rest (Shinozaki, 2015), which is a selection result in its own right. Compliance is the second recurring constraint. Lead firms push hidden sustainability compliance costs onto producers (Ponte, 2022b), and informal actors who cannot be monitored are excluded outright (Narula, 2019), though standards regimes have also opened markets (World Bank, 2020). The enabling-conditions literature adds digital access, infrastructure, institutions and the state’s several roles (Oliveira et al., 2021; Foster et al., 2018; Vandenberg et al., 2015; De Marchi and Alford, 2022; Reardon et al., 2021).

The bulk of the interview instrument on firm readiness and compliance cost comes from this literature (Supplementary Appendix S1). It is also the default against which our results should be read, since if integration failed in this portfolio the first expectation is that it failed for these reasons. It gives the question we test under H2, whether the obstacle in any non-integrating case was a supplier-side one.

Two further strands bear on how integration should be read, and the appendix examines them. Whether a firm connects as an input supplier or as a distributor is usually a background fact, but the spillover evidence runs through backward linkages (Javorcik, 2004; Orlic et al., 2018). Whether that logic transfers depends on whether the buyer can yet say what it would buy. Where a grade or a variety exists, the incentive to transmit capability follows. Where the input is a new recycled or regenerative material whose grade the buyer has not defined, it does not, however far the supplier has taken quality (Gall et al., 2020). The condition `DOWNSTREAM` measures the distinction and Appendix F.2 tests it.

What happens once a relationship exists also bears on how it should be counted. Sellers cut marginal relationships first under a shock (Macchiavello and Morjaria, 2015), and SMEs are the more exposed party (Juergensen et al., 2020; Asgary et al., 2020). Supplier performance nonetheless does not deteriorate on average after a multinational relationship ends, which suggests that firms retain the capabilities they built (Masso and Vahter, 2023; Alfaro-Ureña et al., 2022). If that is right, a lapsed relationship is not necessarily a failure, and a snapshot count of current integration is a poor summary of what a programme achieved. That is why our outcome keeps “integrated then lapsed” apart from “never integrated”, and why H3 expects programme benefits for enterprises whose relationship never formed.

2.4 What the buyer decides: governance, relational contracting and internal decision-making

The literature treats the buyer’s side of the relationship more thinly and more abstractly. The governance tradition (Gereffi et al., 2005) characterises the lead firm’s relationship to its chain along a spectrum from market to captive and hierarchical forms, and the typology helps to predict what a supplier can learn and

how easily it can exit (Arora et al., 2021; Keijser et al., 2021). The relational-contracting literature is more precise about how ties acquire value. Trade volume is constrained by the value of the relationship, and that value rises with its age (Macchiavello and Morjaria, 2015). Relational capital also operates at entry, as well as in maturity, through the institutional bridges that match the right firms (Crescenzi and Harman, 2023; Abe, 2015; Hennart, 2020). The base-of-the-pyramid literature adds a distinction we will need. Including small informal suppliers in a formal chain is a paradox that direct supplier development can resolve and indirect, intermediated development cannot (Brix-Asala et al., 2021).

What the value-chain literature largely brackets is that the lead firm is not a single actor. Decisions about a supplier are divided between a head office and country businesses and between a sustainability strategy and monthly commercial targets, and the people who hold a relationship leave. Li and Cuervo-Cazurra (2024) come closest, in an account in which the multinational’s internal organisation determines what happens to suppliers, whatever its external power.

These internal organisational processes have been studied in the operations and supply-chain literature. Sustainability requirements fail to cascade through a multinational’s supply network when its procurement unit is not engaged (Villena, 2018). Purchasing and sustainability managers make sense of emerging-market suppliers differently (Xiao et al., 2018). Relationship quality is carried by individual boundary spanners and, in emerging markets, by interpersonal ties at lower organisational levels (Zhang et al., 2011; Huang et al., 2016). Gains come from supplying any “superstar” firm, whether or not it is a multinational, so that what matters is whether the connection is commercially serious (Amiti et al., 2024).

We therefore built the interview instrument to ask who in the buyer was committed and at what level, as well as whether the buyer was committed at all (Supplementary Appendix S1, Table S1.4). This strand is the origin of the localisation corollary of H1, of the distinction between a local team rated as supportive and a local team that has already acted, and of the treatment-at-the-margin claim that Section 2.7 sets aside as untestable on our data.

2.5 Impact credentials: environmental upgrading and the social enterprise

A distinct literature asks what changes when the integration in question is environmental, and three of its findings bear on this paper. Suppliers absorb the costs of environmental upgrading while lead firms capture its returns (Ponte, 2022b). Environmental and economic upgrading are distinct and both are located in chain governance (Krishnan et al., 2023). Supplier upgrading is driven principally by anticipated buyer requirements, which leaves suppliers exposed if the demand does not materialise (De Marchi and Di Maria, 2019). Unilever’s own sustainable tea sourcing is the canonical illustration, with Rainforest Alliance certification becoming a condition of market access (Archer and Elliott, 2021). The drivers of greening sit largely outside the supplier’s control, in the lead firm’s emissions accounting and reputational exposure (Steenbergen and Saurav, 2023; Jensen and Whitfield, 2022; Brucal et al., 2019).

The other half of this portfolio was funded for social impact, and the smaller literature on procuring from social enterprises reaches a similar conclusion from the buyer’s side. A multinational can hold a social logic and a market logic together in an inclusive supply chain, but the market logic dominates and the coexistence depends on specific mechanisms inside the firm (Bruton et al., 2025). Procurement functions that buy from social enterprises do so through individual relationships, trust and supplier development that mitigate the tension between the two logics, often with a third party in the relationship (Meqdadi et al., 2020). Social enterprises themselves trade power for impact in their chain relationships and spend heavily on the social capital that keeps them viable (Taylor and Rosca, 2023; Tang, 2018).

Whether the supplier’s credential is environmental or social, then, the literature expects the commercial decision to be taken on commercial terms, and our green and social contrasts test that expectation. To our knowledge this literature contains no measured comparison of integration rates between green and social suppliers, meaning those selected for their environmental and for their social impact, within a single buyer’s portfolio. It studies what happens to green suppliers once connected, and much less whether they connect.

If lead-firm demand governs green integration more than supplier upgrading does, as Crescenzi and Harman (2026) discuss, the relevant condition is the supplier’s alignment with what the buyer is currently trying to buy. That changes as corporate sustainability priorities shift with leadership and strategy. This reasoning yields the credential corollary of H2 and two conditions. GREEN records pre-existing credentials at entry, and PRIORITY_FIT records alignment with the buyer’s prevailing sustainability priority.

2.6 From literature to instrument

Table 1 summarises how each strand maps onto the hypotheses that follow and onto the conditions of the configurational analysis, and Supplementary Appendix S1 gives the interview questions each strand generated. The main results use five conditions. Five more, in parentheses, are defined in Appendix C and enter only the robustness specifications. The evidence base is far richer on what integration does than on how it comes about, and our data address how it comes about. It is also far richer on the supplier’s constraints than on the buyer’s decision, and our results concern the buyer’s decision.

Table 1: From literature to instrument

Theme (Section)	Hypothesis	QCA conditions (robustness conditions in parentheses)
Returns to a buyer link (2.1)	H3; monitoring-data test (Appendix F.3)	—
Selection vs. treatment (2.2)	H1, H2, H3	UL_LINK_START, TIE_PRIOR
What the supplier brings, and afterwards (2.3)	H2; H1 (tenure); direction (Appendix F.2)	DOWNSTREAM (SCALE, CERTIFIED, CODIFIED)
What the buyer decides (2.4)	H2; H1 corollary (localisation)	PRIORITY_FIT (GOVERNANCE, CODIFIED)
Impact credentials (2.5)	H2 corollary (green and social)	GREEN, PRIORITY_FIT (CERTIFIED, GREEN_SECTOR)

Notes: the main conditions are defined in Section 4.2.2 and the robustness conditions in Appendix C. SCALE is a demonstrated route to volume at application; CODIFIED a specification for the transacted input already in existence; CERTIFIED a named third-party certification held at entry; GREEN_SECTOR a core activity in a recognised green sector at entry; GOVERNANCE the governance mode at the Unilever interface. Supplementary Appendix S1 pairs each evidence item with the interview question it became.

2.7 Hypotheses

We organise the empirical strategy around one question from the comparative literature. Does post-project integration reflect selection, the programme connecting with firms that are already a good fit, or treatment, the programme and the relationship building capability that was not there before? Experimental studies of buyer access and management support find treatment effects (Atkin et al., 2017; Bloom et al., 2013; Alfaro-Ureña et al., 2022), and structural studies of large-buyer entry emphasise sorting (Iacovone et al., 2015). We frame the hypotheses as selection expectations. The treatment account has its own observable implications in our data, and we examine each of them. They are an enterprise whose project the buyer’s side saw no use for at the start and that then integrated, within-project engagement adding to the match at entry as a predictor, time in the programme predicting integration, and benefits reported by enterprises that never integrated. We cannot observe relationship building, because the record of what happened during a project is the programme team’s retrospective assessment (Section 3.1). Three hypotheses follow from the review, and Section 7 assesses each.

H1 (Selection at the gate.) Integration is decided by what an enterprise brings to the programme and, specifically, by whether the buyer’s side saw a use for what it proposed when it entered. The programme team’s record rates each project, after the event, on whether the proposed work matched a Unilever business or sustainability objective and how strong that potential match was. We call this the match at entry. It is a different thing from a prior relationship with the buyer, which the same record holds separately as whether the enterprise was known to someone in the business, or already working with it, when its first project was approved. It is also different from how far Unilever’s local team supported the project, which the tracker rates as local buy-in at start. We expect the match to act as a minimum threshold. Enterprises whose project offered Unilever nothing it could use do not integrate, and neither a stronger match above that threshold nor time in the programme substitutes for one that was absent at the start. That expectation runs against the relational-contracting account, in which the value of a buyer–supplier relationship accumulates with its age (Macchiavello and Morjaria, 2015; Iacovone et al., 2015). A prior relationship, where one exists, is expected to raise the probability of integration above

the threshold without being necessary for it. As a corollary, an enterprise put forward by someone in a Unilever market should integrate more often than one that arrived by a central route or an open call, because such a referral records that somebody in the country business had already seen something in the enterprise it could use. The routes into the programme varied over the decade, from the partners' own networks to open challenge calls shortlisted locally, so the corollary concerns the origin of a referral, whatever the selection procedure.

H2 (The buyer decides.) Past the gate, the decision sits with the buyer and not with the seller. Among enterprises the buyer had matched to a use at entry, no combination of what the enterprise brings, as the administrative and application record holds it at entry, is sufficient for integration. The condition set describes the seller's side of the transaction as it stood at entry. It covers what the enterprise offered, where in the chain it sat, what credentials it held, how well its theme fitted what the buyer said it wanted and, on the interviewed cases, what reach its founder had. The enabling-conditions literature is premised on characteristics of this kind determining participation (Pleticha, 2021; Abe, 2015; Cohen and Levinthal, 1990), and we test that premise. Where it fails, the governance typology suggests a reason. Of the three determinants of governance form, only supplier capability is an attribute of the supplier, and the other two describe the transaction as the buyer defines it (Gereffi et al., 2005). We then use the interviews to ask what the buyer's side of the same transaction looked like and whether the deciding conditions were on that side. A corollary follows for the credentials the programme selects on. An environmental or social credential marks entry to the programme and should not, on its own, raise the probability that the buyer trades with the enterprise, because the decision to buy is taken on other terms (Crescenzi and Harman, 2026; Bruton et al., 2025).

H3 (Participation produces benefits that integration does not measure.) What the programme produces for an enterprise, such as a capability, credibility with other buyers or access to other markets, is separable from whether the buyer integrates it. We therefore expect to observe such benefits among enterprises that never integrated, as the accelerator literature finds for ventures selected on fit with their parent (Shankar and Shepherd, 2019; Cohen et al., 2019) and the durability literature finds for suppliers whose relationship ended (Masso and Vahter, 2023). We test the hypothesis on the interviewed cases, where we can also ask how often the stricter claim holds, that something would not have existed without the programme.

We examined two further expectations and carry them in the appendix, because their results do not change what the paper concludes. We tested the green corollary of H2 as a hypothesis in its own right, on three measures of green across the three samples, with the credential separated from sector membership and certification. The result is a null on every basis, reported in full in Appendix F.1 and summarised in Section 5.7. Backward-linkage theory predicts an advantage for enterprises that supply inputs over those that distribute finished goods (Javorcik, 2004). Chain direction is a null with an unstable sign, and Appendix F.2 reports it. A further set of conditions the literature suggests, certification, green-sector membership, a codified specification, a route to scale and the governance mode, enters only the robustness specifications of Appendix C. Two predictions the literature invites cannot be tested here. The first is whether within-project engagement builds a link that was absent, since the tracker's within-project fields record, after the event, what the buyer was doing about an enterprise, with no separate record of programme effort (Section 5.6). The second is whether integration is a stepping stone to other buyers, which the monitoring data allow us to describe but not to test (Appendix F.3).

3 Setting and data

3.1 The programme and its records

TRANSFORM began in 2015 as a partnership between Unilever and the UK’s Department for International Development, and EY joined as a core partner in 2020. It ran in two generations, which the programme labels T1 (2015–2019) and T2 (2020–2026), and closed in 2026. It grew out of a 2014 letter of intent between the Secretary of State for International Development and Unilever’s chief executive, which committed the two to a joint initiative on social business models (DFID and Unilever, 2014). At its launch a year later the language was of using the partners’ “resources, networks and know-how to boost social enterprises and change behaviour across Africa and Asia” (TRANSFORM, 2015). Its stated aim is to support enterprises whose products and services improve health, livelihoods or the environment for low-income households in sub-Saharan Africa and South Asia. Integration into Unilever’s supply and distribution chains appears in none of those statements. In the first generation it was a spillover benefit the partners welcomed. In the second it became an informal target, visible in the regional challenge calls’ statement that Unilever “is interested in how impact enterprises can provide new routes to market, become suppliers, incubate innovation and create purpose-driven brand stories” (TRANSFORM, 2024), and it never became a formal target.

The programme is a partnership between its three members, with joint governance, as its own account describes. Consensus among the partners was needed for any core change, and on the selection of enterprises “EY would act as an advisor while Unilever and the FCDO would vote because they were the cash-contributing partners” (Unilever, FCDO and EY, 2026, p. 27). Decisions about which enterprises to fund were joint, and the development partner decided some of what happened to an enterprise. Throughout the paper we distinguish two actors. The programme team is the joint FCDO–Unilever–EY body that selected and supported enterprises and compiled the records we use. The buyer is Unilever’s commercial business, which decided whether to trade with them.

The programme provides a package of support. A grant funds a defined project of one to two years. The median award recorded in the tracker is about £212,400. Projects selected by the global panel, and those a market team referred up to it, were funded at about that level. The regional panels chose smaller awards, about £100,000, so as to fund more projects. A Unilever project lead is assigned to the enterprise for the life of the project. The programme describes the lead as the enterprise’s champion over twelve to eighteen months (Unilever, FCDO and EY, 2026), and the lead is the person through whom any commercial relationship with Unilever’s country business would be brokered. EY provides business coaching on request, which not every enterprise took up, and for some enterprises a block of consulting on financial modelling, investment readiness or sales. The programme’s support hub adds learning modules and introductions.

Selection moved over the decade from sourcing through the partners’ own networks to regionally led open challenges, in which Unilever’s market teams, EY and the British High Commissions shortlist enterprises locally. That shift bears on the localisation result of Section 5.4. We treat this package of support as the programme treatment. Every enterprise the programme admits receives it, its intensity varies in ways the programme’s records do not capture systematically, and we observe no enterprise that applied and was refused. The comparisons below are therefore among treated enterprises.

The analysis uses four sources. The first is the programme’s confidential project-level evaluation database, which we call the tracker and which covers every enterprise funded from T1 through T2. It records each project’s phase and pillar and its referral route. It records the programme team’s retrospective rating of how strongly the proposed project matched a Unilever business or sustainability objective at project start, during the project and after it, which the tracker labels the strength of the Unilever link, and the team’s rating of the Unilever local team’s buy-in at project start. It records a current integration status and, in a column added in September 2026, whether the enterprise was known to someone in the business, or already working with it, when each project was approved. The programme team compiled the tracker in June 2025, from its knowledge of each relationship and from the project files, and updated it to March 2026. None of its ratings is a contemporaneous record, and the separation it draws between what was true at the start of a project and what happened during it is the team’s own, drawn after the event. That provenance matters for how the tracker is read, and Sections 3.2 and 5.2 return to it.

The second source is the project document corpus. It holds the application form, milestone reports and final report for every enterprise whose folder we could match to the tracker, 314 files across about 103 folders.

The applications were written at the time, by the enterprises, and in some cases with the programme team’s help. The third source is the monitoring data, a cross-section of cumulative totals for jobs and beneficiaries reported by each enterprise. The fourth is nineteen recorded and hand-coded interviews, twelve with founders and seven with Unilever and programme staff, which provide evidence on the processes behind the observed outcomes.

The tracker records 150 projects across 109 organisations. Removing the programme partners that were never candidate Unilever suppliers, and merging one organisation that appears under two names, leaves the 90 supported enterprises to which every tracker-based figure in this paper refers (Section 3.3; Appendix A gives the derivation row by row). We refer to enterprises throughout by a pseudonym made of a business-model label and a letter, and we hold the key. Table 2 lists the twelve interviewed enterprises under those labels, with their region, pillar family, chain direction and outcome. Some enterprises ran more than one project, and the two quantitative methods collapse their projects differently. The regression needs one observation per enterprise, so it keeps the project with the best outcome and reads the covariates from that project. The configurational analysis works at the level of the enterprise, so it takes the strongest link recorded at any project start and the earliest start to the latest end. A multi-project enterprise can therefore carry different values under the two constructions, which is why one comparison, the cohort contrast, points in opposite directions across them (Section 6). We estimate models on the complete cases of the population and state the estimation N with each model.

Table 2: The twelve interviewed enterprises

Enterprise	Region	Pillar family	Chain direction	Tracker status	Integrated
Agri A	East Africa	Livelihoods	Upstream	Integrated, later discontinued	Yes
Agri B	East Africa	Livelihoods	Upstream	Not integrated	No
Agri C	West Africa	Livelihoods	Upstream	Not integrated	Yes ^o
Agri D	South Asia	—	Upstream ^d	No tracker row	No ^o
Platform A	East Africa	Livelihoods	Downstream	Integrated	Yes
Platform B	East Africa	Livelihoods	Downstream	Not integrated	No
Platform C	East Africa	Livelihoods	Downstream	Not integrated	No
Recycling A	South Asia	Environmental	Upstream	Integrated, later discontinued	Yes
Recycling B	South Asia	—	Upstream ^d	No tracker row	No ^o
Refill A	South-East Asia	Environmental	Downstream	Not integrated	No
Refill B	South Asia	Energy ^p	Downstream	Integrated	Yes
Refill C	Southern Africa	Environmental	Downstream	Trying to integrate	No

Notes: enterprises are listed under the pseudonyms used throughout the paper, the first word of which is the business model; labels, regions and the tracker columns match the enterprise file of the replication deposit. Pillar family is the tracker’s classification: Environmental is the green side of the portfolio contrast of Section 5.7, and the other families are the social side, with energy counted among them. Chain direction is the research team’s coding (Section 3.3); the DOWNSTREAM condition of the configurational analysis is coded separately, from the application documents, and differs for Platform B (Appendix C). Tracker status is the March 2026 integration field, and an enterprise counts as integrated if the tracker records it as integrated, whether or not the relationship later lapsed. ^oOutcome set by the overrides of Section 3.2: Agri C supplied Unilever before it applied, and Agri D and Recycling B have no tracker row. ^dNo tracker row, so chain direction is read from the enterprise’s own documents. ^pThe tracker record for Refill B aggregates six projects run by the parent organisation and takes its pillar family from them; the interviewed case is the refill business.

3.2 The outcome, and how far the sources agree on it

The primary outcome is the tracker’s integration status as at March 2026, a four-category variable coded by the programme team: (0) not integrated, no formal Unilever commercial relationship; (1) trying to integrate, negotiations or pilots under way with no confirmed contract; (2) integrated but later discontinued, a formal supply or distribution relationship established and since lapsed; (3) integrated, an active formal commercial relationship as at March 2026. “Ever integrated” collapses the ladder to a binary, 1 for categories 2 and 3 and 0 otherwise, and an enterprise appearing in several tracker rows takes its maximum level. Our outcome is whether the enterprise was ever integrated. That is broader than the tracker’s own question, which is whether the TRANSFORM project produced an integration, and where the two diverge we use ours.

We code three enterprises by explicit override for that reason. We recode Agri C from 0 to 1, since it was already a major supplier of an agricultural commodity to Unilever when it applied, which the tracker does not count. We code Agri D and Recycling B, which have no tracker row, as 0, since their own accounts, and for Agri D the programme staff, describe attempts in progress and no commercial relationship. One further override concerns the buyer’s record of prior relationships (Section 3.1). That record shows Refill B’s parent organisation as having no prior relationship with Unilever on any of its six projects, while its own proposals describe five years of collaboration with Unilever teams and its founder’s history is a Unilever career (Section 5.3). We code it as already working with Unilever and report the programme’s coding as a sensitivity.

Three of the paper’s constructs are measured by more than one source, and the sources mostly agree. On the outcome, the founders’ own accounts of the interviewed enterprises agree with the tracker on ten of twelve under each of two alternative definitions. On green status and chain direction, documents and interviews agree on ten of twelve and nine of ten.

The entry measures are the exception, and the reason is that the two records measure different things. The tracker’s rating at project start is the team’s view of whether the proposed project matched a use inside Unilever. The enterprises’ applications record whether a relationship with Unilever already existed. The two coincide on 48 per cent of the 56 enterprises we hold both for. The disagreement runs one way, with a match rated where the documents show no relationship in 24 cases and the reverse in 5, which is the pattern a match rated on the project and a relationship recorded on the enterprise would produce. The buyer’s own record of prior relationships, added in 2026, asks the documents’ question and agrees with them on 69.6 per cent of the 46 enterprises that carry both (Appendix D.3). Two readings of the match rating itself remain open. It may record a real property of the project, a use the buyer’s side could see and the enterprise had no reason to document. It may instead record hindsight, a team rating the match in 2025 with the outcome in view. We keep the tracker as the outcome and report the enterprises’ contemporaneous record beside it wherever both exist. Sections 5.2 and 5.3 report the gate on the match and test a prior relationship on all three records.

3.3 Who is in the sample

Three decisions determine the population, each taken after an audit of the tracker against the document corpus. We exclude eighteen tracker rows that are programme partners, universities, implementing NGOs, research bodies and utilities, none of which was ever a candidate Unilever supplier. We merge one enterprise that appears under two spellings of its name. We code an enterprise with a blank in the March 2026 integration field as not integrated, which is what the programme lead confirmed a blank records, and 35 enterprises are in that position. Appendix A lists the partners and records the one unresolved grouping decision. It also describes a further variable constructed for this study, which the programme does not record, each enterprise’s chain direction, upstream as a supplier of inputs or downstream as a distributor of Unilever’s products. Of the 90 enterprises, 39 are upstream and 51 downstream.

3.4 The document corpus

The corpus is the enterprises’ own record of themselves and their projects, and we use it in two ways. The application form and proposal say what the enterprise was when it applied, covering its business, its existing relationship with Unilever if any, its credentials and its stated route into Unilever’s chain. They are the source for every configurational condition. We code the conditions from application-stage documents only, so that a condition cannot be satisfied by evidence of what happened afterwards (Section 4.2.3). The applications are contemporaneous where the tracker is retrospective, and they were written without knowledge of the outcome. They are not independent of the programme, since the programme team sometimes worked through a proposal with an enterprise before it was submitted, so they record the enterprise’s account of itself as it stood when the project was agreed. The milestone and final reports say what the project did and what the enterprise reported delivering, which is the treatment as delivered in the terms of Section 3.1. We use them descriptively, for the form each relationship took (Section 5.8), and they would be the source for extending the participation results beyond the interviewed cases. The corpus holds an application-stage document for 56 of the 90 enterprises, and the gap varies with cohort (Section 8).

3.5 The interviews

We conducted nineteen semi-structured interviews, twelve with founders or senior staff of supported enterprises and seven with Unilever staff involved in TRANSFORM, who were pillar leads, local procurement and sustainability contacts and programme managers. Introductions to founders came through the programme team, so the enterprises we could interview are those with which the programme still had a working contact, and Section 4.2.1 describes what that selects on. All twelve founder interviews attach to enterprises in the deep tier, which Table 2 lists with their region, pillar family, chain direction and outcome. Ten of the twelve have tracker rows, Agri D and Recycling B do not, and Section 4.2.1 describes how they are used.

The schedule follows a lifecycle-first “grand tour” structure (Supplementary Appendix S1). The founder narrates the business and its financing before being asked when TRANSFORM entered and what it offered that other funders did not, so we hear the firm’s own account before the interviewer’s frame is imposed. Targeted questions then tracked the firm’s journey against the literature of Section 2, and later interviews moved from open elicitation toward probing the study’s own propositions. We hand-coded the verbatim transcripts in NVivo against a fifty-code codebook built from the same literature themes that generated the questions and the configurational conditions (Table 1). Several conditions, the interview-coded one in particular, were settled while the interviews were under way. We refer to interviews by respondent role. They enter the analysis in four ways, as the source of one configurational condition, as two alternative definitions of the outcome, as the coded record of what participation produced, and as the mechanism evidence against which we read the configurational results (Appendix E).

3.6 Three samples

The analysis works at three levels over the same population. Table 3 lists the samples once, and every later statement of an N refers back to it. The samples are nested because each deeper one requires more evidence about an enterprise, first a tracker row, then a matched application document, then a founder interview. Each is the widest sample on which its kind of claim can be made, which is why the paper reports three and why not every result appears on every tier. The nesting has one exception. Agri D was interviewed and its documents read, but it has neither a tracker row nor a matched folder, so it enters the interviewed cases by the override of Section 3.2 alone. Ten of the twelve sit inside the document-coded 56. The interviewed cases carry one founder interview each, and that one-to-one correspondence is what lets an interview code serve as a condition and an interview account redefine the outcome. The seven buyer-side interviews attach to no enterprise. They are evidence about the buyer’s organisation, and the only evidence in the study from inside it. The three samples also differ in where their conditions are read from. The portfolio sample reads them from the buyer’s tracker, the document-coded sample from the enterprises’ own documents, and the interviewed cases from those documents together with the founders’ own accounts. In that order the evidence becomes progressively more independent of Unilever’s judgement.

Table 3: The three samples, and the interviews

Sample	N	Integ.	Who is in it	Conditions from	Used for
Portfolio sample	90	14	every supported enterprise, all project types; programme partners excluded	the tracker	the regression; necessity and sufficiency at portfolio scale; the during-project, cohort, direction and referral comparisons; the monitoring-data comparison
Document-coded sample	56	10	the enterprises with a matched application document	their own application documents	sufficiency at scale on an independent coding; the green contrast; agreement with the tracker
Interviewed cases	12	5	the enterprises closely read and with one founder interview each	application documents and founder transcripts	the configurational analysis with an interview-coded condition; the alternative outcome definitions; what participation produced; the cases the QCA does not explain
<i>Alongside all three.</i>					
Buyer-side interviews	7	—	pillar leads, procurement and sustainability contacts, programme managers	—	how selection and commercial sign-off work inside the buyer, and why relationships lapse

Notes: the portfolio sample is the population of Section 3.3, with the Agri C override of Section 3.2. The regression and the configurational analysis share that population and differ only in how they collapse an enterprise’s several projects; models are estimated on its complete cases, and the estimation N is stated with each model. The nineteen interviews carry 415 coded references against a fifty-code book, 268 from the twelve deep-tier founders and 147 from the staff; all nineteen are used in the case analysis, the 21 references of the Recycling B transcript included (Appendix E).

4 Empirical strategy

4.1 Three tiers and an interview strand

We approach the integration question on three nested samples of the same TRANSFORM portfolio, with an interview strand running alongside all three. Each sample supports a kind of claim the others cannot, and Appendix A.1 records where the design departs from the project’s inception report. On the portfolio sample, the regression asks which starting characteristics are on average associated with integration. The configurational analysis puts the same tracker material into enterprise-level sets, and from them it produces the necessity result, the sufficiency null and the during-project gate. The document-coded tier applies the coding pipeline of Section 4.2.3 to the 56 enterprises whose document folder we can match, so that we read their conditions from their own applications without reference to Unilever’s assessments. The interviewed cases are the twelve enterprises whose record we have read closely and whose founders we have interviewed, one interview each. It is the only tier at which claims about combinations of conditions can responsibly be made, and the only one on which an interview code can serve as a condition. The interview strand describes process. It is the only strand with evidence on the buyer’s own decision-making, which the portfolio data record only as an outcome. Because no combination of conditions is sufficient at portfolio scale (Section 5.5), we use the interviews to explain what happens after the entry threshold.

Qualitative Comparative Analysis is set-theoretic in procedure but rests on close case knowledge (Ragin, 2008; Schneider and Wagemann, 2012). We therefore treat the two configurational tiers and the interviews together as the case-based component of the design, and the regression as the variable-based one. The design is multiphase in the terms of Creswell and Plano Clark (2011). Early interviews helped develop the constructs and conditions, later interviews explained and tested the patterns the structured analyses produced, and the sequencing followed the phased release of Unilever’s data as much as any prior plan (Appendix A, Figure 2).

No single measure records entry-stage selection. We therefore observe it through several related but distinct measures (Section 3.2). The tracker’s match at entry, rated retrospectively by the programme team, records whether the proposed project matched a business or sustainability use inside Unilever. A prior relationship between the enterprise and Unilever, meaning that the enterprise was already known to the buyer or already working with it before the project, is recorded separately, in the tracker’s prior-relationship record and in the enterprises’ application documents (UL_LINK_START). TIE_PRIOR, coded from the founder interviews, records the founder’s prior reach into the buyer’s world, and it is broader than a prior enterprise relationship. Two further constructs are each read from more than one source. Chain direction is read from the research team’s public-source coding, the documents (DOWNSTREAM) and the interview codes for integration mode. Green status is read from the buyer’s pillar classification, the documents’ GREEN condition and the interview codes. The outcome is the tracker’s, and on the interviewed cases the interviews provide a broader and a narrower alternative. When sources disagree we report the disagreement and interpret it. Because the match at entry, the prior relationship and the founder’s reach are three distinct constructs, differences among them are not measurement disagreement over whether a single link existed at entry. The relationship among them is itself informative. Where founders and buyer staff give different accounts of the same failure, we report both without privileging either. The interviews enter the analysis in the four ways listed in Section 3.5. The last of these, the mechanisms behind the cases the QCA does not explain, is what set-theoretic multi-method research identifies as the deviant cases to trace (Schneider and Rohlfing, 2013, 2019). The strands are nested, since ten of the twelve interviewed enterprises sit inside the document-coded tier, which is drawn from the portfolio, so agreement across strands is evidence of coherence across levels of observation.

4.2 The configurational analysis

4.2.1 Case selection

We selected the interviewed cases on two criteria. The first is the nested comparative design of the inception report, with a core of circular-economy enterprises clustered by business model and a set of contrast cases. The core is a refill cluster (Refill A, Refill C, Refill B), chosen so that firms similar in model but divergent in outcome can be set against one another, a recycling case (Recycling A) and two agricultural-supply cases (Agri C, Agri D). The contrast cases (Platform A, Agri A, Platform B, Agri B, Platform C) establish whether

what the clusters reveal is specific to circular-economy enterprises. The second criterion is variation on the three dimensions the portfolio analysis identified as candidate predictors, referral source, local buy-in at entry and starting link strength. Selection is purposive on business model and on the conditions. The design sought variation on integration, and has it, since five of the twelve are integrated, but no enterprise was chosen because it had integrated or because it had not.

Interview access constrains the set. Four enterprises named in the inception design are not in it because no interview could be secured. Introductions to founders came through the programme team, so the enterprises we interviewed are those with which the programme still had a working contact, and that is a selection in its own right. An enterprise whose relationship ended badly, whose founder has moved on or which has nothing to say about Unilever is less likely to be in the set than one with a relationship to describe. Reachability is therefore not plausibly independent of how a relationship ended, and the sample is purposive within the limits of who was available.

How the interviewed cases differ from the rest of the portfolio bounds what the interview evidence can support. On the portfolio frame, the ten interviewed enterprises with a tracker row differ from the other eighty on two dimensions, both by design. Two-fifths of them ever integrated, against eleven per cent of the rest, and every one arrived with a Medium or High starting link, against seventy per cent of the rest. On chain position, multi-project status, pillar, cohort and region they are unremarkable. They under-represent the locally originated referral route, so the interviews cannot corroborate or contest the localisation result (Appendix E.2). Because the sample is selected on circular-economy criteria, and circular models are overwhelmingly upstream, its integration rates and its upstream configurations are properties of the selection, and we never cite them as portfolio evidence. Because selection was on business model while the GREEN condition is coded on credentials at entry, the two differ under coding for two of the twelve. Each of the twelve has at least one application-stage document, a founder interview and a known integration status as of March 2026, read from the tracker for nine and set by override for Agri C, Agri D and Recycling B (Section 3.2). The twelfth founder interview, with Recycling B, joined the analysed set after the other eleven were selected and adds a non-integrated case that changes no interpreted result (Appendix C.4). Appendix C, Table 13, gives the twelve cases with their codes.

4.2.2 Conditions

We selected the conditions against the literature of Section 2 and kept them to a small binary set, a constraint at this sample size (Marx and Dusa, 2011). Four are coded from the enterprises’ application-stage documents and one from the founder interviews. Five more, a demonstrated route to scale, a codified specification for the transacted input, third-party certification, green-sector membership at entry, and the governance mode at the Unilever interface, are defined in Appendix C. Each of them substitutes for a main condition in a robustness specification, and none changes the result.

UL_LINK_START captures whether a relationship between the SME and Unilever existed at entry, as the SME’s own documents record it, which corresponds to the captive or relational versus market distinction of Gereffi et al. (2005). It is 1 where the SME had a Unilever contact from outside the selection process, a draft or active supply or distribution arrangement, or a prior project relationship at entry. A pillar lead, reviewer or coach named on a proposal does not count, because such contacts were part of selection.

DOWNSTREAM captures the SME’s stated pathway into the chain, coded from proposal language at application. It is 1 where the SME pitched on-selling, distributing or retailing Unilever-branded products, and 0 where it pitched as an input supplier or a non-distribution innovator. The regression takes upstream as its reference category and reports a coefficient on downstream, so the two describe the same contrast in the same direction.

PRIORITY_FIT captures alignment between the SME’s stated impact theme and Unilever’s prevailing strategic priority at entry. We read it from a purpose-built priority-intensity matrix, scored 0–3 by year and pillar, which tracks the shifting flagship priorities across the Polman, Jope and Schumacher chief-executive regimes. The condition is 1 only where the intensity in the entry year reaches 3.

GREEN captures pre-existing green credentials at entry: 1 where the SME held recognised environmental certification, was founded on an explicit sustainability mission, or operated in a recognised green sector before entry.

TIE_PRIOR is the one condition coded from the interviews. It records whether the founder arrived

with an international network through which a buyer could be reached. It measures the same moment as `UL_LINK_START`, the documented relationship with Unilever at application, in a different way. It is 1 where the founder’s transcript has the code for an internationally networked founder or a founder of Global North origin, or where the founder had worked inside the buyer, following Filatotchev et al. (2009), Drori et al. (2009) and Colonnelli et al. (2026). It exists only where a founder was interviewed. It is more exposed to hindsight than the document-coded conditions, because it is coded in 2026 about entry decisions taken years earlier. It never enters a model alongside `UL_LINK_START`, for which it substitutes.

4.2.3 Coding conditions from documents

We produce the condition codes through a three-stage pipeline that follows the LLM-assisted deductive-coding approach of Xiao et al. (2023). In the first stage, regular expressions surface candidate evidence passages from each firm’s documents into an auditable per-firm workbook. In the second, a large language model (Claude Sonnet 4.6, at temperature 0, with the full exchange archived) proposes a binary code for one firm and one condition at a time, against that evidence alone and a calibration rubric, and returns a confidence rating, its reasoning and the decisive quote. In the third, a human coder verifies the code against the source documents and records any overrides. The evidence passed to the model is restricted to application-stage documents, so that a condition defined at the point of application cannot be satisfied by evidence of what happened afterwards. We also coded the same enterprises from their whole folders, and Section 6 reports what the restriction changed, which is a measure of how much hindsight enters when the whole record is used.

On the twelve interviewed enterprises, a second member of the research team checked the model’s codes against the source documents. The check took every cell the model returned at low confidence, 19 of the 96, and a random sample of 35 of the remaining 77, 54 cells in all. It confirmed 48 of the 54 codes and revised six, an agreement of 88.9 per cent. The revisions move the solution terms and coverage of several robustness specifications but leave every necessity conclusion, the sufficiency null and every hypothesis verdict unchanged.²

4.2.4 Analytical strategy

We apply crisp-set QCA using the QCA package in R (Dusa, 2019), in which every firm is either fully in or fully out of each condition and of the outcome. A condition is necessary if the outcome almost never occurs without it, and sufficient if firms that possess it are nearly all integrated. A condition can be one without being the other. We judge necessity against a consistency threshold of 0.90 and sufficiency against 0.80, where consistency is the share of cases that behave as the claim says they should. We report each necessity result beside its triviality, the share of all cases the condition covers, since a condition present everywhere is trivially necessary. We extract sufficient combinations by Boolean minimisation (Ragin, 2008). Each combination the minimisation returns is a term, the set of terms is the solution, and we report each solution’s coverage, the share of integrated firms it reaches. Even three yes-or-no conditions can combine in eight ways, and twelve cases fill only six of them, so every model on the interviewed cases uses three conditions at most. We report the parsimonious solution, which makes whichever assumption about the unobserved combinations gives the simplest result.

The choice of a configurational method is more than a concession to sample size. The phenomenon is one in which several conditions are believed to operate together, and in which removing any single one leaves most of the outcome standing. A method that reports the average contribution of each condition separately cannot express that. We analyse non-integration alongside integration (Schneider and Wagemann, 2012), and test robustness to the condition set and to the definition of integration in Section 6 and Appendix C.

The document-coded tier runs the same procedure at scale, and it allows two things the interviewed cases cannot. The 56 enterprises are coded on the same five conditions by the same pipeline, so the tier gives the integration rate for each condition on its own, on a coding independent of the programme team. It

²The check was not blind, since the verifier could see the model’s code, so it measures agreement with the source documents and does not yield an inter-rater statistic. The codes that enter the analysis are the model’s, checked in this way, together with three outcome overrides and two DOWNSTREAM rulings made by the researchers and recorded in Table 13; Section 6 reports what happens when the low-confidence cells are treated as unknown. Appendix B gives the stages in full, worked examples and the pipeline’s limits.

also lets a four-condition specification run, since 56 cases populate the sixteen combinations well enough for minimisation to be meaningful. We run every specification built from document-coded conditions on both tiers, with the same thresholds, and we report each solution with the enterprises it covers, so that a term resting on a single enterprise is visible as such. The tier’s limit is the outcome. Only the tracker’s status is available for it, so the document-coded results test whether the enterprises’ own account of themselves at entry sorts the programme team’s outcome. The interviewed cases allow what the document-coded tier cannot, since each founder describes the relationship in their own terms and the outcome can be redefined there. Section 5.8 sets out three definitions, the tracker’s, the enterprise’s own broader account, and a strict account that counts only a direct commercial relationship. We re-run every deep-tier specification under each of the three (Appendix C), so that we can test whether the results depend on the definition of the outcome.

4.3 The regression

The regression is associational and portfolio-wide. The QCA identifies which combinations of conditions are sufficient for integration in closely observed cases. The regression estimates which individual characteristics are, on average, associated with a firm ending up integrated. We estimate it on the portfolio sample of Table 3, with one best-outcome project row per enterprise. The core specification is a logit for ever-integration on what was fixed before the project began, to which the within-project measures are then added:

$$\Pr(\text{INT}_i = 1) = \Lambda(\alpha + \beta \text{LINK}_i + \beta_b \text{BUYIN}_i + \beta_d \text{DOWN}_i [+ \delta_1 \text{ENGAGE}_i + \delta_2 \text{LINKDUR}_i]), \quad (1)$$

where INT_i is ever-integration and DOWN_i is 1 for a downstream distributor.

The equation has two blocks, and the distinction between them is the distinction H1 rests on. The first block holds what was settled before the project began, at selection. LINK_i is the tracker’s rating of the strength of the Unilever link at project start, which is the team’s assessment of how well the proposed project matched a use inside Unilever (the match at entry, Section 3.1). BUYIN_i is its rating of the Unilever local team’s buy-in at project start, which is how far the country business supported the project. They measure different things, and the paper keeps the tracker’s names for them throughout, the link at start and local buy-in at start. The second block, in brackets, holds what the tracker records of the same link during the project, after selection. LINKDUR_i is the strength of the link during delivery and ENGAGE_i the engagement of Unilever’s local team. All four are recorded as None, Low, Medium or High and scored 0 to 3, and they are graded for 90 and 85 of the portfolio respectively (Appendix D, Table 19). We test whether the within-project block improves the fit once the pre-selection block is in, which H1 predicts it will not.

Our preferred specification, Model P, is the binary logit that retains only the pre-selection block, cohort, local buy-in at start and link strength at start. We drop the within-project block because its two measures are recorded over the window in which integration happens, so they cannot be treated as pre-project predictors of it (Appendix D.2). We check Model P in four ways, all reported in Appendix D.3. We compare it with the model that includes the within-project block, by a likelihood-ratio test and the Akaike criterion. We re-estimate it with a Firth penalty, and on its own complete cases. We also re-estimate it with the enterprises’ document-coded relationship in place of the tracker’s rating, and with the buyer’s record of a prior relationship added to the pre-selection block, since the match and a prior relationship are different things and the paper tests both (Sections 5.2 and 5.3). The tracker records integration as four steps, not integrated, trying, integrated then lapsed, and integrated. Appendix D also reports ordered logistic models on all four steps (McCullagh, 1980; Venables and Ripley, 2002), in four specifications, each also estimated on the yes-or-no outcome. The gate result, that no enterprise below the threshold integrated, is a two-way table and does not depend on any of these choices.

We report the mechanism tests, on referral origin, starting-link level and time in the programme, as cross-tabulations with Fisher’s exact test, which computes the probability of a table at least as uneven as the observed one under independence and is exact however small the cells. Where the cells permit, we add a penalised logit, since with 14 events the events-per-variable budget does not support a multivariate mechanism model. All regression results are exploratory and suggestive of pattern.

5 Results

We report the results by claim. Section 5.1 gives the integration rate and the lapses. Sections 5.2 to 5.4 take H1 in three parts, the match at entry, a prior relationship with the buyer, and where the matches that held were made. Sections 5.5 and 5.6 take H2, first the finding that no combination of enterprise characteristics is sufficient for integration and then the buyer’s decision as the interviews describe it. Section 5.7 takes the corollary of H2 on impact credentials. The section closes with the definition of integration (Section 5.8), its association with reported impact (Section 5.9) and, for H3, the benefits of participation for enterprises that did not integrate (Section 5.10). Each subsection uses all the evidence that bears on it, and where two sources measure the same construct differently we report the disagreement. The samples are those of Table 3.

5.1 Integration rates and lapses

Table 4 reports the distribution of integration outcomes across the 90 enterprises of the portfolio sample as at March 2026.

Table 4: Distribution of Integration Status (March 2026, portfolio sample, $N = 90$)

Integration Status	<i>N</i>	%
Not integrated	71	78.9
Trying to integrate	5	5.6
Integrated but later discontinued	7	7.8
Integrated	7	7.8
<i>Ever integrated</i>	<i>14</i>	<i>15.6</i>

Just under four-fifths of enterprises (78.9 per cent) have not established a formal commercial relationship with Unilever. Fourteen, one in six, integrated at some point. Exactly half of those no longer maintain an active relationship as at March 2026, which leaves 7.8 per cent currently integrated. Whether one in six is a low rate depends on the benchmark, and there is no established one. Integration was an intention of the programme’s partners and never a target. We know of no published figure for the share of an impact accelerator’s ventures, or of a public–private supplier-development programme’s enterprises, that reach a commercial relationship with the corporate partner. The rate is lower than an analysis confined to the enterprises Unilever had explicitly assessed would report, because most first-generation enterprises carry a blank status, which records non-integration (Section 3.3). The two readings share the same count of integrations and differ only in the denominator.

The lapses deserve attention, because half of the ever-integrated relationships had ended by March 2026 and neither the tracker nor the document record gives the reason. Where the interviews allow the reason for a lapse to be reconstructed, founders and Unilever staff point to the same class of cause, decisions or changes inside Unilever, and neither points to deterioration in the supplier’s performance. Founders give more specific accounts than staff. One upstream recycling enterprise’s material reached a flagship pack, and the relationship produced no repeat order. The founder’s first explanation is that the two individuals who held the relationship inside the buyer had both moved on, so that “there is no one holding the torch for me” within Unilever. The founder’s other explanations are a corporate reweighting from sustainability to earnings under new leadership, and the retail underperformance of the product the material went into. An upstream agricultural supplier describes personnel churn removing the relationship owner, after which nobody enforced the sourcing arrangement downward and “the co-packers are a little bit more free to do their own stuff.”

Unilever and programme staff describe the same cause at portfolio level. A manager says of enterprises whose theme was cut from the focused pillars that they “got impacted, and then the integration then gets impacted, because it’s no longer . . . relevant to Unilever’s strategy.” Staff could not give a case-level account. Asked why a repeat order never came, a regional programme manager says “we don’t know, right, what didn’t really work,” and a senior programme manager offers two reasons without choosing between them, a restructure and “a commercial decision on, kind of, price.” That reflects what the people we interviewed knew. The commercial team that placed the order was not among them, so it is not necessarily what the

buyer as an organisation knows. It is consistent, though, with what another senior programme manager says of the outcome, that “if you look at our log frame, we don’t even count integration, it’s not even in there”. Integration was not a monitored indicator, so the programme kept no record of why a relationship ended. That is one reason the two sides can hold different views of whether an enterprise was integrated.

5.2 The gate at entry

Enterprises whose project the programme team saw no use for at entry almost never integrate. The tracker rates each project, after the event, on how well it matched a Unilever business or sustainability objective. None of the 24 enterprises rated None or Low ever integrated, against 14 of the 66 rated Medium or High (Fisher exact $p = 0.017$; it is the appropriate test when a cell is empty, as one is here). The result is a gate. Above it, Medium and High matches integrate at 10.5 per cent (2 of 19) and 25.5 per cent (12 of 47), a difference a penalised logit cannot distinguish from zero ($p = 0.201$). The operative distinction is therefore between a project the buyer’s side could use and one it could not. Three of the 24 enterprises below the gate were rated Medium or High during their project, so the team did see a use emerge in a few cases. None of the three integrated.

The regression and the configurational analysis give the same result on the same table. A Firth-penalised logit, which corrects the bias an ordinary logit shows when events are few and a cell is empty, returns a coefficient on the match at start of +2.61 with a 95 per cent interval of [+0.52, +7.47] ($p = 0.009$). In the set-theoretic reading, the match is necessary for integration, since no enterprise without one integrated (consistency 1.000). It is a weak predictor among the 66 enterprises that have it, of which 14 integrate (coverage 0.212). Neither reading depends on how an enterprise’s several projects are collapsed. On the regression’s one-row-per-enterprise construction the split is 26 below and 64 above, and the two constructions differ on two non-integrated enterprises. We use the portfolio construction throughout because the necessity statistics are computed on it.

The paper’s preferred specification, Model P, retains only what was fixed before the project began. The reason is that the tracker’s four ratings of the link measure one underlying quantity, and two of them are recorded over the window in which integration occurs (Appendix D.2). On Model P each step up the four-point match scale multiplies the odds of integration by about four and a half, and adding the during-project assessments does not improve fit (Appendix D.3, Table 23). That a match was recorded before the project began does not make it exogenous. Enterprises whose projects the buyer’s side could already use differ in ways we do not observe. The coefficient therefore identifies selection, which is what H1 asserts, and it cannot be read as what a match would do for an enterprise that lacks one.

The same file records what the match consisted of, as the team’s category for what a project offered Unilever. Projects that offered the commercial business a function it could buy or sell through, a route to market or a supplier of an input, integrated at 10 of 37 (27.0 per cent). Projects that offered progress toward a sustainability objective, a plastics-collection target or a refill model, integrated at 3 of 25 (12.0 per cent). None of the 12 projects offering behaviour change, a product test, staff development or a reputational benefit integrated. The categories are the team’s, applied after the event, and the cells are small, so the comparison is descriptive (Appendix D, Table 22). It is consistent with the account of the buyer’s decision in Section 5.6.

5.3 A prior relationship with the buyer

A prior relationship with the buyer is a different thing from a match, and the records measure it separately. The buyer’s own record states whether an enterprise was known to someone in the business, or already working with it, when its first project was approved (Section 3.1). It holds a value for 72 of the 90 enterprises, and Table 5 gives integration by that record. Enterprises with any prior relationship integrated at 9 of 25 (36.0 per cent), against 4 of 47 (8.5 per cent) for those without (Fisher exact $p = 0.008$). Those already working with Unilever integrated at 6 of 14, against 7 of 58 ($p = 0.015$).

A relationship raises the rate and is not a gate. It is present in 9 of the 13 integrators with a known value (consistency 0.692), and 4 enterprises integrated from no prior relationship at all. Every enterprise with a prior relationship was rated High on the match, so the relationship operates within the set the buyer’s side had already matched. Above the threshold, enterprises with a relationship integrated at 9 of 25, against 4 of 31 ($p = 0.058$). Because the two measures are nested, Model P cannot separate them as independent

predictors. Adding the relationship beside the match roughly halves the match coefficient, from +1.17 to +0.64. It returns a relationship coefficient of +1.00 $[-0.51, +2.69]$ ($p = 0.197$) on 71 enterprises and 12 events, so that neither is distinguishable from zero on its own (Appendix D.3, Table 21).

Table 5: Integration by Prior Relationship with Unilever at Entry (buyer’s record, $N = 90$)

Prior relationship at first entry	Not ever	Ever integrated	N	% integrated
No	43	4	47	8.5
Known to the business, not working together	3	3	6	50.0
Known, working status unclear	5	0	5	0.0
Already working with Unilever	8	6	14	42.9
Unknown or blank	17	1	18	5.6
Total	76	14	90	15.6

Notes: the programme team’s record of whether the enterprise was known to someone in the business, or already working with it, when its first project was approved, added to the tracker by the programme lead in September 2026 after reading a draft of this paper, and retrospective like the rest of the tracker. One enterprise is recoded from “No” to “Already working” on its own documents and its founder’s account (Section 3.2); on the programme’s coding the two-way comparison below is 8 of 24 against 5 of 48 ($p = 0.025$). The unknowns are mostly Indian enterprises, for which the programme lead had no visibility. Among the 72 with a known value, enterprises with any prior relationship integrated at 9 of 25 against 4 of 47 (Fisher exact $p = 0.008$), and those already working with Unilever at 6 of 14 against 7 of 58 ($p = 0.015$). A prior relationship is present in 9 of the 13 integrators with a known value (consistency 0.692), so it is not necessary for integration.

The enterprises’ own applications record the same thing from the other side, whether a relationship with Unilever existed when they applied. The two records agree on 32 of the 46 enterprises that carry both (69.6 per cent; Appendix D.3, Table 20). On the documents alone, integration is 22.7 per cent (5 of 22) for enterprises with a documented relationship, against 14.7 per cent (5 of 34) without. Re-estimating Model P with the document-coded relationship in place of the tracker’s match keeps the sign at about a third of the magnitude (+0.52 against +1.50; Table 23). That comparison sets a relationship against a match, and the result is what the nesting above would lead one to expect. On the interviewed cases the document-coded relationship is present in three of the five integrators, and enterprises with it integrate at three of six against two of six without.

Two readings of the match rating itself remain open, as Section 3.2 set out. It may record a use the buyer’s side could see and the enterprise had no reason to document, or it may record hindsight, a team rating the match in 2025 with the outcome in view. The pattern of disagreement between the team’s rating and the documents favours the first reading, since nineteen of the twenty-four enterprises the team rated as a match without a documented relationship never integrated. The team’s rating was therefore not confined to the enterprises that succeeded. The hindsight reading cannot be excluded on the record we hold. H1 is therefore supported on the programme team’s record for the match. Its relationship half is supported on the buyer’s record and, more weakly, on the enterprises’ contemporaneous one.

The interviews give a third record of the same moment, whether the founder had reach into the buyer’s world before the enterprise existed, and it predicts integration better than the documented relationship. TIE_PRIOR (Section 4.2.2) is 1 for seven of the twelve founders and substitutes for UL_LINK_START in the model. It is present in four of the five integrators (consistency 0.800). Entered with DOWNSTREAM and GREEN, it returns two terms that each rest on a single case, Platform A and Agri C, at consistency 1.00 and coverage 0.400. No combination built on it is therefore sufficient for integration beyond an individual enterprise. One of the two terms should not be credited to the specification at all, since Agri C was supplying Unilever before it applied and enters the outcome by the override of Section 3.2. The condition is inferred from the founders’ own histories, where they had worked and whom they knew before the enterprise existed. Those are the facts in a retrospective interview least open to being reconstructed around the outcome.

The transcripts record what that reach consisted of. One founder had spent two decades inside the buyer before founding the enterprise that now trades with it, and describes the counterparts across the table as “ex-colleagues” in “a small community” who “bump into each other regularly in different social gatherings”. A senior manager of a downstream distribution platform had been a TRANSFORM grantee at a previous venture and had run an FCDO challenge fund before that, and the platform’s country chief executive had worked for Unilever. A senior manager of an upstream agricultural supplier had worked on the funder’s side of the programme: “I was at FCDO, and actually, that’s kind of how I know TRANSFORM, because I was

the donor side for TRANSFORM.” Two of the three enterprises integrated, both on the first term. The third is upstream and did not integrate, so in this solution prior reach is associated with integration among distributors only.

One cell of the condition is a researcher ruling, and Appendix C.8 gives it. On the code as first applied, prior reach combined with downstream position was a path through two cases, at a coverage of 0.600. Applying the same reading to the two founders with a Unilever career removes that path and leaves the two single-case terms above. We therefore rely on the necessity result, four of five integrators against seven of twelve cases, which does not move. Reach is a different thing from the acquaintance the buyer’s record measures, and on the 9 interviewed enterprises with a value on that record the two agree in 4.

One further pattern, on the strict definition of Section 5.8, rests on three cases. Of the three enterprises with a direct commercial relationship with Unilever, one had it before the programme, and the other two, both commercial platforms, are led by people who had worked for Unilever. The third enterprise led by a former Unilever employee reached a corporate commitment that was never exercised, and no order. Prior reach into the buyer appears to be associated with reaching the point at which a commercial decision is taken, and Section 5.6 describes who takes it.

Public professional profiles give an approximate reading of how common a career inside the buyer is across the portfolio. A desk search for the founders and chief executives of the 90 enterprises, 76 of which have an identifiable founder, finds a Unilever career for one founder, a Unilever internship for a second and a Unilever award for a third, and two enterprises embedded in structures Unilever set up or staffed. The interviews had found two further Unilever careers among the twelve interviewed founders and executives that public profiles do not show, so the desk count is a floor. Whatever its exact level, a founder’s own history inside the buyer is uncommon in this portfolio and accounts for at most two of the fourteen integrators. The reach TIE_PRIOR records is broader than a career inside the buyer, and its coding across the portfolio from the same sources is in progress.

Two sensitivities on the buyer’s record move little. Coding the 18 unknown values, mostly Indian enterprises the record does not cover, as no relationship gives 9 of 25 against 5 of 65 ($p = 0.002$). Taking the programme’s coding without the one override of Section 3.2 gives 8 of 24 against 5 of 48 ($p = 0.025$). Table 21 carries both variants of the regression.

5.4 Where the matches were made

The corollary of H1 concerns where the matches that held were made, and in this portfolio they were made in the local market. How an enterprise entered the programme predicts integration more strongly than anything else we can measure about its entry. The collapsed referral variable of the regression does not show it, because it merges locally originated referrals with weakly performing regional-partner routes. Separating them, enterprises referred up by a Unilever market or regional team to the global selection panel integrate at 6 of 14, against 4 of 41 for enterprises the programme’s global pillar leads brought in and 4 of 35 for every other route (Fisher exact $p = 0.018$ across the three routes; Table 6). The two non-local routes integrate at similar rates. Pooling them, the comparison is 6 of 14 against 8 of 76 (Fisher exact $p = 0.007$; Firth coefficient 1.82, 95 per cent interval [0.56, 3.09], $p = 0.005$). The difference is therefore between a referral that began with someone in a market and every other way of arriving, central or regional.

The buyer’s record of prior relationships gives a local referral its content. Every one of the 14 enterprises a market or regional team referred up was already known to the business or working with it (8 already working, 6 known), and 6 of the 9 integrators with a prior relationship arrived by that route. The global route brought in 6 enterprises with a prior relationship, of which 3 integrated, and 26 without, of which 1 did. The regional and open routes brought in 5 with a relationship, none of which integrated, and 21 without, of which 3 did. A local referral and a prior relationship are close to the same set of enterprises. The referral result is therefore largely a statement about where the buyer already knew the enterprise, in the country business and before the programme. The programme lead’s own reading is that an enterprise a market team put forward and the global panel then selected had to fit both local and strategic objectives. That is consistent with the account in Section 5.6 of how a contract was reached inside the buyer.

A rating of local support does not carry the same information as a referral. The tracker also holds the programme team’s rating of the Unilever local team’s buy-in at project start, a four-point score of how far the country business supported the project (Section 4.3). Entered together in one Firth-penalised logit, the

Table 6: Referral Origin and Prior Relationship at Entry (ever integrated, $N = 90$)

Referral origin	No relationship		Relationship		Unknown		% integrated (all)
	N	Integrated	N	Integrated	N	N	
Local-referral-to-global	0	0	14	6	0	14	42.9
Global-led	26	1	6	3	9	41	9.8
Regional / Other	21	3	5	0	9	35	11.4

Notes: every enterprise in the portfolio sample carries a referral route once an enterprise’s projects are collapsed, so the table covers all 90. Relationship is any prior relationship on the buyer’s record at the enterprise’s first entry (Table 5). Fisher’s exact test on the three routes, all enterprises, returns $p = 0.018$; the binary comparison used in Section 7, local-referral-to-global against every other route, is 6 of 14 against 8 of 76 with $p = 0.007$ (Firth coefficient 1.82, 95 per cent interval [0.56, 3.09], $p = 0.005$), and the two tests differ because the second pools the non-local routes. Every enterprise a Unilever market or regional team referred up to the global panel was already known to the business or working with it, and of the nine integrators with a prior relationship, six arrived by that route. The three routes collapse eleven detailed channels, which Appendix D.7 gives in full and crosses with the tracker’s donor-selection marks; there the EY-led route is 0 of 4 and the FCDO route, on either of the two records that carry it, is 1 of 20 against 13 of 70 (Section 6).

locally originated referral keeps its positive coefficient and falls just short of significance at the five per cent level ($p = 0.056$), while the buy-in rating adds no information ($p = 0.62$). A referral that began in a market records that somebody in the country business already knew the enterprise and what it could be used for, which a rating of support does not capture. It is the same thing the interviews describe when programme staff make internal sponsorship the determinant of whether an enterprise is integrated (Section 5.6).

Two checks qualify the result (Appendix D.6). First, within cohort, the route’s record rests on the enterprises that entered in the first generation. None of the five it brought in during the second generation, by which time selection had been devolved to local panels, integrated. Second, the broader proposition that devolving selection to local teams improves integration reverses within the second generation, so the pooled advantage of locally selected enterprises is a composition effect of cohort. We therefore support the localisation corollary of H1 for the narrow referral route, and only weakly, since the result rests on thin cells and on first-generation entrants. We find no detectable effect of devolved selection as such. If the route marks demand that already existed locally, routing more enterprises through market teams would not create that demand. The referral result is then a further instance of the selection account of H1, and it carries no implication that the programme could raise integration by changing its referral routes.

5.5 No sufficient combination of enterprise characteristics

At portfolio scale, no combination of conditions is sufficient for integration in any of the three specifications we estimate. The three are a baseline built from the tracker’s own fields, an engagement specification on the 43 enterprises with a computable project duration, and a direct test of the case-study conditions on all 90. The result holds whichever of the sample decisions of Section 3.3 is relaxed.

The document-coded sample at $N = 56$ gives the same result in more detail. Its baseline model does return a solution, of three terms. Every enterprise that has one of the three combinations integrated, but each term covers a single enterprise (Platform A, Platform L and Recycling L), and together they reach three of the ten integrated enterprises, a coverage of 0.300. Of the six specifications that run on the full 56, two return a solution, and all but one of the four terms those two produce rest on a single enterprise (Appendix C). No specification on this sample reaches more than three of its ten integrators, four of the ten are reached by none, and every term that reaches one holds one enterprise or two. The combinations that do separate integrators therefore describe individual enterprises.

The document-coded sample also gives the integration rate for each condition on its own, on an independent coding (Appendix C, Table 17), and three of these rates bear on the hypotheses. A documented relationship with Unilever at entry raises the integration rate by eight percentage points. That is in the direction of the buyer’s own relationship record, and short of the precondition the team’s match rating shows (Sections 5.2 and 5.3). Downstream position outperforms upstream, a contrast that is unstable across samples and is reported in Appendix F.2. The integration rate does not differ by priority fit, and it is lower for enterprises with green credentials (Section 5.7).

On the interviewed cases, twelve enterprises of which five integrated, the same result holds with fewer cases per term. Every model on this tier runs at three conditions, and we run eight, the baseline, the

interview-coded substitution of Section 5.3 and six robustness specifications. Consistency is 1.00 wherever a solution exists. The best-covering document-coded specification reaches three of the five integrated cases (a coverage of 0.600), two specifications return no solution, and the terms that appear combine two or three conditions. The baseline is one of the two without a solution, because the single term it returned earlier, priority fit without green credentials, rested on Agri C alone. The entry-stage coding now places that enterprise outside priority fit. The terms that do appear each rest on one or two cases, and Section 6 shows what happens when the nineteen low-confidence cells are treated as unknown.

No specification reaches any of the seven non-integrated enterprises, Refill A, Platform B, Recycling B, Agri B, Platform C, Refill C and Agri D, which at a consistency of 1.00 is what a solution for integration must show. What the configurational analysis leaves unexplained on this tier is why those seven did not integrate, and Section 5.6 takes that up.

The two larger samples qualify the deep-tier result. Priority fit does not raise the integration rate in the document-coded sample. The association that survives there is the absence of green credentials, which the green contrasts at portfolio scale (Section 5.7) and the staff interviews on how the buyer accounts for sustainability corroborate. Even in a case set selected for circular-economy enterprises, green credentials are not associated with integration. One result is inconsistent with the corollary of H2. On the interview coding, green credentials given as the reason for selection, combined with a prior international tie, are sufficient for integration in two of the five integrated cases on this tier, both of them upstream. We report it in Appendix C.8 because the truth table it comes from is saturated and the term is closer to a description of two enterprises than to a configuration.

The result is unchanged when the conditions are replaced by the three determinants of governance form in Gereffi et al. (2005). These are how complex the information that must pass between the parties is, how far that information can be written into a specification, and whether the supplier can already meet it. We code all three for the twelve interviewed enterprises from their entry-stage and closeout documents, and we derive the five-fold governance form from that framework’s own table instead of assigning it, with a quoted passage behind each of the seventy-two cells (Appendix F.5). The forms do not separate the outcome at either moment, which is to be expected on twelve cases across five categories. The change the coding does show is on the supplier’s side. Supplier capability changes in four of the twelve enterprises, upward in three of those four. By project close four enterprises, Refill A, Recycling B, Agri B and Agri C, are at the modular form the framework associates with arm’s-length supply to a codified specification. One of the four records a Unilever transaction in its closeout report, and it buys through Unilever’s distributors and not from Unilever. Three of the four are among the seven the specifications do not reach.

The closeout documents, which are contemporaneous project records, record what is missing in the same terms the founders use in retrospect. The processors Agri B worked to “specify variety, grade, size, moisture, delivery window and volume” and “will not sign a binding purchase agreement, commit to a price in advance, or provide any pre-financing”. The consignment that case turns on was rejected with “no appeals process, no independent quality assessment and no mediation”. Refill C’s milestone report records the bulk supply agreement as still unfinalised and lists its blockers as the buyer’s own sales, commercial and marketing sign-offs, with no capability gap recorded anywhere in the document. The three determinants govern whether a transaction can be conducted at arm’s length. They do not address whether the buyer chooses to conduct it, and in this portfolio that is where the cases differ.

5.6 The buyer’s decision in the interviews

The tracker holds the portfolio-scale form of what the interviews describe. It grades, for each enterprise, the strength of the Unilever link during the project and the engagement of Unilever’s local team. On both, the outcome is more sharply divided than on anything measured at entry (Appendix D, Table 24). Not one of the 23 enterprises rated None or Low on link strength during the project ever integrated, against 8.7 per cent of those rated Medium and 27.3 per cent of those rated High. Of the 17 rated None on local engagement, one integrated, Agri C, whose supply relationship predates the programme. The association cannot be read as engagement causing integration. Both ratings were made after the event by the team that coded the outcome. Both record the team’s view of how far the project mattered to Unilever, in a few cases action already under way and mostly potential, so an enterprise on its way to a contract will show a strong link during the project close to by construction. What the pattern does show is a second threshold, on the

buyer's side. Apart from Agri C, enterprises toward which the buyer's organisation did nothing during the project did not integrate, whatever their own capability or fit.

Interviewees describe three factors that the condition set does not measure, whether a commercial sign-off route existed, whether a champion had the authority to use it, and whether a budget holder had a live use for the enterprise. Asked why a well-disposed Unilever counterpart could still not move a decision, one refill founder gives the reason: "it's too hard. It's too hard. You need someone from supply chain, someone from finance, someone from master data, someone from... well, multiple people from marketing. Everybody has to buy into this thing." In the cases the founders describe, a supply or distribution contract required simultaneous assent from four or five functions with unrelated objectives, and the joint probability of that assent is not a function of the supplier's characteristics. How far that holds across the portfolio we cannot say, since a pilot or a grant-funded project needs no such assent and the interviews cover twelve enterprises.

On the founders' and staff's accounts, the events that decided the outcome in those cases were specific to the buyer and the moment, and no firm-level covariate could capture them. An upstream agricultural supplier applied to grow a vegetable crop and learned mid-project that the buyer had not yet qualified a variety of it. Another went "6 months or more" without an order because the buyer "tried to change the formulation" and then reverted. In a senior programme manager's account, a country sub-portfolio's prospects were fixed on the funder's side, because the FCDO country team the programme asked was a health team: "when you ask a health team, what would you like to focus on, the answer is health ... and Unilever, at the same time, was stopping funding any health."

The interview coding shows how general these accounts are across the twelve (Figure 1; Appendix E.2). Six of the twelve founders and four of the seven staff describe a gap in the buyer's own procurement or resource allocation, with staff describing the cause and founders its effect, resource that "is not available because it's been allocated to short-term delivery on an existing program". Five founders name a specific individual inside the buyer as decisive, three describe losing that person and four the buyer's restructuring. The two code groups that describe the internal organisation of the buyer and the mechanics of the buyer link are the largest in the coding. We therefore read the configurational null as a finding that the deciding conditions are internal to the buyer's organisation and specific to a moment in time, and are not recorded in any administrative data on the supplier.

Across the non-integrated enterprises we interviewed, the proximate reason was more often a decision the buyer did not take than a capability the enterprise lacked. One refill founder had "buy-in from the international team, and we just never got buy-in from the local team," whose contact was invested but not "[s]enior enough to drive the decision within the business unit". A downstream distribution platform was told, after presenting to the buyer's senior management, "this is the number one thing we need to pursue," and then "nothing ever came of that," partly because the project lead who had introduced them "left the team." A regional programme manager describes the same process from inside. Relevance to a priority the leadership has already chosen "increases a lot of internal sponsorship ... which is one of the biggest determinants of whether an SME actually gets integrated or not," and "plays a huge role in which SMEs get selected," so the same factor affects both selection and integration.

The programme reliably produces a local champion, but a champion, however committed, cannot on their own get the buyer's procurement and commercial functions to act. That limit appears in six of the twelve founder interviews and four of the seven staff interviews. It appears as a champion holding "a dotted line relationship" to the team that would place the order, as a market team with buy-in "only so far that they could push," and, from a founder whose relationship did convert, as a pilot "approved and greenlit by the head of sustainability." It is absent from every condition set we estimate. We report it as an observation for testing on a larger interview sample and not as a variable, because the passages separate the seniority of the person holding the relationship, the location of the decision right and whether a commercial sign-off process existed. The quotations in this subsection come from nine people, six founders and three staff, with no more than five from any one of them (Appendix E.2.3).

Several of the same enterprises formed relationships with other buyers, and their accounts of those relationships are consistent with this reading. An upstream recycling enterprise whose material had reached one Unilever pack was taken up by another global buyer, through a component maker, in four months, where the Unilever process had taken two and a half years. The founder's explanation is that at the second buyer "I was directly in touch with the decision maker," who "just decided that this has got to be done," whereas at Unilever the champion sat in another country and the local brand team had only "a dotted line relationship"

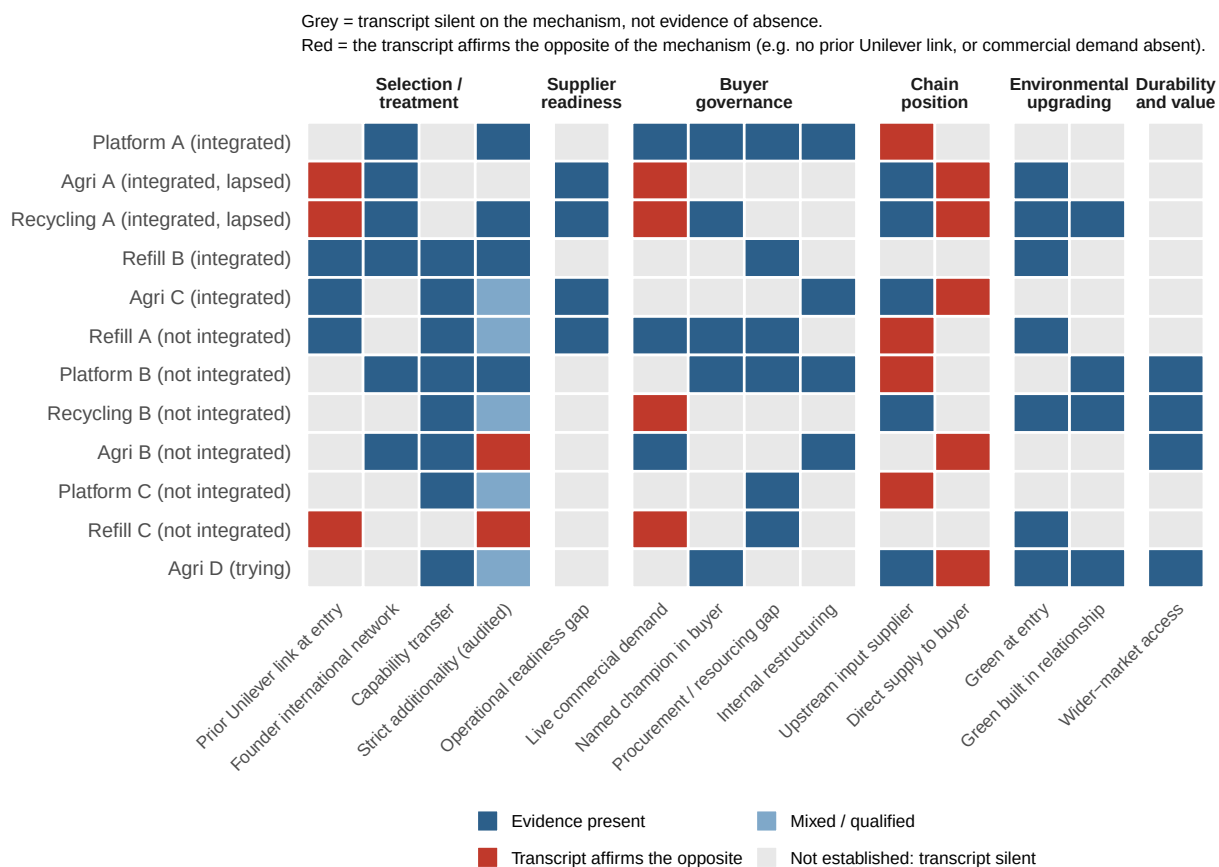


Figure 1: Interview evidence by literature-derived mechanism: which mechanisms each founder describes, interviewed enterprises

Notes: the fourteen mechanisms most central to the hypotheses and to programme value, grouped by the literature domains of Section 2, read against each founder's transcript. Each row is an interviewed enterprise, labelled with its integration outcome from the tracker (with the three overrides of Section 3.2); the label describes the case and does not enter the coding. A cell is dark where the transcript evidences the mechanism, light where the evidence is mixed or qualified, red where the transcript affirms the opposite of the mechanism (for instance that no prior Unilever link existed, or that commercial demand was absent), and grey where the transcript is silent; silence is never treated as absence. The buyer-governance block is the densest and is dense on both sides of the integration line, which is the sufficiency null seen case by case; the founder-network column is the TIE_PRIOR condition of Section 5.3; the benefit columns are read in Section 5.10. Appendix E.2 extends the grid to all twenty-eight mechanisms. The figure covers all twelve interviewed enterprises, including Recycling B, with the rows carrying the enterprises' pseudonyms.

to her. An upstream agricultural aggregator kept, after the programme ended, the processor in Unilever’s chain that the programme had introduced it to, and says that with other corporates it now walks away from ninety-day payment terms. A refill enterprise had piloted with two other multinationals before Unilever, and a downstream platform lists other consumer-goods manufacturers in its pipeline. Where a relationship formed quickly, the founder attributes it to direct access to the person who could decide, which is also how the non-integrated cases explain why theirs did not.

Founders and staff are two sources on the same relationships, and where they disagree we report the disagreement (Section 4.1). The two groups broadly agree on the mechanisms most relevant to the configurational result. Live procurement demand is raised in three of twelve founder interviews and all seven staff interviews, the buyer’s resource-allocation gap in six of twelve and four of seven, and a named individual inside the buyer in five of twelve and, counting project-lead support, all seven (Appendix E.2, Table 28). Two themes are raised almost entirely by staff, the tension between the funder’s and the buyer’s objectives and the shifts in the buyer’s sustainability priorities, which founders are not positioned to observe.

The two groups disagree on attribution. For two of the seven non-integrating enterprises, staff attribute non-conversion to the supplier’s readiness where the founder attributes it to the buyer’s conduct. A senior programme manager says of one early-stage platform that “the gap between where they were and where they needed to be for Unilever to integrate was too big,” where that founder’s account is of a refusal of systems access and a senior endorsement from which “nothing ever came.” Each group has an interest in the attribution that favours it, and our evidence cannot adjudicate between them. The choice of account decides whether the policy reading is that suppliers should be better prepared or that the buyer’s conversion process should be fixed. The same division exists in the literature, between the supplier-barriers account of Section 2.3 and the buyer-governance account of Section 2.4.

The founders’ accounts of the seven non-integrated enterprises allow us to place the binding constraint in each case. That classification is our synthesis across cases and is not a coded variable. On it, six of the seven divide as follows. Four describe a constraint inside the buyer’s organisation, a procurement function never engaged, a champion without the standing to commit the buyer, a refusal of technical access, or resources allocated elsewhere. One describes an upstream and structural constraint, scarce certified seed, aggregation economics, a consignment rejected on moisture content. One describes an external shock interrupting an unfinished process whose champions remained in place. We leave the seventh, Recycling B, unclassified because its transcript does not have the mechanism codes the synthesis rests on. Buyer-side organisational factors therefore explain most of the failures in these cases, though not all of them.

The same contrast appears when we set the enterprises a specification does reach against those it does not. We confine this comparison to the seven document-derived specifications, so that a case is not counted as explained because its founder had an international network. Five of the eight enterprises those specifications do not reach describe a procurement or resourcing gap, against one of the four they do reach. Three of the four reached describe a gap in the enterprise’s own operational readiness, against one of the eight unreached (Appendix E.2, Table 32). The four reached enterprises are all integrators and only one integrator is unreached, so this is close to a comparison of integrators with the rest. The enterprises that did integrate report readiness gaps more often, which is evidence against reading non-integration as a capability failure. On twelve cases these are descriptions. What they show is that the enterprises the conditions cannot reach are the ones whose accounts are dominated by decisions inside the buyer.

5.7 Impact credentials and integration

The programme selects enterprises for their environmental or social impact, and the corollary of H2 is that the credential which earns entry to the programme should not, on its own, raise the probability that the buyer trades with the enterprise. Because the result is a null, the definition of green matters more than it would for an effect, so we measure green in three ways. On the portfolio sample we use the programme’s own pillars. Every enterprise was funded under a primary theme, and the themes divide into the environmental pillars (Recycling, Refill and Regenerative Agriculture) and the social ones (health and digital health, water and sanitation, farmer and retailer livelihoods and the future of work), with a small energy pillar beside them. The portfolio contrast is therefore between enterprises funded for environmental impact and enterprises funded for social impact.

On the document-coded sample we use the GREEN condition of Section 4.2.2, coded from each enter-

prise’s own application, which is a different thing from the pillar since it records credentials the enterprise held at entry. On the interviewed cases the interview coding separates credentials that pre-dated the programme from greening that emerged through it, and agrees with the document-coded condition on ten of the twelve. Appendix F.1 separates certification from sector membership within the GREEN condition, reports each pillar family on its own, and records what the interviews say about credentials deepened during a relationship.

Table 7: Integration rates of green and social enterprises, and by green credentials at entry, across three samples

Sample	Green, and what it is compared with	N	Green	Comparison
Portfolio sample	environmental pillars, against the social pillars (tracker)	90	0.152	0.158
Document-coded sample	green credentials at entry, against none (GREEN, from the application)	56	0.133	0.231
Interviewed cases	green credentials at entry, against none (GREEN, from the application)	12	0.286	0.600

Notes: the outcome in every row is the tracker’s ever-integrated status (Section 3.2). In the portfolio row green is the programme’s own Recycling, Refill and Regenerative Agriculture pillars and the comparison is the social pillars, with the small energy pillar counted among them (5 of 33 against 9 of 57); in the two document-coded rows green is the GREEN condition of Section 4.2.2 and the comparison is the enterprises without green credentials at entry. The interviewed cases are ruled off because they were selected on circular-economy criteria; their rates are properties of the case selection and not of the portfolio. Rates by pillar family, and the two keyword readings of the pillar field, are in Appendix F.1.

We find no higher rate of integration for green enterprises on any basis, and on one basis a lower rate that does not reach significance (Table 7). On the programme’s own classification, enterprises funded under the environmental pillars and enterprises funded under the social ones integrate within a percentage point of each other, 15.2 against 15.8 per cent ($p = 1.00$), and no pillar family is distinguishable from any other ($p = 0.80$; Appendix F.1). On the enterprises’ own documents the gap is 9.8 points against green, 13.3 against 23.1 per cent. On 56 enterprises that difference is within what chance would produce (Fisher $p = 0.49$), and it appears only when conditions are coded from application-stage documents (Appendix B). Among the interviewed cases, the absence of green credentials is in the only surviving sufficient term, and on the strictest definition of integration the three enterprises with a direct commercial relationship are all outside a green sector (Section 5.8). Those twelve were chosen on circular-economy criteria, so their rate has no weight, but the configurational evidence is consistent with the portfolio evidence. The corollary therefore holds. A credential is associated with entry into the programme and does not raise the probability of integration, and the null does not depend on whether green is read from the buyer’s label, the enterprise’s application or the founder’s account.

Unilever and programme staff give the reason, and it applies to social credentials as well as environmental ones. Asked about the social and environmental lens these enterprises bring, a Unilever manager answers that “[t]hat isn’t incorporated into a normal financial P&L . . . [t]he social portion of it and the sustainability portion of it isn’t integrated, and therefore . . . [f]rom a business-as-usual operations, it will never be executed.” A green or a social credential counts with the budget that funds pilots and does not count with the budget that signs contracts, which is why the environmental and social halves of the portfolio integrate at the same rate. The programme’s own guide places the same boundary on the procurement side, that “impact enterprises may need to charge a premium for their products or services, while cost is always one of the key criteria reviewed by procurement” (Unilever, FCDO and EY, 2026).

Founders describe the boundary from outside it. One refill founder distinguishes integration from its appearance, since enterprises that are truly integrated “will have access to not just . . . the sustainability department, but also into the commercial departments, like sales, marketing, product, packaging”. A regional programme manager names the incentive that governs the commercial side, “if their turnover becomes low, they are not going to be questioned around how sustainable the partner is.” Our priority-fit condition is our measure of the buyer demand that the environmental-upgrading literature makes decisive (De Marchi

and Di Maria, 2019; Ponte, 2022a). On the entry-stage coding it makes no difference in the document-coded sample, and it is part of the interviewed cases' single surviving term together with the absence of green credentials. Twelve cases cannot test the mechanism further. Appendix F.1 adds two refinements to the null. Green characteristics in this portfolio are mainly selected at entry and sometimes deepened during the relationship, once at the buyer's explicit request, with no higher rate of integration in either group. Type and chain position are nearly collinear here, so the green contrast and the chain-direction contrast of Appendix F.2 cannot be separated in this sample.

5.8 Definitions of integration

Every number above takes Unilever's tracker as the definition of integration. The tracker's working rule appears to be that goods demonstrably reached a Unilever product or that a contract was signed. That is one defensible rule among several. Because the interviews give the enterprises' own account of the same relationships, we can test how sensitive the findings are to alternative definitions of integration. On the interviewed cases we hold three definitions. The first is the tracker's. The second is the enterprise's own account, which counts any route to Unilever the founder reports, including supply to a processor in Unilever's chain and registration as a vendor. The third is a strict account, which counts only a direct commercial relationship with Unilever, meaning a contract, purchase order or payment between the enterprise and Unilever itself, in either direction, and not a flow that reaches Unilever through a processor, converter or distributor. The tracker's definition remains the outcome throughout, and the other two are sensitivity checks. They give 5, 7 and 3 integrated enterprises out of twelve, and each interview-coded definition agrees with the tracker on 10 of the 12.

The disagreements differ in direction. Refill A and Agri B report a route the tracker does not count, a vendor registration with no transaction behind it and a supply relationship with a processor in Unilever's chain. Agri A and Recycling A are tracker-integrated but describe supply through an intermediary, so they drop out under the strict reading. The strict reading is deliberately narrow. It excludes Recycling A, which the programme presents as its example of a supplier partnership, with Unilever's country business sourcing recycled plastic through its local supply chain from the enterprise (Unilever, FCDO and EY, 2026). The strict definition does not question that relationship. Under it the enterprise is not counted because the founder describes the material as reaching Unilever through a converter, with no contract or purchase order from the brand, so that on its own account the enterprise supplies Unilever's chain without supplying Unilever directly. The tracker and the programme count it and the strict reading does not, which is a difference of definition. The tracker therefore uses a broader definition than direct commercial supply, but a narrower definition than any relationship the enterprise reports.

Setting the three definitions against what each founder describes shows what the disagreements are about (Table 8). One supplies Unilever directly as a first-tier supplier, three reach Unilever through a processor, converter or recycler, and two are commercial platforms that buy Unilever products and resell or distribute them. One is registered as an approved vendor with no transaction behind the registration, one holds a corporate commitment, a first right of refusal that was never exercised, with no payment in either direction, and three describe attempts and pilots that produced no relationship. The mode of integration follows chain position, with supply through an intermediary for the upstream agricultural and recycling cases and platform integration for the downstream ones. The founders' accounts of the mode match the direction coded from their applications in nine of the ten cases to which a mode is assigned (Section 6).

The programme's own account recognises the same variety. Its guide distinguishes supplier partnerships, advertising partnerships, corporate-responsibility partnerships and innovation partnerships as ways an enterprise has been brought into a partner's value chain, and names enterprises in this portfolio under each (Unilever, FCDO and EY, 2026); the tracker's single integration status combines those four into one.

Two things follow for what an integration count means. First, a direct commercial relationship and first-tier supply are different things. Of the three enterprises with a direct commercial relationship under the strict definition, two are platforms that buy from Unilever and are therefore its customers. Platform A, for example, is integrated on every definition in use, its founder describes purchases and distribution of Unilever products, and the programme's guide records commercial agreements with Unilever in four countries (Unilever, FCDO and EY, 2026). It is a distributor, and the returns literature of Section 2.1 concerns suppliers. The one enterprise supplying Unilever commercially among the twelve, Agri C, had that relationship before the

Table 8: What each definition of integration counts, by the relationship the founder describes ($N = 12$)

Enterprise	Relationship described	Counted as integrated by			Payment or order evidence
		Tracker	Enterprise	Strict	
Agri C	Direct first-tier supplier; the relationship pre-dates the programme	•	•	•	evidenced
Platform A	Commercial platform: buys and distributes Unilever products	•	•	•	evidenced
Refill B	Commercial platform: buys and resells Unilever products	•	•	•	evidenced
Agri A	Supply through an intermediary	•	•	—	evidenced, lapsed
Recycling A	Supply through an intermediary	•	•	—	goods flow, no contract with the processor
Agri B	Supply to a processor in Unilever’s chain	—	•	—	
Refill A	Registered as an approved vendor; no transaction	—	•	—	none
Refill C	Corporate commitment (a first right of refusal, not exercised); no payment flow	—	—	—	none
Platform B	Attempts and pilots; no relationship formed	—	—	—	none
Platform C	Attempts and pilots; no relationship formed	—	—	—	none
Agri D	Attempts in progress; no relationship formed	—	—	—	none
Recycling B	Attempts in progress; no relationship formed	—	—	—	none
Counted		5	7	3	

Notes: the three definitions are the tracker’s; the enterprise’s own account, which counts any route to Unilever the founder reports; and the strict account, which counts a direct commercial relationship only, a contract, order or payment between the enterprise and Unilever itself. The relationship column is a descriptive classification of what each founder describes, and every assignment rests on quoted transcript evidence held in the interview-coding registry. The payment column records what the founder reports; “none” means no payment or order was described in either direction, which is not evidence that none occurred.

programme, which is why Section 3.2 codes it integrated against the tracker. Its founder places the purchasing leg with co-packers who “try to buy cheaper somewhere else”, so the interview coding marks the relationship as mediated. The interview code for direct supply is applied to no founder transcript. The endline reports of the eight integrated enterprises that have one document no first-tier supply relationship created by a project (Appendix F.4). On the founders’ own accounts, therefore, what the tracker calls integration is almost never supply, since of the tracker’s five integrated enterprises one is the pre-existing supplier, two are resale platforms and two are intermediated flows. This pattern is consistent with the inclusion paradox described in the base-of-the-pyramid literature (Brix-Asala et al., 2021), in which intermediated supplier development produces participation in a corporate commitment more often than a first-tier supply relationship.

Second, the integration measure combines several economically different types of relationship. They are a first-tier supply relationship, resale platforms, intermediated flows and, depending on whose definition is in use, states of registration or commitment that involve no transaction. Six of the twelve describe no payment or order in either direction, and two of those six are counted as integrated under at least one definition in use. The tracker counts the intermediated flows it can trace and does not count the registrations it cannot, so its rule follows traceability more than the existence of a commercial relationship. Recycling A is counted with no contract or purchase order from the brand, and Refill A is not counted while listed as an approved vendor. The endline reports show the same at project close, with the caution that several integrations came after the reports were written. Of the eight integrated enterprises across the portfolio with an endline report, four describe a commercial platform and four record no payment flow with Unilever in either direction at the time of writing. Recycling A’s report names the brands it supplies recycled plastic to, Unilever not among them, with Unilever and the programme appearing as the funder (Appendix F.4). This is a second reason the count is a poor measure for a programme of this kind, and it does not depend on the selection result of Section 5.2.

The sensitivity analysis re-runs the configurational results under each definition. Under the enterprises’ broader account solutions become easier to find and coverage rises to 0.857. Under the strict account three of the eight specifications return no solution. Across the three definitions 24 of the 33 solution terms rest on a single enterprise, and no definition produces a combination that both covers most integrators and rests on more than a handful of them. The sufficiency null of Section 5.5 therefore holds under all three definitions.

One necessity result holds under the strict definition, and it is the only one in the study that is far from trivial. For direct integration, the absence of green-sector membership at entry is present in every case that integrated directly (consistency 1.00), while it is present in 58 per cent of all cases. All three enterprises with a direct commercial relationship therefore operate outside a recognised green sector. On three cases this is a hypothesis. It does not extend to the 56 document-coded enterprises, where only the tracker’s outcome is available and the condition is present in seven of the ten integrators (Appendix C). It is nonetheless the clearest instance of the green null in our data. In a case set selected for circular-economy business models, the enterprises that reached a direct commercial relationship are the ones that were not in that sector.

We do not enter direct or intermediated as a condition anywhere, because it is a property of the outcome and predicting an outcome with itself would be circular. Whether enterprises classified as integrated have better outcomes is the question the next two subsections take up.

5.9 Integration and reported impact

We test whether integrated enterprises report more jobs and more people reached than non-integrated ones, which the monitoring data allow us to do directly. Appendix F.3 reports the exercise in full. Our prior expectation is a null, since the employment gains in the linkage literature are estimated on firms that become first-tier suppliers (Alfaro-Ureña et al., 2022), and Section 5.8 has shown how seldom what this portfolio calls integration is first-tier supply. The data can only compare enterprises that integrated with enterprises that did not, observed once, in cumulative totals reported during and at the close of projects, in many cases before the relationship the tracker records had formed. Table 9 reports the comparison as we tighten it step by step. It runs from the raw cross-section to a comparison confined to enterprises that passed the entry gate and matched on cohort, chain direction and scale, and then adds the pillar family under which the enterprise was funded.

The beneficiaries difference is not distinguishable from zero once the first controls are added, and it remains so under every further restriction. The jobs difference is robust to every restriction of the design,

Table 9: What integration is worth: rate ratios as the design tightens

Design	n	Integrated	Direct jobs	Beneficiaries
Cross-section	86	14	10.66 [4.11, 27.67]	3.77 [1.15, 12.34]
+ cohort, direction, scale	55	10	18.85 [5.55, 64.02]	1.16 [0.24, 5.70]
Gate-restricted	40	10	15.32 [2.78, 84.57]	1.07 [0.18, 6.31]
Gate-restricted, matched	35	7	15.03 [4.92, 45.95]	1.19 [0.22, 6.46]
Gate-restricted + pillar family	40	10	11.42 [1.95, 66.76]	—

Notes: quasi-Poisson rate ratios with 95 per cent intervals, integrated against not, on the 86 enterprises the monitoring workbook matches to the portfolio sample; the outcome is the tracker’s ever-integrated status. A rate ratio of 2 means integrated enterprises report twice the total of non-integrated ones. Scale is the application-stage SCALE condition of Appendix C, coded for 55 of the 86. The gate is a Medium or High Unilever link at start. Matching is coarsened exact matching on cohort, chain direction and scale, cohort being the generation in which the enterprise entered because the totals are cumulative. The permutation test on the gate-restricted jobs estimate, which we prefer to the model-based p -value with fourteen treated enterprises, gives $p = 0.07$ without pillar family and $p = 0.13$ with it. Two placebo contrasts, the pillar-family panels and an exposure-adjusted specification that could not be estimated are in Appendix F.3.

and it is not robust to holding fixed what the enterprise was funded to do. With pillar family in the model the rate ratio falls from 15.3 to 11.4 and the permutation test moves from $p = 0.07$ to $p = 0.13$. Dropping one family at a time moves the estimate between 9.8 and 46.4, a wider range than the whole ladder produces.

The reason lies in what the monitoring instrument records for each family. Enterprises funded to reach people report people, and enterprises funded to process material report staff. Of the 32 environmental enterprises, 25 report a positive jobs total, against 7 of the 21 in health and WASH, and the median health and WASH enterprise reports zero jobs and zero beneficiaries. Two further properties of the data limit the reading. The programme’s own attribution measure, “individuals employed as a result of TRANSFORM funding”, is identical to total employment for every enterprise in the source workbook, so the programme’s claim for itself cannot be separated from the enterprise’s headcount, and the totals are unadjusted for exposure. Integrated enterprises therefore report more jobs than non-integrated ones. Once the entry gate, observable scale and the pillar they were funded under are held fixed, that gap cannot be separated from the buyer’s selection and from what the enterprise was funded to do. Its magnitude is in any case too large to read as an effect of integration. Given how seldom integration here is first-tier supply, that is what the linkage literature predicts, and it means that the jobs an integration count appears to bring reflect which enterprises the buyer chose as much as anything the programme did.

A second monitoring workbook lets us ask the question the other way round, whether the integrated enterprises were larger during the grant, in the capital they raised from other funders and in their own sales. Across the matched portfolio the median integrated enterprise raised GBP 279,450 in private capital, against a median of zero for the rest, and that is the only difference the workbook shows. It is a first-generation difference. Among second-generation enterprises that passed the entry gate, a third of each group raised any private capital, grants and own revenue do not differ, and in Firth logits on the gate-passing enterprises neither capital nor revenue predicts integration (Appendix F.3.2). Most of the private capital raised by the integrated first-generation enterprises was reported from 2022 onwards, after their projects had closed. It therefore reads as capital following the relationship, or the enterprise’s survival, and it cannot have been a condition for either. Scale during the grant does not sort the enterprises the buyer could use into those it traded with and those it did not, which is the null of Section 5.5 on a fourth kind of evidence.

5.10 Benefits of participation

Discussions of programmes like TRANSFORM often combine three distinct outcomes, and the interviews let us separate them. Commercial integration is the outcome the tracker records. Programme benefit is what an enterprise keeps whether or not it integrated, namely a capability, credibility with other buyers, access to markets beyond Unilever, or a green capability built through the relationship. Strict additionality is a stronger, counterfactual standard, which requires that the enterprise, its route to market or a capability would not have existed or scaled without the programme. The codebook has a code for each of the four

benefits and one for catalytic additionality. We re-read the additionality passages one by one against the codebook’s counterfactual rule to reach a strict verdict for each enterprise (Appendix E.2).

Table 10 sets the twelve interviewed enterprises out case by case, split by whether they integrated, because that split is the test of H3. If what the programme produces is separable from what the buyer decides, enterprises that did not integrate should report benefits, and the table shows which do. We report the five columns separately and never sum them, because four are coded interview references and the fifth is an audited counterfactual judgement.

Table 10: Programme benefits reported by the twelve interviewed enterprises, by integration outcome

Enterprise	Outcome	Capability transfer	Credibility	Wider market	Green built	Strict additionality
<i>Integrated or formerly integrated (5)</i>						
Platform A	integrated	—	•	—	—	met
Agri A	lapsed	—	•	—	—	no counterfactual stated
Recycling A	lapsed	—	•	—	•	met
Refill B	integrated	•	—	—	—	met
Agri C	integrated	•	•	—	—	mixed
<i>Did not integrate (7)</i>						
Refill A	not integrated	•	•	—	—	mixed: partnership expected anyway
Platform B	not integrated	•	—	•	•	met: platform would not exist
Recycling B	not integrated	•	—	•	•	mixed: reached anyway, 3–4 years later
Agri B	not integrated	•	•	•	—	no additionality passage coded
Platform C	not integrated	•	—	—	—	mixed: “it wasn’t from scratch”
Refill C	not integrated	—	—	—	—	not met: “the same... 12 months later”
Agri D	trying	•	•	•	•	mixed: no explicit counterfactual
Integrated (of 5)		2	4	0	1	met in 3
Did not integrate (of 7)		6	3	4	3	met in 1

The outcome column is the tracker’s integration status as at March 2026, with the three overrides of Section 3.2 (Agri C integrated; Agri D and Recycling B not integrated, from their own accounts); it is not read from the founders’ interviews or the enterprises’ documents. A bullet records that the code is applied to that founder’s transcript; a dash records that the transcript is silent, which is not evidence of absence. The four benefit codes are capability transfer, credibility or brand signalling, wider-market access and green capability built. The strict-additionality column is the verdict from re-reading every catalytic-additionality passage against the codebook’s counterfactual rule, that the enterprise, route to market or capability would not have existed or scaled without the programme; “mixed” marks a capability claim made beside statements that the enterprise or the relationship would have continued regardless. Platform B’s green entry is weak: the capability was self-initiated in response to a logistics constraint, was an operational arrangement and not a standard, and did not outlive the operation. No test is computed on five against seven cases and none should be.

Benefits are reported on both sides of the integration line, and on two of the four dimensions more often by the enterprises that did not integrate. Capability transfer is evidenced for six of the seven non-integrators against two of the five integrators. Examples are a business case rebuilt with the accelerator’s consultants, a commercial model imposed on what had been a grant-funded habit, and technical standards explained by the buyer’s own staff. Access to markets beyond Unilever is evidenced for four of seven against none of five, including one aggregator that became an official supplier to a processor in the buyer’s chain and kept that relationship after the programme ended. Credibility is reported more often by integrators, four of five

against three of seven. The wider-market contrast is four evidenced cases against five silent transcripts, and interview order alone could produce that silence, since founders who integrated spent the interview on the Unilever relationship while the others described what else the programme was good for. On five and seven cases nothing here is a test. We report the counts and do not compare them statistically.

Strict additionality is scarcer, and the distinction matters for how the programme is judged. Passages about additionality are common. The catalytic-additionality code is applied to eleven of the twelve founder transcripts, across twenty-eight references. Re-read against the counterfactual rule, four of the twelve meet it, three of them integrators. Among the seven that did not integrate only one does, a digital credit platform whose founder says “without a doubt, we wouldn’t have built [it] and kind of taken it to market without the TRANSFORM funding,” and whose parent firm funded its other platforms from other sources. Four non-integrators are mixed, with a capability claim made beside statements that the enterprise or the relationship would have continued regardless. One founder describes acceleration only. Asked how the firm would have evolved without the programme, the founder answered “it would be the same, it would just be... probably 12 months later.” In one transcript no additionality passage is coded. That strict additionality concentrates among integrators is consistent with the selection account of Section 5.2. The comparison is weak, though, since on the seven document-derived specifications it is two of four enterprises a configuration reaches against two of seven it does not (Appendix E.2, Table 32).

The pattern also bears on the null of Section 5.5. Capability transfer is reported by six of the seven enterprises no document-derived specification reaches, against one of the four a specification does reach. A gap in the enterprise’s own operational readiness is reported by three of the four reached, against one of the seven. The capability benefits are therefore reported mostly by the enterprises that do not appear in the integration measure.

The evidence therefore supports the weaker claim. For the seven enterprises that did not integrate, founders report a capability gained in six cases, access to other buyers in four and credibility in three, and something that would not otherwise have existed in one. That is close to the pattern Masso and Vahter (2023) report for suppliers whose multinational relationship ended, where productivity does not fall on average because firms keep what they built. Extending it to enterprises whose relationship never formed is our inference from interview evidence. One of the seven reports no benefit of any kind and describes the programme as capital that arrived about a year early.

The integration count omits enterprises that report the programme’s clearest benefits, and it includes one whose relationship pre-dates the programme. Whether that pattern holds beyond the twelve interviewed enterprises is a question the milestone and final reports of the wider portfolio could answer, since they record capability and market outcomes for every project. We have not coded them for this purpose and note it as the natural extension.

6 Robustness and sensitivity

The configurational strand supports the paper’s central negative claim. With twelve cases on one tier and fifty-six on the other, it is the strand most exposed to the analyst’s choices, so we vary them and report what changes, and Appendix C gives every specification in full. The sufficiency null is robust to four choices, the condition set, the definition of the outcome, the treatment of documentary silence and the case set. We run eight specifications on the interviewed cases and six on the whole document-coded sample. They substitute a governance measure for the documented relationship at entry or a scale pathway for green credentials, separate certification from green-sector membership, or add a codified-specification condition. Across all of them, consistency is 1.00 wherever a solution exists, and no document-coded specification reaches a coverage above 0.600. Under the enterprises’ broadest account of themselves coverage rises to 0.857. Under their strictest, three of the eight specifications return nothing. Across the three definitions, twenty-four of the thirty-three solution terms rest on a single case.

Nineteen of the ninety-six condition cells on the interviewed cases were coded 0 at low confidence because the application is silent, and all nineteen were among the cells verified by hand (Section 4.2.3). Treating them as unknown removes the solution of one specification. Redrawing them at random ten thousand times, at each of three priors, leaves the median best coverage at 0.600, which is still the canonical maximum and now rests on M7 alone (Appendix C). Adding the twelfth interviewed enterprise as a non-integrated case changes no reported specification.

Necessity is where a small sample most easily misleads, so every necessity result is reported beside its triviality. The one portfolio-scale result that clears the threshold, a match at entry, has a triviality of 0.733, so it is necessary but common, and it does not predict which of the enterprises that have it integrate (Section 5.2).

Because the paper reads three sources against one another, how far they agree bears on every result. On almost everything they agree closely (Section 3.2). The two document-coded tiers assign identical values on every shared cell. The founders’ accounts of how their enterprise reaches Unilever match the direction coded from their applications for nine of the ten enterprises to which a mode is assigned, and the interview code for green credentials matches the document-coded condition for ten of the twelve. The two interview-coded definitions of integration each agree with the tracker on ten of twelve, and they were built twice, from the interview corpus and from the endline reports, with the two codings agreeing cell for cell.

They diverge at entry, and the divergence is one of construct. The programme team’s rating of the match at project start and the enterprise’s own application record of a relationship agree on 48 per cent of the fifty-six enterprises we hold both for, a κ of 0.06, with the disagreement one-sided. The buyer’s own record of a prior relationship, which asks the documents’ question, agrees with them on 69.6 per cent (Section 5.3). Set against the agreement everywhere else, the divergence between the match rating and the documents is what two different constructs would produce, and Section 5.3 discusses the two readings of the match rating itself. The founder and staff interviews are a further pair of sources on the same relationships. Section 5.6 reports where they agree and where, for two of the seven non-integrating enterprises, they give incompatible accounts of the same failure.

The coding decision with the largest effect on the results was restricting the evidence to what was available at application. Shown the whole record, a coder can satisfy a condition defined at application with evidence of what happened afterwards. Coding the same fifty-six enterprises twice moved twenty cells one way and eleven the other, with the movement largest on the documented relationship at entry, where it was one-sided. For that reason every condition is coded from application-stage documents alone, and the document-coded green contrast is negative on the applications and flat on the full record (Appendix B).

Appendix D reports three further checks. Cohort shows no association, with 4 of 32 first-generation enterprises ever integrating against 10 of 58 in the second (Fisher exact $p = 0.76$). Cohort is read from the project the regression keeps for each enterprise, and Section 3.1 sets out why the enterprise-level construction reverses the comparison. Thirty-five enterprises have a blank in the tracker’s integration field, which the programme lead confirmed records non-integration (Section 3.3), and most of them are first-generation. Had those enterprises been dropped instead of coded as not integrated, the first generation would have shown the higher integration rate of the two, because its four integrations would be set against a much smaller denominator. The cohort comparison therefore depends on coding blanks as non-integration, and the count of integrations is the same on either reading. Tenure shows no association either. Time in the programme

is unassociated with integration (Spearman $\rho = 0.11$, $p = 0.39$, on the 59 enterprises with a computable duration), which is a loose test of the tenure component of H1.

On the coding, the interview coding is single-coder with no reliability statistic. A second member of the research team checked the machine-suggested condition codes on the interviewed cases against the source documents, taking all 19 low-confidence cells and a random sample of 35 of the other 77, 54 in all. The check confirmed 48 and changed six (an agreement of 88.9 per cent). The six revisions move the solution terms and coverage of several deep-tier specifications without altering any necessity conclusion, the sufficiency null or any hypothesis verdict (Appendix B).

One check bears on how the funders read the count. The programme's lead told us that some enterprises entered TRANSFORM without any prospect of a Unilever relationship, selected for a donor, and the tracker records this in its notes and referral fields. Twenty of the ninety have the mark. Their integration rate is 5.0 per cent (1 of 20) against 18.6 per cent (13 of 70) for the rest, which is the direction the account implies and, on one event in twenty, not distinguishable from zero (Fisher exact $p = 0.179$). The referral channels identify the same twenty, and the four enterprises the EY route brought in integrated at 0 of 4 (Appendix D.7). Removing the donor-selected twenty leaves 70 enterprises with 13 integrations, and the entry result is unchanged. Not one enterprise below the threshold integrated, 0 of 15 against 0 of 24, with a Firth coefficient of +2.29 ($p = 0.032$). The Fisher test rises to $p = 0.056$, which reflects the loss of nine below-threshold enterprises and no change in the pattern. One donor-selected enterprise, Agri A, did integrate.

7 Discussion

Table 11 records the verdict on each of the three hypotheses of Section 2.7. The discussion follows the three forms the inception design proposed for the findings, which are confirmation and refinement, extension, and contradiction, in that order.

Table 11: Hypotheses and verdicts

	Claim	Verdict	Evidence
H1	Integration is settled at entry, by whether the buyer’s side saw a use for what the enterprise proposed	Supported on the programme team’s record for the match; a prior relationship raises the rate on the buyer’s and the enterprises’ records and is not necessary	0 of 24 without a match at entry integrated against 14 of 66 with one ($p = 0.017$); Medium and High matches indistinguishable above it ($p = 0.201$); no association with time in the programme ($\rho = 0.11$). A prior relationship on the buyer’s record: 9 of 25 against 4 of 47 ($p = 0.008$), present in 9 of 13 integrators; on the enterprises’ own documents the gap is eight points and half the integrators show no relationship; on the interviewed cases the founder’s prior network is closest to necessary (four of five integrators). Corollary: enterprises referred from a Unilever market integrate at 6 of 14 against 8 of 76 ($p = 0.007$), every one of them already known to the business, and the rating of local buy-in adds nothing beside the route
H2	Past the gate, the buyer decides: no entry-stage combination is sufficient	Supported; buyer-side constraints explain most, but not all, of the unexplained cases	No sufficient combination at $N = 90$; at $N = 56$ no specification reaches more than three of the ten integrators and every term holds one enterprise or two; holds on three outcome definitions and with silence treated as unknown. Of the seven interviewed non-integrators, four describe a constraint inside the buyer, one an upstream constraint, one an external shock, and one (Recycling B) carries none of the mechanism codes the classification rests on. Corollary: enterprises funded under the environmental pillars integrate at 15.2 per cent against 15.8 for the social ones ($p = 1.00$), and green credentials at 13.3 against 23.1 on the enterprises’ own documents ($p = 0.49$)
H3	Participation produces benefits that integration does not measure	Supported on the interviewed cases	Of the seven interviewed non-integrators, six report a capability kept, three credibility and four access to other buyers; on the strict counterfactual test one of the seven, and four of the twelve, describe something that would not otherwise have existed

7.1 Confirmation and refinement: selection at entry

The gate at entry confirms the sorting account of large-buyer entry (Iacovone et al., 2015) in a specific form. The sorting variable is the buyer’s judgement of the project, whether the buyer’s side could see a use for what the enterprise proposed. The programme team rated that judgement, and the enterprise’s own documents had no reason to record it. The team compiled that record after the event, and it also recorded the outcome, so part of the perfect gate may be hindsight (Section 5.3). A prior relationship with the buyer is a different sorting variable, and the three records that measure it agree on its role. It raises the probability of integration among the enterprises the buyer had matched, and it is not necessary, since 4 enterprises integrated from none.

The experimental literature describes a later stage of the same sequence. Atkin et al. (2017) and Bloom et al. (2013) allocate buyer access or advice to firms and measure what follows, and they do not observe how

a buyer decides whom to allocate it to. Where that decision is observed, as here, it does most of its work before the programme starts.

The evidence also refines the sorting account. The entry effect is a threshold, since nothing below a Medium match integrated and Medium and High integrate at rates a penalised logit cannot distinguish. A model of the buyer’s interest as an input to be built up is therefore describing something these data do not show. Nor does the association strengthen with time in the programme. Macchiavello and Morjaria (2015) show relationship value rising with a relationship’s age, which is the theoretical basis for expecting a multi-year programme to convert weak ties into strong ones, but time in the programme is unassociated with integration here ($\rho = 0.11$, $p = 0.39$). We hold that result loosely, since the tenure measure assigns time-so-far to unresolved projects.

The localisation corollary of H1 is supported in part. A referral that began in a Unilever market outperforms every other route, and a rating of local support entered beside it adds nothing. Route and generation are entangled, though, and the broader proposition that devolving selection raises integration reverses within the second generation (Appendix D.7). The buyer’s record of prior relationships gives the referral its content. Every enterprise a market team referred up was already known to the business, so the route marks a country business that already knew what the enterprise could be used for. We do not take from it that devolving selection would raise integration rates.

7.2 Extension: integration depends on coordination inside the buyer

The governance tradition treats the lead firm as a coherent actor that governs its chain (Gereffi et al., 2005). The interviews describe a buyer that is internally fragmented, in which integration requires simultaneous assent from functions with unrelated objectives and the relationship is held by individuals who leave (Section 5.6). In the interview coding the codes for a champion inside the buyer, for the champion’s loss, for the buyer’s resource-allocation gap and for its restructuring together account for 61 of the 415 references (Appendix E). Operations and supply-chain research has examined these internal coordination problems. It describes procurement units that must be engaged before sustainability requirements move (Villena, 2018), purchasing and sustainability managers who read the same supplier differently (Xiao et al., 2018), and boundary spanners and lower-level ties through which the relationship is held (Zhang et al., 2011; Huang et al., 2016). The value-chain literature has not examined them, and our extension is to bring them into its account of the lead firm.

Internal fragmentation explains the sufficiency null. If integration is the joint outcome of several functions’ independent assessments, its probability is low, weakly related to any single supplier characteristic and sensitive to personnel churn, which is what we observe. It also runs against a result in the relational-contracting literature. Macchiavello and Morjaria (2015) show relationship value accumulating with a supplier’s reliability. Here, relationships are repeatedly broken from the buyer’s side, by its own churn, before that value can accrue. Fragmentation also identifies a variable the administrative record lacks. The tracker records whether a champion existed inside the buyer. The failures the interviews describe turn on whether that champion could get procurement and the commercial functions to act, which the tracker does not record.

The programme team reached a neighbouring conclusion on its own. Its guide names “strong, committed project leads” as “the single most important driver of success” and advises partnerships to “clarify organisational structure and where decision makers sit” and to align internal teams that “often operate with different risk thresholds” (Unilever, FCDO and EY, 2026). The guide’s advice is to build a more committed champion. Our evidence is that a champion, however committed, could not on their own get the buyer’s procurement and commercial functions to act (Section 5.6).

Buyer-side constraints explain most, but not all, of the non-integrated interviewed cases. One of the seven describes constraints of the kind the barriers literature expects, upstream of anything the buyer decides (Rahman et al., 2020), and one an external shock (Asgary et al., 2020). Neither a purely supplier-side nor a purely buyer-side account fits all seven cases. For two of the seven, staff and founder give incompatible accounts of the same failure (Section 5.6).

A further extension follows from the first. Colonnelli et al. (2026) show that access to capital in Africa follows the founder’s prior connections, and that the founder’s stated preference and the firm’s quality do not predict it. Our entry result has the same form. A founder’s prior reach into the buyer’s world predicts the attention of a corporate buyer, before the programme has done anything, and a connection substitutes

for a demonstrated capability. If the two results reflect one selection process, an enterprise without that reach is disadvantaged in the market for capital and in the market for buyer attention alike. A programme that supplies capital or buyer access without supplying the connection then addresses only the consequence of that disadvantage. This is a conjecture, and it is the extension we would most like to see tested elsewhere.

A last extension concerns how the buyer accounts for sustainability. Section 5.7 reports the reason Unilever managers give for the green null, that sustainability “isn’t incorporated into a normal financial P&L,” so that a green credential is valued by the teams that fund pilots and is not part of the procurement decision. The green null therefore measures a boundary inside the buyer. Villena (2018) finds the same boundary limiting multinationals’ sustainability efforts in general, because the procurement function that contracts with suppliers is not the function that selects them on sustainability grounds. The boundary applies to the social half of the portfolio in the same terms, since the manager’s sentence covers “the social portion” as well. The programme’s own guide states the procurement side of it, that impact enterprises may need to charge a premium “while cost is always one of the key criteria reviewed by procurement” (Unilever, FCDO and EY, 2026). An impact enterprise built to deliver a social or environmental outcome has, by construction, a cost structure that a commodity supplier does not, and the programme’s support is meant to close that gap. In the cases we observe, where the gap was closed it was closed by pilot funding.

7.3 Contradiction: supplier characteristics and green credentials

The first contradiction is that no entry-stage combination is sufficient, against a literature premised on supplier characteristics determining participation, where the characteristics are capability, absorptive capacity, compliance, finance and readiness (Pleticha, 2021; Abe, 2015; Cohen and Levinthal, 1990). No combination of the conditions we can observe administratively is sufficient for integration, in either document-coded sample, under any of the five sample constructions, on any of the three definitions of integration, or with documentary silence treated as unknown (Sections 5.5 and 6). The only terms that clear the consistency cut hold one enterprise or two.

Low coverage might be read as a failure of the analysis, since no specification reaches more than three of the ten integrators in the larger sample. Full consistency with low coverage means that each sufficient term describes one or two particular integrators, and that no term describes a path the other integrators follow. That is the configurational form of the finding that these conditions do not decide integration. The conditions were the ones the supplier-side literature treats as decisive, coded from what the enterprises said about themselves at entry. We tried fourteen specifications, three definitions of the outcome and the treatment of silence as unknown before we accepted the result.

Coverage would rise with a condition set that describes the buyer, coded for every enterprise, and we lack one. The tracker’s during-project fields do describe the buyer, but they sort the outcome sharply because they record revealed demand (Section 5.6), and the interview codes that describe the buyer’s decision cannot be applied at portfolio scale. The governance typology suggests why a supplier-side set should return a null. Gereffi et al. (2005) derive governance form from transaction complexity, codifiability and supplier capability, and only the third is an attribute of the supplier. A condition set built from supplier attributes is therefore missing two of the three determinants by construction. The interviews indicate the same thing in concrete terms, since no supplier characteristic could have fixed a crop variety the buyer had not yet qualified. We examined one expectation of the same literature and set it aside. Backward-linkage theory predicts an advantage for enterprises that supply inputs (Javorcik, 2004), and chain direction is a null here with a sign that changes across samples (Appendix F.2). Type and chain position are nearly collinear in this portfolio (Section 5.7).

The second contradiction is that green enterprises do not integrate more often. On their own account of their credentials the point estimate is negative, though it is not distinguishable from zero. The environmental-upgrading literature would lead one to expect green suppliers to face distinctive barriers (Ponte, 2022b; De Marchi and Di Maria, 2019; Krishnan et al., 2023). On the programme’s own classification, enterprises funded under the environmental and the social pillars integrate within a percentage point of each other, and on the enterprises’ own documents green enterprises integrate less, 13.3 against 23.1 per cent (Section 5.7). The interviewed cases are consistent with the portfolio result. The interviews support one half of the expectation. Green credentials are mainly selected at entry and sometimes deepened during the relationship, in three of the seven green-selected cases and once at the buyer’s own request, so environmental upgrad-

ing is observed here. It is not associated with a commercial premium (Section 5.7; Appendix F.1). The programme’s funders asked us whether selecting on green credentials selects enterprises the buyer is more willing to integrate. On the evidence we have it does not, and on the enterprises’ own account of their credentials it selects enterprises the buyer is somewhat less willing to integrate. That is the pattern the cost-shifting account predicts for green suppliers who sit upstream (Ponte, 2022b), and the procurement boundary described in the previous subsection is the mechanism we propose for it.

8 Limitations

The design supports description and pattern-finding. It does not support causal inference, and several features of the data bound what even the descriptive evidence can support.

The largest limit is that the programme randomised neither which enterprises it selected nor how intensively it engaged them, and we observe no exogenous variation in any predictor. Every estimate in this paper is therefore associational. We sought a comparison group of enterprises that applied and were not selected, and under the funders’ data-protection terms those records could not be shared. The accelerator literature now has exactly that design (Assenova and Amit, 2024; Yu, 2020), and it is the benchmark against which the present evidence should be read. The selection-versus-treatment interpretation therefore rests on the pattern of associations within the observed portfolio. The claim that follows from it, that integration counts are a poor evaluation metric, is an argument that the evidence supports without establishing it.

The sample is small and the outcome is rare, since the portfolio holds ninety enterprises and fourteen integration events. We state the complete-case sample for each model beside it. With a dozen events the coefficients are imprecise, and the significance of the starting-link coefficient depends on which handful of enterprises a complete-case rule keeps. The two-way table on which the gate result rests, none of the 24 enterprises below the threshold integrating, does not depend on those choices, which is why Section 5.2 leads with it.

The programme team coded the outcome, and it is a party with an interest in the programme’s apparent effectiveness. The team compiled the whole tracker, the rating of the match at the start of each project included, in 2025 with the outcomes in view (Section 3.1). The buyer’s record of prior relationships was added in 2026, after the team had read a draft of this paper, and the enterprises’ contemporaneous applications are the independent check on both. We can check the outcome from outside the programme only for the twelve interviewed enterprises, where the founders’ accounts agree with the tracker on ten of twelve under each of two definitions. The during-project ratings have the same problem to a greater degree. They record the team’s assessment of an enterprise’s fit for the organisation as the project ran, made by the team that also coded the outcome. They therefore measure the buyer’s revealed demand, with programme effort not separately recorded, which is why the treatment-at-the-margin claim cannot be tested here.

Several variables are measured with error. The link predictors are ordinal labels treated as continuous scores. The match rating and the enterprises’ own record of a relationship agree on 48 per cent of the cases we hold both for ($\kappa = 0.06$), which reflects a difference of construct as much as measurement error (Section 3.2). The entry-stage conditions are known least well for the oldest enterprises, since first-generation folders far more often hold an outcome without an application document. The green constructs under-capture agricultural certification on two integrated upstream suppliers, a bias that runs against the green category and cannot produce the null we report.

Each method then has its own cost. In the regression, the events-per-variable ratio is far below the conventional floor, we did not pre-register a specification, and the significance levels are not corrected for the number of models examined. In the configurational analysis, twelve cases are at the small end of what crisp-set methods support (Schneider and Wagemann, 2012), every model runs at three conditions, the surviving terms rest on one or two cases, and calibration and thresholds are analyst choices. `PRIORITY_FIT` is conceptually close to the outcome.

The machine-assisted coding behind the conditions is fast, consistent and documented. On the interviewed cases a second reader verified the low-confidence cells and a random sample of the rest against the source documents, confirming 48 of the 54 cells checked and revising six (an agreement of 88.9 per cent). The revisions shift the terms and coverage of some robustness specifications but leave the necessity findings, the sufficiency null and the hypothesis verdicts intact. That check was not blind, so it yields no inter-rater statistic, and the codes for the other document-coded enterprises were not verified by hand (Appendix B).

The interview evidence is limited by the sequence of the fieldwork, by single coding and by access. Unilever’s data reached us in stages over the fieldwork and the schedule moved with them, so later founders were asked about mechanisms the earlier ones were not. That is one reason an absent code records only that a mechanism was not raised (Appendix E). One researcher coded the nineteen interviews, typically one informant per enterprise, with no reliability statistic, and the one condition built from them is retrospective. On access, four enterprises named in the inception design are outside the case set because no interview could be secured. Introductions came through the programme team, so the set is confined to enterprises

the programme could still reach. The interviewed enterprises over-represent integrators and enterprises that arrived with a High match, and they under-represent the local referral route. Retrospective accounts may also favour the founder, since a founder explaining a stalled relationship may attribute it to the buyer's process. That Unilever staff describe the same friction from the other side is some reassurance, and where the two sides disagree about a case we have reported the disagreement without resolving it.

Two further qualifications concern the study as a whole. The design was iterative and partly inductive. Several relationships we report, the localisation result among them, were arrived at through that cycle without being posited in advance, so we present them as hypothesis-generating. One mechanism we set out to examine, the non-financial support each enterprise received, could not be answered from programme records. A second, equity, does not arise, because the programme takes no equity in the enterprises it supports.³ Finally, all of the above concerns a single buyer and programme, with one funder consortium and a portfolio concentrated in sub-Saharan Africa and South Asia. The mechanisms we describe, especially the dependence of integration on a senior commercial sign-off inside the buyer's business unit, are plausibly specific to this lead firm's procurement and decision structures.

³The interviews describe one exclusivity term, a right of first refusal, among the twelve, and no equity arrangement; the non-financial support each enterprise received (coaching, consulting, introductions) is recorded neither in the tracker nor systematically in the project documents, so its intensity cannot be compared across enterprises.

9 Conclusion

This paper asked what the number of small firms that integrate into a multinational’s chain measures, when that number is read as a measure of what a supplier-development programme achieved. We examined the question using the internal record of TRANSFORM, a decade-long partnership between Unilever, the FCDO and EY in the form of an impact accelerator. We framed it as a question about selection and treatment, that is, whether integration is determined mainly by what an enterprise brings to the programme or by what subsequently happens within the programme and the buyer organisation. All three hypotheses of Section 2.7 are supported. In the programme team’s record, none of the 24 enterprises whose project offered Unilever nothing it could use at entry integrated, against 14 of the 66 whose project did. Integration is unassociated with time in the programme, and a prior relationship with the buyer raises the rate above that threshold without being necessary for it. No combination of the characteristics recorded at entry is sufficient for integration. Interviewees attribute the outcome to decisions inside the buyer, where a contract needed several functions with unrelated objectives to agree at the same time and the champion the programme produced could not on their own get procurement to act (Section 5.6). The benefits founders most often report from participation are reported whether or not the enterprise integrated, and most often by the enterprises that did not (Section 5.10).

The findings bear on how programmes of this kind are run and how they are judged. An integration count records the lead firm’s procurement decisions at least as much as anything the programme did, and it adds together relationships of different kinds, so a headline integration rate is not a measure of programme performance alone. Enterprises report that the programme gave them capability, credibility with other buyers and access to markets beyond Unilever. These benefits are measurable and are reported whether or not the enterprise integrated, though on the strict counterfactual test few of them would not otherwise have existed. A programme should be judged on the benefits its participants keep, measured on both sides of the integration line, with integration reported as one outcome among them.

The evidence suggests that selection panels should establish, before an enterprise is admitted, a named commercial owner inside the buyer with the authority to sign a supply or distribution agreement, since the champions the programme produced, however committed, could not on their own get procurement and the commercial functions to act (Section 5.6). The programme’s own guide concludes that committed project leads were its single most important driver of success. Our evidence is that commitment was not enough without that authority. Where a programme selects on environmental or social credentials, the same reasoning suggests pairing the credential with a procurement specification the buyer’s commercial function has already agreed to buy against, because the credential alone influences sustainability funding without affecting the commercial procurement decision. The programme’s tracker should record the match and any prior relationship at entry, from both sides and at the time. It should keep its within-project fields prospective, so that a later evaluation can separate demand from effort, and it should record which functions inside the buyer signed and when. The monitoring instrument should separate the jobs the programme claims from the enterprise’s headcount. Finally, the records of unsuccessful applicants should be retained, under whatever data-protection terms the funders can agree, since that would do more for the evaluability of the instrument than any other single change.

The paper makes three contributions to the literature. The first is a portfolio-scale null on the supplier-side conditions the enabling-conditions tradition treats as decisive, with a reason from the governance typology for why such a null should be expected. The second is evidence that commercial integration turns on coordination across functions, budgets and individuals within the buyer, an extension we have not seen made in the value-chain literature. The third is a measured integration gap between green and social suppliers inside one buyer’s portfolio, which shows that impact credentials do not by themselves convert into commercial integration. The design choice that proved most useful was to read entry from more than one record, the buyer’s rating of the match, its record of prior relationships, the enterprise’s contemporaneous application and the founder’s own history, and to treat the interview coding as data. The comparison across records separated two things the tracker’s own label runs together, whether the buyer could use the project and whether it already knew the enterprise, and it showed that the first is the threshold and the second raises the rate above it. The interview coding let the interviews explain the cases no configuration reaches. The paper has three main limitations. It offers no causal identification, and every estimate is associational. The fourteen integration events and twelve interviewed cases are few, and the outcome and interview coding are

retrospective and single-coder. Everything here concerns a single buyer, programme and funder consortium.

Each of these limitations could be addressed by further work. The most important is a counterfactual comparison. Comparing admitted enterprises with those that applied and were refused would establish what an integration count is worth in a way no within-portfolio analysis can. Reading a second buyer’s supplier-development record the same way would establish whether these internal organisational constraints are specific to Unilever. The entry gate could be tested more cheaply by coding the founder’s prior international reach for all ninety enterprises and matching the portfolio to the deal records of Colonnelli et al. (2026). Coding the milestone and final reports for the benefits the interviews identify would show whether benefit without integration is the common pattern across the portfolio. Following the relationships that lapsed would show whether suppliers keep what a multinational relationship built. A researcher embedded in the buyer’s country business and the programme team for the life of a project could observe the integration decision while it is being made, and record the mechanism this paper infers from interviews, including for the enterprises that do not integrate. An integration count is easy to collect and easy to report. What the programme produced lies mostly outside that number, in the capabilities, credibility and markets its participants kept whether or not the buyer ever bought from them, and a programme that wants to know what it achieved should count those too.

CRedit author statement

Harman: conceptualisation, methodology, software, validation, formal analysis, investigation, data curation, writing (original draft), writing (review and editing), visualisation.

Shukla: conceptualisation, methodology, software, validation, formal analysis, investigation, data curation, writing (original draft), writing (review and editing), visualisation, project administration, funding acquisition.

Data availability

The data underlying this paper were shared by the TRANSFORM programme under confidentiality terms. A pseudonymised extract of the programme’s portfolio tracker, cleared for release by the programme, is available with a codebook in the replication deposit that accompanies this paper. It holds two files, one row for each of the 150 funded projects and one for each of the 90 enterprises in the portfolio sample, keyed by the pseudonyms under which enterprises appear in the text and carrying the tracker variables this paper uses together with the research team’s coding of chain direction. Enterprise and project names, staff names, countries and every free-text field have been removed and dates are coarsened to years; region is kept, because it is needed to check the chain-direction and localisation results. The files reproduce the outcome distribution of Table 4, the entry gate, the referral origins of Table 6 and the cohort comparison of Section 6. No other TRANSFORM material can be shared. That covers the tracker as received, the project documents (applications, milestone reports and final reports) and the monitoring data, whose totals are enterprise-identifying at these magnitudes. The analysis code will be shared with the deposit. The interview transcripts cannot be shared; the coded interview matrix will be shared at the level of codes by pseudonym. The interview schedule, with the evidence behind each question, is Supplementary Appendix S1, a separate file.

References

- Abe, Masato**, “SME Participation in Global Value Chains: Challenges and Opportunities,” in Asian Development Bank and Asian Development Bank Institute, ed., *Integrating SMEs into Global Value Chains: Challenges and Policy Actions in Asia*, Manila and Tokyo: Asian Development Bank and Asian Development Bank Institute, 2015, pp. 27–65.
- Alfaro-Ureña, Alonso, Isabela Manelici, and Jose P Vasquez**, “The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages*,” *The Quarterly Journal of Economics*, 01 2022, *137* (3), 1495–1552.
- Amiti, Mary, Cédric Duprez, Jozef Konings, and John Van Reenen**, “FDI and superstar spillovers: Evidence from firm-to-firm transactions,” *Journal of International Economics*, 2024, *152*, 103972.
- Archer, Matthew and Hannah Elliott**, “‘It’s up to the market to decide’: Revealing and concealing power in the sustainable tea supply chain,” *Critique of Anthropology*, 2021, *41* (3), 227–246.
- Arora, Amit, Anshu Arora, Julius Anyu, and John R. McIntyre**, “Global Value Chains’ Disaggregation through Supply Chain Collaboration, Market Turbulence, and Performance Outcomes,” *Sustainability*, 2021, *13* (8).
- Asgary, Ali, Ali Ihsan Ozdemir, and Hale Özyürek**, “Small and Medium Enterprises and Global Risks: Evidence from Manufacturing SMEs in Turkey,” *International Journal of Disaster Risk Science*, February 2020, *11* (1), 59–73.
- Assenova, Valentina A. and Raphael Amit**, “Poised for Growth: Exploring the Relationship Between Accelerator Program Design and Startup Performance,” *Strategic Management Journal*, 2024.
- Atkin, David, Amit K. Khandelwal, and Adam Osman**, “Exporting and Firm Performance: Evidence from a Randomized Experiment,” *Quarterly Journal of Economics*, 2017, *132* (2), 551–615.
- Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts**, “Does Management Matter? Evidence from India,” *Quarterly Journal of Economics*, 2013, *128* (1), 1–51.
- Bolwig, Simon, Peter Gibbon, and Sam Jones**, “The Economics of Smallholder Organic Contract Farming in Tropical Africa,” *World Development*, 2009, *37* (6), 1094–1104.
- Brambilla, Irene, Daniel Lederman, and Guido Porto**, “Exports, Export Destinations, and Skills,” *American Economic Review*, 2012, *102* (7), 3406–3438.
- Brix-Asala, Carolin, Stefan Seuring, Philipp C. Sauer, and Axel Zehendner**, “Resolving the Base of the Pyramid Inclusion Paradox through Supplier Development,” *Business Strategy and the Environment*, 2021.
- Brucal, Arlan, Beata Javorcik, and Inessa Love**, “Good for the environment, good for business: Foreign acquisitions and energy intensity,” *Journal of International Economics*, 2019, *121*, 103247.
- Bruton, Garry D., Jorge H. Mejia-Morelos, and David Ahlstrom**, “Multinational corporations and inclusive supply chains: how conflicting social and market logics can coexist,” *Journal of World Business*, 2025, *60* (6), 101663.
- Carballo, Jeronimo, Ignacio Marra de Artiñano, Gianmarco IP Ottaviano, and Christian Volpe Martincus**, “Linkages with multinationals: the effects on domestic firms’ exports,” 2021.
- Cohen, Susan L., Daniel C. Fehder, Yael V. Hochberg, and Fiona Murray**, “The Design of Startup Accelerators,” *Research Policy*, 2019, *48* (7).
- Cohen, Wesley M. and Daniel A. Levinthal**, “Absorptive Capacity: A New Perspective on Learning and Innovation,” *Administrative Science Quarterly*, 1990, *35* (1), 128–152.

- Colonnelli, Emanuele, Marcio Cruz, Mariana Pereira-Lopez, Tommaso Porzio, and Chun Zhao,** “Startups in Africa,” NBER Working Paper 35261, National Bureau of Economic Research 5 2026. Revised August 2026.
- Crescenzi, Riccardo and Oliver Harman,** *Harnessing global value chains for regional development: How to upgrade through regional policy, FDI and trade*, Routledge, 2023.
- and —, “Greening Global Value Chains,” 2026. Cambridge Elements, forthcoming.
- , **Arnaud Dyèvre, and Frank Neffke,** “Innovation Catalysts: How Multinationals Reshape the Global Geography of Innovation,” *Economic Geography*, 2022, 98 (3), 199–227.
- Creswell, John W. and Vicki L. Plano Clark,** *Designing and Conducting Mixed Methods Research*, 2 ed., Thousand Oaks, CA: SAGE Publications, 2011.
- and —, *Designing and Conducting Mixed Methods Research*, 3 ed., Thousand Oaks, CA: SAGE Publications, 2018.
- DFID and Unilever,** “British Government and Unilever join forces to help world’s poor,” Press release, 22 September 2014, GOV.UK 2014. Accessed 4 September 2026.
- Drori, Israel, Benson Honig, and Mike Wright,** “Transnational Entrepreneurship: An Emergent Field of Study,” *Entrepreneurship Theory and Practice*, 2009, 33 (5), 1001–1022.
- Dusa, Adrian,** *QCA with R: A Comprehensive Resource*, Cham: Springer, 2019.
- Filatotchev, Igor, Xiaohui Liu, Trevor Buck, and Mike Wright,** “The Export Orientation and Export Performance of High-Technology SMEs in Emerging Markets: The Effects of Knowledge Transfer by Returnee Entrepreneurs,” *Journal of International Business Studies*, 2009, 40 (6), 1005–1021.
- Foster, Christopher, Mark Graham, Laura Mann, Timothy Waema, and Nicolas Friederici,** “Digital Control in Value Chains: Challenges of Connectivity for East African Firms,” *Economic Geography*, 2018, 94 (1), 68–86.
- Gall, Markus, Michael Wiener, Cecilia Chagas de Oliveira, Reinhold W. Lang, and Erik G. Hansen,** “Building a Circular Plastics Economy with Informal Waste Pickers: Recyclate Quality, Business Model, and Societal Impacts,” *Resources, Conservation and Recycling*, 2020, 156, 104685.
- Gereffi, Gary, John Humphrey, and Timothy Sturgeon,** “The Governance of Global Value Chains,” *Review of International Political Economy*, 2005, 12 (1), 78–104.
- Hennart, Jean-François,** “More than intent: A bundling model of MNE–SME interactions: Jean-François Hennart,” *Journal of International Business Studies*, 2020, 51 (7), 1176–1194.
- Heyman, Fredrik, Fredrik Sjöholm, and Patrik Gustavsson Tingvall,** “Is there really a foreign ownership wage premium? Evidence from matched employer–employee data,” *Journal of International Economics*, 2007, 73 (2), 355–376.
- Horváth, Dóra and Roland Zs. Szabó,** “Driving forces and barriers of Industry 4.0: Do multinational and small and medium-sized companies have equal opportunities?,” *Technological Forecasting and Social Change*, 2019, 146, 119–132.
- Huang, Ying, Yadong Luo, Yi Liu, and Qian Yang,** “An Investigation of Interpersonal Ties in Interorganizational Exchanges in Emerging Markets,” *Journal of Management*, 2016, 42 (6).
- Iacovone, Leonardo, Beata Javorcik, Wolfgang Keller, and James Tybout,** “Supplier Responses to Walmart’s Invasion in Mexico,” *Journal of International Economics*, 2015, 95 (1), 1–15.
- Javorcik, Beata Smarzynska,** “Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages,” *American Economic Review*, 2004, 94 (3), 605–627.

- Jensen, Federico and Lindsay Whitfield**, “Leveraging participation in apparel global supply chains through green industrialization strategies: Implications for low-income countries,” *Ecological Economics*, 2022, *194*, 107331.
- Juergensen, Jill, José Guimón, and Rajneesh Narula**, “European SMEs amidst the COVID-19 crisis: assessing impact and policy responses,” *Journal of Industrial and Business Economics*, September 2020, *47* (3), 499–510.
- Keijser, Charlotte, René Belderbos, and Micheline Goedhuys**, “Governance and learning in global, regional, and local value chains: The IT enabled services industry in South Africa,” *World Development*, 2021, *141*, 105398.
- Koleša, Sabina**, “Global Value Chains: Government Policies for Enhancing the Role of Small and Medium Enterprises in Global Value Chains-A Case Study of Slovenia,” *Management (18544223)*, 2018, *13* (1).
- Krishnan, Aarti, Valentina De Marchi, and Stefano Ponte**, “Environmental Upgrading and Downgrading in Global Value Chains: A Framework for Analysis,” *Economic Geography*, 2023, *99* (1), 25–50.
- Li, Cheng and Alvaro Cuervo-Cazurra**, “How subsidiary and supplier misbehavior lead to corporate social responsibility performance improvements in multinationals,” *Journal of International Business Studies*, June 2024, *55* (4), 470–491.
- Macchiavello, Rocco and Ameet Morjaria**, “The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports,” *American Economic Review*, 2015, *105* (9), 2911–2945.
- Marchi, Valentina De and Eleonora Di Maria**, “Environmental Upgrading and Suppliers’ Agency in the Leather Global Value Chain,” *Sustainability*, 2019, *11* (23), 6530.
- **and Matthew Alford**, “State policies and upgrading in global value chains: A systematic literature review,” *Journal of International Business Policy*, March 2022, *5* (1), 88–111.
- Marx, Axel and Adrian Dusa**, “Crisp-Set Qualitative Comparative Analysis (csQCA), Contradictions and Consistency Benchmarks for Model Specification,” *Methodological Innovations Online*, 2011, *6* (2), 103–148.
- Masso, Jaan and Priit Vahter**, “Joining and exiting the value chain of foreign multinationals and performance of their local suppliers: evidence from interfirm transaction data,” *Industrial and Corporate Change*, 01 2023, *32* (4), 848–875.
- McCullagh, Peter**, “Regression Models for Ordinal Data,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 1980, *42* (2), 109–142.
- Meqdadi, Osama, Thomas E. Johnsen, and Mark Pagell**, “Relationship configurations for procuring from social enterprises,” *International Journal of Operations & Production Management*, 2020, *40* (6), 819–845.
- Narula, Rajneesh**, “Enforcing higher labor standards within developing country value chains: Consequences for MNEs and informal actors in a dual economy,” *Journal of International Business Studies*, December 2019, *50* (9), 1622–1635.
- Oliveira, Luis, Afonso Fleury, and Maria Tereza Fleury**, “Digital power: Value chain upgrading in an age of digitization,” *International Business Review*, 2021, *30* (6), 101850.
- Orlic, Edvard, Iraj Hashi, and Mehtap Hisarciklilar**, “Cross sectoral FDI spillovers and their impact on manufacturing productivity,” *International Business Review*, 2018, *27* (4), 777–796.
- Paus, Eva A. and Kevin P. Gallagher**, “Missing Links: Foreign Investment and Industrial Development in Costa Rica and Mexico,” *Studies in Comparative International Development*, March 2008, *43* (1), 53–80.

- Pleticha, Petr**, “Who Benefits from Global Value Chain Participation? Does Functional Specialization Matter?,” *Structural Change and Economic Dynamics*, 2021, 58, 291–299.
- Ponte, Stefano**, “The Hidden Costs of Environmental Upgrading in Global Value Chains,” *Review of International Political Economy*, 2022, 29 (3), 818–843.
- , “The hidden costs of environmental upgrading in global value chains,” *Review of International Political Economy*, 2022, 29 (3), 818–843.
- Ragin, Charles C.**, *Redesigning Social Inquiry: Fuzzy Sets and Beyond*, Chicago: University of Chicago Press, 2008.
- Rahman, Mahfuzur, Mansura Akter, Kenny Odunukan, and Shamim Ehsanul Haque**, “Examining economic and technology-related barriers of small- and medium-sized enterprises internationalisation: An emerging economy context,” *Business Strategy & Development*, 2020, 3 (1), 16–27.
- Reardon, Thomas, Lenis Saweda O. Liverpool-Tasie, and Bart Minten**, “Quiet Revolution by SMEs in the midstream of value chains in developing regions: wholesale markets, wholesalers, logistics, and processing,” *Food Security*, December 2021, 13 (6), 1577–1594.
- Schneider, Carsten Q. and Claudius Wagemann**, *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*, Cambridge: Cambridge University Press, 2012.
- **and Ingo Rohlfing**, “Combining QCA and Process Tracing in Set-Theoretic Multi-Method Research,” *Sociological Methods & Research*, 2013, 42 (4), 559–597.
- **and —**, “Set-theoretic Multimethod Research: The Role of Test Corridors and Conjunctions for Case Selection,” *Swiss Political Science Review*, 2019.
- Seitz, Nikolaus et al.**, “Corporate Accelerators: Design and Startup Performance,” *Small Business Economics*, 2023.
- Shankar, Raj K. and Dean A. Shepherd**, “Accelerating Strategic Fit or Venture Emergence: Different Paths Adopted by Corporate Accelerators,” *Journal of Business Venturing*, 2019, 34 (5), 105886.
- Shinozaki, Shigehiro**, “Financing SMEs in Global Value Chains,” in Asian Development Bank and Asian Development Bank Institute, ed., *Integrating SMEs into Global Value Chains: Challenges and Policy Actions in Asia*, Manila and Tokyo: Asian Development Bank and Asian Development Bank Institute, 2015, pp. 66–100.
- Steenbergen, Victor and Abhishek Saurav**, “The effect of multinational enterprises on climate change: Supply chain emissions, green technology transfers, and corporate commitments,” 2023.
- Tang, Christopher S.**, “Socially responsible supply chains in emerging markets: some research opportunities,” *Journal of Operations Management*, 2018, 57, 1–10.
- Taylor, Kelsey M. and Eugenia Rosca**, “Sink, swim, or drift: how social enterprises use supply chain social capital to balance tensions between impact and viability,” *Journal of Supply Chain Management*, 2023, 59 (2), 62–86.
- TRANSFORM**, “UK’s Department for International Development, Unilever, and the Clinton Giustra Enterprise Partnership join forces to improve health and well-being for 100 million people in Africa and South Asia,” Press release, 25 September 2015, PR Newswire 2015. Accessed 4 September 2026.
- , “TRANSFORM: funding programme to impact enterprises, South Africa and Zambia,” Regional challenge call, as reproduced by fundsforNGOs 2024. Accessed 4 September 2026.
- Unilever, FCDO and EY**, “How to collaborate for impact: a practical guide to building and managing successful public-private partnerships. Insights from 10+ years of TRANSFORM,” TRANSFORM play-book, launched at the FCDO Global Partnership Conference, May 2026 2026. Accessed 10 September 2026.

- Vandenberg, Paul, Naoyuki Yoshino, Akira Goto, Patarapong Intarakumnerd, and Jeffrey Miyamoto**, “Policies to Enhance SME Internationalization,” in Asian Development Bank and Asian Development Bank Institute, ed., *Integrating SMEs into Global Value Chains: Challenges and Policy Actions in Asia*, Manila and Tokyo: Asian Development Bank and Asian Development Bank Institute, 2015, pp. 117–143.
- Venables, William N. and Brian D. Ripley**, *Modern Applied Statistics with S*, fourth ed., New York: Springer, 2002.
- Villena, Verónica H.**, “The Missing Link? The Strategic Role of Procurement in Building Sustainable Supply Networks,” *Production and Operations Management*, 2018.
- World Bank**, *World Development Report 2020: Trading for Development in the Age of Global Value Chains*, Washington, DC: World Bank, 2020.
- Xiao, Chengyong, Miriam Wilhelm, Taco van der Vaart, and Dirk Pieter van Donk**, “Inside the Buying Firm: Exploring Responses to Paradoxical Tensions in Sustainable Supply Chain Management,” *Journal of Supply Chain Management*, 2018.
- Xiao, Ziang, Xingdi Yuan, Q. Vera Liao, Rania Abdelghani, and Pierre-Yves Oudeyer**, “Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding,” in “Companion Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI ’23 Companion)” Association for Computing Machinery New York, NY, USA 2023, pp. 75–78.
- Yu, Sandy**, “How Do Accelerators Impact the Performance of High-Technology Ventures?,” *Management Science*, 2020, 66 (2).
- Zhang, Chun, Sridhar Viswanathan, and John W. Henke**, “The Boundary Spanning Capabilities of Purchasing Agents in Buyer–Supplier Trust Development,” *Journal of Operations Management*, 2011, 29 (4), 318–328.

A Data, samples and design

A.1 The sequencing of the design

In the terms of Creswell and Plano Clark (2018) the design is primarily exploratory sequential. Early interviews helped develop the constructs and QCA conditions, and later interviews helped explain and test the patterns emerging from the quantitative and configurational analyses. Because the stages informed one another, the study is a multiphase design (Creswell and Plano Clark, 2011), whereas a convergent design would have run the strands independently. The phased release of Unilever’s data determined its sequencing (Figure 2) as much as any prior plan did.

The design follows the three-step build set out in the project’s Inception Report of October 2025, in which a literature review establishes the constructs, portfolio data test them, and interviews validate and extend them. It retains that report’s nested comparative design for the case component. It departs from the report in three respects. The portfolio-scale tiers were added after Unilever released the integration tracker. The machine-assisted coding pipeline was built to address a bottleneck the inception did not foresee, and then scaled to the document-coded sample. A comparison group of unsuccessful applicants could not be obtained under the funders’ data-protection terms, and that is the single largest constraint on what the design can claim (Section 8).

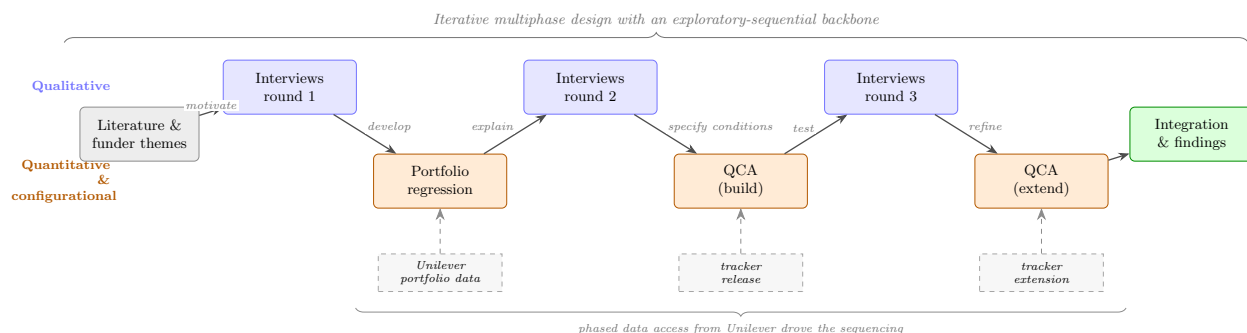


Figure 2: Sequencing of the mixed-methods design

Notes: qualitative phases (interviews) and quantitative and configurational phases (portfolio regression, QCA) alternate. Early interviews develop the constructs, QCA conditions and hypotheses that the structured analyses operationalise, an exploratory-sequential logic (qual → QUAN); later interviews explain and test the resulting patterns (QUAN → qual). The sequencing was shaped by the phased release of Unilever portfolio data and was not fixed in advance.

A.2 Who is in the sample

Section 3.3 states the three decisions that determine the population, and this subsection records them in full, beginning with the arithmetic. The tracker’s sheet holds 168 rows, of which one is a sub-header and seventeen name no organisation, leaving 150 project rows across 109 distinct organisations. Merging one organisation that appears under two spellings of its name gives 108, and removing the eighteen partner organisations gives the 90 supported enterprises. The eighteen partner rows are excluded by a ruling about the organisation, recorded in the project registry and applied to every tracker-based analysis alike.⁴ We merge the one enterprise that appears under two spellings of its name. We code an enterprise with a blank in the March 2026 integration field as not integrated, which is what the programme lead confirmed a blank records, and 35 enterprises are in that position.

⁴Four universities; seven NGOs or implementing organisations, one of which appears twice; three research or consultancy bodies; two water utilities; and one sanitation programme. Recycling M, which a name pattern flags as an NGO, was reviewed and retained as a portfolio enterprise. The three tracker rows carried under the parent organisation are kept as three enterprises in the portfolio frame, while the interviewed case Refill B is that parent.

A.3 Chain direction as a constructed variable

Section 3.3 names the direction variable. It classifies each enterprise as upstream or downstream. An upstream enterprise supplies inputs that feed into Unilever’s manufactured products, such as recycled materials, agricultural commodities and EPR credits. A downstream enterprise distributes, sells or otherwise gets Unilever’s branded products to market, and the category includes refill stations, last-mile distributors and digital commerce platforms. We built the variable from enterprise profiles in TRANSFORM publications, Unilever press releases and enterprise websites, and verified it against project proposals where available. Of the 90 enterprises in the portfolio sample, 39 are upstream and 51 downstream. Every enterprise could be classified, and none mixes the two across its projects.

B The coding pipeline

This appendix sets out the three stages of the machine-assisted pipeline of Section 4.2.3. It gives worked examples of how each condition was coded from the enterprises’ documents, reports what restricting the evidence to application-stage documents changed, and states the pipeline’s limits. The reliability checks, which are the agreement between tiers, the agreement between full-corpus and application-stage coding, and the share of low-confidence cells, are in Appendix C.

B.1 The three stages

The first stage extracts evidence from each firm’s application, milestone and final-report documents, whose text is extracted programmatically (PDF, DOCX and XLSX alike). A condition-specific dictionary of regular expressions then surfaces candidate evidence passages. The dictionary holds Unilever-partnership language for `UL_LINK_START`, Unilever brand names and distribution terms for `DOWNSTREAM`, impact-theme and entry-year markers for `PRIORITY_FIT`, and certification names and green-sector terms for `GREEN`. Each match is returned as an excerpt with ± 200 characters of surrounding context and the matched term bracketed for legibility, and near-duplicate excerpts are removed. The output of this stage is an auditable, per-firm workbook of extracted passages organised by condition. The evidence passed to the model is restricted to documents classified as application-stage, so that a condition defined at the point of application cannot be satisfied by evidence of what happened afterwards. The comparison below measures what that restriction changed.

In the second stage, the extracted evidence is passed, one firm and one condition at a time, to a large language model (Claude Sonnet 4.6, Anthropic) together with a calibration rubric for that condition, and the model proposes a code. The rubric states the condition’s theoretical basis, an explicit code= 1 rule, and an explicit code= 0 rule. For `PRIORITY_FIT` it also supplies the priority-intensity matrix, a pillar glossary, and a regime rationale. The model is instructed to code the condition using only the supplied evidence and not to invent evidence. It returns a structured JSON object with the suggested binary code, a confidence rating (high / medium / low), one to three sentences of reasoning, the single most decisive verbatim quote, and a free-text field for data gaps or alternative interpretations. For `PRIORITY_FIT` the object also exposes the intermediate matrix lookup, which is the inferred entry year, the identified pillars, the regime and the intensity. All calls are issued at temperature 0 for reproducibility, and the full prompt-and-response exchange is archived to a JSON log for audit. The entry-stage run issued 96 calls on the interviewed cases and 56 on the document-coded tier, and the remaining document-coded cells carried forward unchanged from the full-corpus run.⁵

In the third stage a second member of the research team verifies the suggested codes against the underlying source documents and records any overrides. On the twelve interviewed enterprises the verification covered every cell the model returned at low confidence, 19 of the 96, all of them coded 0, and a random sample of 35 of the 77 cells it returned at medium or high confidence, 54 cells in all. The sample was drawn with a fixed seed so that the draw can be reproduced. The verifier confirmed 48 of the 54 codes and changed six, an agreement of 88.9 per cent, with 18 of the 19 low-confidence cells confirmed and 30 of the 35 medium- and high-confidence cells confirmed. The six changes are the CODIFIED codes for Refill C and Refill A and the

⁵Estimated API cost of the entry-stage run: approximately USD 2.

SCALE codes for Refill A and Recycling B, each raised from 0 to 1, the DOWNSTREAM code for Refill B, raised to 1, and the PRIORITY_FIT code for Agri C, lowered to 0. The DOWNSTREAM change restates a ruling already recorded in Table 13, so five of the six alter the working dataset. The revisions move the parsimonious solutions and coverage of several deep-tier robustness specifications reported in Appendix C. The baseline now returns no solution, and the terms or coverage of M3, M5, M6 and M7 change, but no necessity conclusion, the sufficiency null, or any hypothesis verdict changes. The random sample is there so that the check says something about the cells the model did not flag as well as the ones it did. The verifier could see the model’s code, so the check measures agreement with the source documents and is not blind double coding. An inter-rater statistic would need a second coder working without the first code. The other human rulings are the two DOWNSTREAM overrides recorded in Table 13 and the UL_BUYIN coding of Appendix C. The codes for the document-coded enterprises outside the interviewed cases were not verified by hand.

B.2 Worked examples

The rules above require a substantive relationship, so a textual mention of a buyer or a credential does not satisfy them. For each of the four conditions we have chosen cases in which a surface reading of the documents would have produced the opposite code, since these show the rules at work. We give one positive and one negative case per condition, drawn from the twelve deep cases and quoting the evidence the coder cited.

UL_LINK_START. Platform B is coded 1. Its proposal records that “[Platform B] has established strong working relationships with Unilever [in its market] and [a named distributor in a named county] who will collaborate on the Platform B pilot,” alongside a grant agreement effective February 2022 and named counterparts engaged in joint decisions. Recycling B is coded 0 despite its proposal referring repeatedly to Unilever. Every reference is prospective, since the proposal “seeks to collaborate,” describes “early-stage engagement,” and aims to “secure long-term purchase agreements”. A second reason for the code is that the one named Unilever individual appears only in the proposal form’s Pillar Lead field, which is an administrative role within the programme.

The other three conditions are coded the same way; the archived version of this appendix gives an example for each, and the calibration rubrics are in the pipeline’s configuration file.

B.3 Full-corpus versus application-stage coding

Every condition in this study is defined at the point of application, but in an earlier pass each was coded from the enterprise’s whole document folder, which holds milestone and closeout reports as well as the proposal. A condition could therefore be satisfied by evidence describing what happened after entry. In one case an enterprise was coded as holding a codified input specification on the strength of a closeout report recording that material had been supplied to a named buyer at food grade. The report records a later achievement, while the condition concerns a specification held at application.

We re-coded the sample from baseline documents alone and retain both codings, because they answer different questions. The baseline-only coding uses what was knowable at application, so it is the only version that can support a claim about selection, and it is the one we report throughout. The full-corpus coding uses the whole record and describes what distinguishes the enterprises that integrated. It describes integrators after the event and cannot serve as a predictor, because the later reports record the outcome the conditions are meant to predict.

The difference between the two codings measures how much hindsight was entering our coding. Twenty condition cells move from 1 to 0 when the later documents are withheld, against eleven moving the other way, so the full record credits more enterprises with satisfying entry-stage conditions than their own applications support. The movement is largest on UL_LINK_START, the condition this paper relies on most, whose integration gap falls from 14.3 to 8.0 percentage points. Two contrasts, GREEN and DOWNSTREAM, grow when the later documents are withheld, as discussed below the table. The measurement is a property of our coding procedure and is uninformative about the programme’s own records.

The movement is not uniform across conditions, and Table 12 shows where it concentrates. Agreement between the two codings runs from 0.84 to 0.93 and κ from 0.66 to 0.83, so the two are far from independent.

Table 12: Agreement between full-corpus and entry-stage codes ($N = 56$)

Condition	Share 1, full	Share 1, entry	Agreement	κ	$1 \rightarrow 0$	$0 \rightarrow 1$
UL_LINK_START	0.500	0.393	0.857	0.714	7	1
DOWNSTREAM	0.214	0.179	0.929	0.774	3	1
PRIORITY_FIT	0.607	0.625	0.839	0.660	4	5
GREEN	0.607	0.536	0.893	0.782	5	1
SCALE	0.286	0.321	0.929	0.831	1	3

Notes: CODIFIED, CERTIFIED and GREEN_SECTOR were not coded on the full corpus.

UL_LINK_START, the condition this paper relies on most, nonetheless has both the lowest agreement and the most one-sided movement, with seven enterprises losing the code when the later documents are withheld against one gaining it. PRIORITY_FIT is the one condition that moves in both directions in roughly equal measure, which is the pattern a coding difference with no hindsight component would produce. Two contrasts grow when the later documents are withheld, GREEN from flat to clearly negative and DOWNSTREAM to nearly double its size. Measurement error alone would not produce that pattern, since noisier codes would shrink every contrast toward zero.

B.4 Limits of the pipeline

The pipeline has three weaknesses that follow from its design. First, the model only sees what the first-stage regular expressions surface, so evidence that no keyword matches is invisible to it, and the pipeline’s recall is limited to what the signal dictionary matches. Second, temperature 0 makes the coding repeatable without making it accurate, since an identical call returns an identical code and a confidently wrong code is reproduced as faithfully as a correct one. Third, the “code only on the supplied evidence” instruction reduces but does not eliminate the risk of the model reading membership into ambiguous text. Human verification in the third stage addresses all three. On the interviewed cases it confirmed 48 of the 54 cells checked and changed six, which moved several robustness specifications without overturning any necessity conclusion, the sufficiency null or a hypothesis verdict. The model provides a fast and documented first reading of a large document corpus, and the human check concentrates on the cells where that reading is least secure. Section 6 reports how much rests on them.

C The configurational analysis

This appendix reports the configurational analysis in full. It covers the conditions that enter only the robustness specifications, every model on both tiers, necessity with a relevance check, the sensitivity of the deep-tier results to the definition of integration, and the reliability checks on the coding.

C.1 Robustness conditions

Five conditions are used only in the robustness specifications. All are coded from application-stage documents by the pipeline of Section 4.2.3 except UL_BUYIN, which is read from the tracker.

SCALE is a demonstrated or specified route to volume at application. It is satisfied by existing throughput at commercial volume, a quantified expansion plan naming geographies and unit counts, replication already achieved across more than one site, or capital already committed to expansion. A stated aspiration to scale does not qualify. It is present for 4 of 12 deep and 19 of 56 document-coded enterprises.

SCALE records a route to volume. Because an enterprise may be too small to integrate whatever route it can describe, we attempted a size-based alternative built from the headcount each enterprise states in its own application. It cannot be built, because the programme’s proposal templates have no employee-count field. They ask for organisational structure in prose, and most applicants answer without a number. Searching every entry-stage document in both tiers under patterns naming the enterprise’s own staff returned candidates for 16 of the 56 document-coded enterprises. Four of these were rejected on inspection, as a savings-group network of 3,000 counted as a payroll, a partner organisation’s staff attributed to the applicant, a percentage

read as a headcount, and a statistic about Asian firms in general. A headcount is therefore established for 12 of the 56 document-coded enterprises and 2 of the 12 interviewed ones. Substituting it for SCALE in a deep-tier specification would run the model on two enterprises, so we do not estimate it.

We can report the agreement between the two measures where both exist. On the twelve document-coded enterprises with both, a headcount of 50 or more agrees with SCALE for nine of the twelve ($\kappa = 0.50$). At a threshold of 20 the agreement is eight of twelve ($\kappa = 0.38$), and at 100 it is nine of twelve ($\kappa = 0.47$). The two measures are moderately related. Two of the disagreements are the largest employer among the twelve, at 400 staff, coded 0 on SCALE, which did not integrate, and an enterprise with 70 staff, also coded 0 on SCALE, which did. Among those twelve, no enterprise below 50 employees integrated and three of the six at or above it did. That is the direction Lisa Hawkes’s point predicts, and twelve enterprises with three events cannot establish it (Fisher $p = 0.18$).

We record the coverage failure because it is a finding about the programme’s instruments of the same kind as the two already in this paper, an integration field that is blank for half the portfolio, and a monitoring workbook whose attribution column duplicates its employment column. The application template does not ask applicants how many people they employ, although the programme selects on readiness to supply at volume.

CODIFIED records that a specification for the transacted input already existed at application. It is satisfied by a Unilever specification or approved grade, a recognised external standard defining the product, a finished Unilever-branded product, or an established commodity grade. It is coded 0 where a specification had to be created during the project, and it is present for 4 of 12 and 6 of 56. This is the transaction-side determinant of governance form in Gereffi et al. (2005), and it is close to uncodeable from applications. The coder returned it at low confidence for 43 of the 56 document-coded enterprises, because a proposal for a product the buyer has not yet specified rarely says so.

CERTIFIED records a named third-party certification, accreditation or audited standard held at entry (Fairtrade, Rainforest Alliance, organic, FSC, B-Corp, food-grade, ISO, GMP, HACCP, GlobalGAP or equivalent). An intention to certify is 0. It is present for 2 of 12 and 2 of 56, and low-confidence for 32 of 56.

GREEN_SECTOR records that the enterprise’s own core activity sat in recycling or materials recovery, refill or reuse, regenerative agriculture, renewables or waste-to-resource at entry, and would have done so had the project not been funded. It is present for 5 of 12 and 23 of 56. **CERTIFIED** and **GREEN_SECTOR** are carved out of **GREEN** and never enter a model alongside it.

UL_BUYIN records the Unilever local team’s buy-in to the project at its start, 1 where the tracker grades it High and 0 at anything less. The buyer records the field, so it cannot inherit **PRIORITY_FIT**’s evidence, which comes from the supplier’s documents. It is fixed at entry, which rules out the engagement field measured during delivery. It is graded for 88 of the 90 portfolio enterprises, so the specifications that use it run on the document-coded tier as well. Its agreement with **PRIORITY_FIT** on the deep tier is seven of ten. It is associated with **UL_LINK_START** ($\phi = 0.48$) without being identical to it, and the two are never entered together.

Earlier drafts carried this condition as **GOVERNANCE**, read as the governance mode at the Unilever interface in the sense of Gereffi et al. (2005), with High buy-in standing for a relational mode and anything less for market or modular. We now retain the field under a different interpretation, for two reasons. The first was visible at the time. A version the research team coded from the enterprises’ entry-stage documents reproduced **PRIORITY_FIT** exactly on the ten deep-tier enterprises where both were coded, and we set it aside in favour of the buyer-side field, which measures buy-in. The second is established in Appendix F.5, which codes that framework’s three determinants properly for the twelve interviewed enterprises and derives the form from them. Of the five enterprises the buy-in rule calls relational, the determinant coding returns market for one, modular for one, captive for one and hierarchy for two, so the two codings agree on none of them. The field measures what the buyer’s local organisation was willing to do at the start of a project, which on the evidence of Section 5.6 is worth having as a condition.

C.2 The interviewed cases and their codes

Table 13 gives the twelve interviewed cases with their entry-stage codes on the five main conditions and the outcome; the case selection is described in Section 4.2.1.

Table 13: QCA dataset for the interviewed cases (entry-stage codes, $N = 12$)

Firm	UL_LINK_START	DOWNSTREAM	PRIORITY_FIT	GREEN	TIE_PRIOR	INTEGRATED
Refill A	1	1	1	1	0	0
Platform A	0	1	0	0	1	1
Agri A	0	0	0	0	1	1
Platform B	1	0	0	0	1	0
Recycling B	0	0	1	1	0	0
Recycling A	1	0	1	1	1	1
Agri B	0	0	0	1	1	0
Platform C	1	1	0	0	0	0
Refill B	1	1	1	1	1	1
Agri C	1	0	0	0	0	1
Refill C	0	1	1	1	1 ^r	0
Agri D	0	0	1	1	0	0

Notes: conditions are coded from application-stage documents only, except TIE_PRIOR, which is coded from the founder interview and is 1 where the founder is internationally networked or of Global North origin. ^rone TIE_PRIOR cell is a researcher ruling: a founder whose Unilever career ended before he applied carries neither NVivo code, and is coded 1 so that a career inside the buyer counts as reach for him as it does for the founder with two decades inside it (Section 5.3). The same five conditions are coded by the same pipeline for the 56 document-coded enterprises, and the two tiers agree on every cell for the ten enterprises they share; the wider dataset is reported in Appendix C. Three outcome cells are researcher overrides against the tracker: Agri C is coded 1, having been a major supplier of an agricultural commodity to Unilever before it applied; Agri D 0, having no tracker row and attempts in progress at the time of interview; and Recycling B 0, likewise having no tracker row, its founder stating that the enterprise has no transactions with Unilever. Two DOWNSTREAM cells are researcher rulings against the document coding: Refill B, on the decision that the case is the parent organisation as a whole, and Platform A, where the entry-stage code was the lone dissenter against the document-coded portfolio tier, the tracker’s value-proposition classification and the enterprise’s own account of its route to Unilever, all of which place it downstream. One UL_LINK_START cell is a researcher ruling: Agri D is coded 0, because its only evidence for a link is the Pillar Lead’s box on the proposal form, a role within selection (Section 4.2.2). Nineteen condition cells in this tier are coded 0 at low confidence because the application documents are silent; Section 6 tests what rests on them.

C.3 Specifications and solutions, both tiers

Table 14 lists every specification and its parsimonious solution on the entry-stage coding. Because twelve cases cannot identify four conditions, we run every deep-tier model at three. Where a four-condition specification had to lose one, we dropped UL_LINK_START, the condition Section 5.3 tests on its own. Consistency is 1.00 wherever a solution exists. Terms resting on a single case are marked.

Across the two tiers, seven of the ten integrated enterprises are reached by some specification, and no single specification reaches more than three of them. On the deep tier no non-integrated enterprise is reached by any specification, which at a consistency of 1.00 it cannot be. The seven enterprises that no specification reaches are therefore Refill A, Platform B, Recycling B, Agri B, Platform C, Refill C and Agri D. On the document-coded tier, four of the ten integrated enterprises (Recycling F, Agri A, Platform I and Refill B) are reached by no specification.

C.4 Robustness to the twelfth case

Recycling B is the twelfth interviewed case, included throughout the twelve-case deep tier reported above. It has application documents but no tracker row, and its outcome is coded from its own account, as Agri D’ is. The founder states that “right now, we don’t have any transactions with Unilever,” the scale gap is unresolved and the one order in prospect is with another buyer, so the enterprise is coded not integrated (TIE_PRIOR = 0, INTEGRATED = 0). Its inclusion leaves the reported specifications of Table 14, the necessity results for integration and every interpreted finding unchanged. It moves denominators, coverage and trivialities slightly. It shifts the RESOURCE_GAP share among non-integrating cases from four of six to four of seven. In the one saturated interview-coded model the text does not interpret (M-I3), it drops the $\sim$ RESOURCE_GAP · DEMAND_ABSENT term, since Recycling B shares that configuration as a non-integrated case.

Table 14: Every specification and its parsimonious solution (entry-stage coding)

Model	Tier	Conditions and solution	Cons.	Cov.	Cases covered
M1	deep	DOWN, PRIO, GREEN: no solution	—	—	—
M1	doc.	UL, DOWN, PRIO, GREEN: ~UL*DOWN*~GREEN [†] + DOWN*PRIO*~GREEN [†] + UL*~DOWN*~PRIO*GREEN [†]	1.00	0.300	Platform A; Platform L; Recycling L
M-I1	deep	TIE, DOWN, GREEN: TIE*DOWN*~GREEN [†] + ~TIE*~DOWN*~GREEN [†]	1.00	0.400	Platform A; Agri C
M5	deep	DOWN, PRIO, SCALE: ~PRIO*SCALE [†]	1.00	0.200	Agri A
M5	doc.	UL, DOWN, PRIO, SCALE: no solution	—	—	—
M6	deep	UL, SCALE, PRIO: SCALE*~PRIO [†] + UL*~SCALE*PRIO [†]	1.00	0.400	Agri A; Refill B
M6	doc.	UL, SCALE, PRIO: no solution	—	—	—
M4	deep	UL, DOWN, PRIO: UL*~DOWN*PRIO [†] + ~UL*DOWN*~PRIO [†]	1.00	0.400	Recycling A; Platform A
M4	doc.	UL, DOWN, PRIO: no solution	—	—	—
M7	deep	UL, DOWN, CODIFIED: ~DOWN*CODIFIED + ~UL*DOWN*~CODIFIED [†]	1.00	0.600	Platform A, Recycling A, Agri C
M7	doc.	UL, DOWN, CODIFIED: UL*~DOWN*CODIFIED	1.00	0.200	Recycling A, Agri C
M3	deep ^a	GOV, DOWN, PRIO: ~GOV*PRIO [†] + ~DOWN*PRIO [†]	1.00	0.400	Refill B, Recycling A
M3	doc.	GOV, DOWN, PRIO: no solution	—	—	—
M2	doc.	GOV, DOWN, PRIO, GREEN: GOV*DOWN*~GREEN DOWN*PRIO*~GREEN	1.00	0.100	Platform A
M-H3	deep	DOWN, CERT, GSEC: no solution	—	—	—
M-H3	doc.	UL, DOWN, CERT, GSEC: no solution	—	—	—

Deep tier $N = 12$, 5 integrated; document-coded tier $N = 56$, 10 integrated. [†]term rests on one case. ^aUL_BUYIN is read from the tracker, which has no row for Agri D or Recycling B and no buy-in grade for one document-coded enterprise, so M3 runs on 10 deep-tier cases and M2 and M3 on 55 document-coded ones; M2 is retired on the deep tier because at three conditions it is M3. “|” separates alternative solutions of equal consistency. Cov. is solution-level coverage. M-I1 is deep-tier only (TIE_PRIOR is coded from interviews). Source: 28_specification_robustness_20260922.xlsx.

C.5 Necessity with a relevance check

Table 15 reports, for INTEGRATED, the necessity consistency of every condition and its negation, beside its coverage and its triviality, which is the share of all cases in which the condition is present. At a triviality of 0.80 or more, a consistency near 1.00 follows largely from the prevalence of the condition. No condition reaches the 0.90 threshold non-trivially on either tier. The portfolio-scale result of Section 5.2, UL_LINK necessary at consistency 1.000 on the tracker’s own coding, has a coverage of 0.212 and a triviality of 0.733 (66 of 90 enterprises have the condition). The condition is therefore necessary and common, and the main text interprets it on that basis.

C.6 Sensitivity to the definition of integration

Section 5.8 codes two alternative outcomes from the interviews for the deep tier. They are sme_reported, any route to Unilever the enterprise itself reports (7 of 12), and sme_direct, a direct commercial relationship only (3 of 12). Table 16 gives every specification under all three. Under the broadest definition solutions become easier to find, and under the strictest three of the eight specifications return none. Counting every

Table 15: Necessity for INTEGRATED, with coverage and triviality (entry-stage coding)

Condition	Deep ($N = 12$, 5 integrated)			Document-coded ($N = 56$, 10)		
	inclN	covN	triv.	inclN	covN	triv.
UL_LINK_START	0.600	0.500	0.500	0.500	0.227	0.393
~UL_LINK_START	0.400	0.333	0.500	0.500	0.147	0.607
DOWNSTREAM	0.400	0.400	0.417	0.300	0.300	0.179
~DOWNSTREAM	0.600	0.429	0.583	0.700	0.152	0.821
PRIORITY_FIT	0.400	0.333	0.500	0.500	0.147	0.607
~PRIORITY_FIT	0.600	0.500	0.500	0.500	0.227	0.393
GREEN	0.400	0.286	0.583	0.400	0.133	0.536
~GREEN	0.600	0.600	0.417	0.600	0.231	0.464
TIE_PRIOR	0.800	0.571	0.583	—	—	—
~TIE_PRIOR	0.200	0.200	0.417	—	—	—
SCALE	0.400	0.500	0.333	0.400	0.211	0.339
CODIFIED	0.400	0.500	0.333	0.200	0.333	0.107
~CODIFIED	0.600	0.375	0.667	0.800	0.160	0.893
CERTIFIED	0.200	0.500	0.167	0.100	0.500	0.036
~CERTIFIED	0.800	0.400	0.833	0.900	0.167	0.964
GREEN_SECTOR	0.200	0.200	0.417	0.300	0.130	0.411
~GREEN_SECTOR	0.800	0.571	0.583	0.700	0.212	0.589
UL_BUYIN ^a	0.400	0.400	0.500	0.700	0.241	0.527

inclN = share of integrated cases with the condition; covN = share of cases with the condition that are integrated; triv. = share of all cases with the condition. ^aUL_BUYIN on 10 deep and 55 document-coded cases. TIE_PRIOR is coded for interviewed enterprises only. Source: 28_specification_robustness_20260922.xlsx, sheet necessity_sweep.

term in Table 16, 24 of the 33 solution terms across the three definitions rest on a single case. For direct integration, ~GREEN_SECTOR is necessary at inclN = 1.000 with a triviality of 0.583. On this strict outcome ~GOVERNANCE (triviality 0.500) and ~SCALE (0.667) are necessary as well.

Table 16: Deep-tier solutions under three definitions of integration ($N = 12$)

Model	Tracker (5 integrated)		Reported (7)		Direct (3)	
	Solution	Cov.	Solution	Cov.	Solution	Cov.
M1	none	—	~PRIO*GREEN	0.143	none	—
M5	~PRIO*SCALE	0.200	DOWN*SCALE + ~PRIO*SCALE	0.286	none	—
M6	SCALE*~PRIO + UL*~SCALE*PRIO	0.400	~UL*~PRIO + UL*PRIO	0.857	UL*~SCALE*PRIO	0.333
M4	UL*~DOWN*PRIO + ~UL*DOWN*~PRIO	0.400	UL*PRIO + ~UL*~PRIO	0.857	~UL*DOWN*~PRIO	0.333
M7	~DOWN*CODIFIED + ~UL*DOWN*~CODIFIED	0.600	UL*CODIFIED + ~UL*DOWN*~CODIFIED	0.571	~UL*DOWN*~CODIFIED	0.333
M3	~GOV*PRIO + ~DOWN*PRIO	0.400	~GOV*PRIO + GOV*~DOWN	0.571	~GOV*PRIO	0.333
M-I1	TIE*DOWN*~GREEN + ~TIE*~DOWN*~GREEN	0.400	~TIE*~DOWN*~GREEN + TIE*DOWN*~GREEN + ~TIE*DOWN*GREEN	0.714	TIE*DOWN*~GREEN + ~TIE*~DOWN*~GREEN	0.667
M-H3	none	—	CERTIFIED	0.286	none	—

Consistency is 1.00 for every solution shown. The M-I1 row carries the researcher ruling on one TIE_PRIOR cell (Table 13). Source: 29_outcome_definition_comparison_20260922.xlsx.

C.7 Reliability of the coding

Restricting the evidence to the application changes the codes, and Appendix B reports by how much, condition by condition (Table 12). That comparison sets, for the 56 document-coded enterprises, the codes returned from the whole document folder against those returned from application-stage documents only. Twenty cells move from 1 to 0 when later documents are withheld and eleven move the other way, and the movement is largest on UL_LINK_START.

The two document-coded tiers agree completely where they overlap: the ten enterprises present in both

carry identical values on every condition and on the outcome.

The coder also flags a cell as low-confidence where the evidence is thin or silent. On the deep tier, 19 of the 96 cells across eight conditions are low-confidence, all coded 0, concentrated in CODIFIED (6), GREEN (4), CERTIFIED (4) and GREEN_SECTOR (4). On the document-coded tier 138 of 448 cells are low-confidence, 137 of them coded 0, concentrated in CODIFIED (43) and CERTIFIED (32). Two integrated document-coded enterprises, Platform I and Recycling F, are low-confidence on almost every condition. Under application-stage evidence most enterprises have one or two documents, so a condition the documents do not mention is coded as absent. The results on CODIFIED, CERTIFIED and GREEN_SECTOR in particular are nulls on conditions that the application documents mostly do not record.

Three condition cells are set by ruling against the document coding. Two of them are DOWNSTREAM cells. Refill B is ruled downstream because the case is the parent organisation, which on-sells Unilever product through its refill technology, as the tracker also records. Platform A is ruled downstream because its entry-stage code of 0 disagreed with the document-coded tier, the tracker’s value-proposition classification and the enterprise’s own account. The third is the UL_LINK_START cell of Agri D, coded 0 because its only evidence for a link is the Pillar Lead’s box on the proposal form. Two outcomes are set by override, Agri C to 1 and Agri D to 0 (Section 4.2.1).

C.8 Interview-coded conditions

Four further conditions are coded from the founder transcripts in NVivo Final Clean Coding v3 and enter deep-tier models only. They are GREEN_SELECTED (code 7.1, green credentials given as the reason for selection), CHAMPION (2.3, a named Unilever champion carrying the relationship), RESOURCE_GAP (5.1) and DEMAND_ABSENT (5.5). All four share the coding convention of TIE_PRIOR, which is 1 where the code is applied to the founder’s transcript, 0 where the founder is interviewed and the code is absent, and blank where there is no interview. A 0 is therefore silence and not established absence, and every one of these is retrospective—coded in 2026 from accounts of decisions taken years earlier by founders who know how the relationship ended.

On necessity, on twelve cases with five integrated, no interview-coded condition approaches necessity for integration. TIE_PRIOR reaches $\text{inclN} = 0.800$, DEMAND_ABSENT and GREEN_SELECTED reach 0.600, and CHAMPION and RESOURCE_GAP reach 0.400. Read against non-integration the ordering inverts, and the one worth recording is RESOURCE_GAP, which is present for four of the seven non-integrating enterprises ($\text{inclN} = 0.571$).

One TIE_PRIOR cell is a researcher ruling, and its boundary matters to the result. The NVivo coding records an international network or Global North origin. On that code a founder with two decades inside the buyer is 1, while a founder whose Unilever career ended some years before he applied, and who found the programme through an open call, is 0. We apply the same reading to both and code the second 1 (Table 13). On the code as first applied, prior reach combined with downstream position was a two-case path at a coverage of 0.600. Applying the reading consistently removes that path, and the specification reported in Section 5.3 falls to two single-case terms, so the paper relies on the necessity result, four of five integrators against seven of twelve cases, which does not move. The leading term also depends on two DOWNSTREAM cells that are researcher rulings (Table 13).

Two models use these conditions. M-I2 substitutes GREEN_SELECTED for GREEN and enters it with DOWNSTREAM and TIE_PRIOR, and M-I3 enters the three buyer-side conditions together. Both return solutions at consistency 1.00. M-I2 gives $\text{GREEN_SELECTED} \cdot \sim \text{DOWNSTREAM} \cdot \text{TIE_PRIOR}$ (coverage 0.400, two cases, Agri A and Recycling A), $\sim \text{GREEN_SELECTED} \cdot \text{DOWNSTREAM} \cdot \text{TIE_PRIOR}$ (0.200, Platform A alone) and a single-case term on Agri C, for a solution coverage of 0.800. M-I3 gives $\text{CHAMPION} \cdot \text{DEMAND_ABSENT}$, covering two cases. Without Recycling B a second term, $\sim \text{RESOURCE_GAP} \cdot \text{DEMAND_ABSENT}$, also appears, and it drops once Recycling B enters the truth table, since Recycling B shares that configuration as a non-integrated case.

We report them here, outside the main text, for a reason of construction. Both truth tables are saturated. All eight configurations are occupied by twelve cases, so every row holds one or two enterprises and the minimisation has no logical remainders to work with. A saturated table has no counterfactual problem, but its solution is a re-description of individual cases, since no configuration is left for it to generalise over. The M-I3 result also has the two signs the reverse of those hypothesised, with DEMAND_ABSENT on the

path to integration and CHAMPION on the path to non-integration. We take that as evidence that the codes record how founders narrate an outcome they already know, and we do not interpret the terms. We retain two results. The first is the necessity ordering above. The second is the result reported in the main text, that on the interview coding green credentials combined with a prior international tie are sufficient for integration in two of the five integrating enterprises, which is the one result on which the interview coding disagrees with H3.

C.9 Bivariate rates on the document-coded sample

Section 5.5 reports three of the bivariate rates on the document-coded sample; Table 17 gives all five.

Table 17: Document-coded bivariate integration rates ($N = 56$, 10 integrated).

Condition	Absent	Present	n present
Relationship at entry (documents)	0.147	0.227	22
Downstream	0.152	0.300	10
Priority fit	0.227	0.147	34
Green	0.231	0.133	30
Scale	0.162	0.211	19

C.10 Sensitivity of the configurational null to the low-confidence cells

The coder flagged nineteen of the ninety-six condition cells on the twelve interviewed cases as low-confidence, and all nineteen are coded 0, concentrated in CODIFIED (6), GREEN (4), CERTIFIED (4), GREEN_SECTOR (4) and UL_LINK_START (1). We ran three checks on an in-memory copy of the canonical dataset, which was not modified. The first is a replication, in which re-running the ten canonical deep-tier specifications reproduces every reported parsimonious solution, ten of ten. The second is an exclusion check, in which treating the nineteen cells as unknown removes cases from the specifications that use them and costs one specification its solution. The third is a bounded simulation, in which we redraw only those nineteen cells independently, ten thousand times at each of three priors, and re-run only the pre-specified specifications. Table 18 reports the distribution of the best coverage achieved. The distribution barely moves as the prior shifts and never establishes a configuration covering most integrators. Section 6 gives the reading.

Table 18: Deep-tier solution coverage when the nineteen low-confidence zero cells are redrawn

Prior	Draws	No sufficient	Median best	5th–95th	Coverage	
		solution	coverage	percentile	≥ 0.5	≥ 0.8
$P(1) = 0.25$	10,000	0%	0.6	0.4–0.8	87%	29%
$P(1) = 0.50$	10,000	0%	0.6	0.4–0.8	80%	36%
$P(1) = 0.75$	10,000	0%	0.6	0.4–0.8	78%	24%

Only the nineteen cells the coder flagged low-confidence *and* coded 0 are redrawn. The outcome, every coded 1, every confident zero, the two DOWNSTREAM researcher rulings and every interview-coded condition are held fixed. Each draw re-runs only the pre-specified canonical specifications, with the conditions the canonical run used; no condition search. Coverage is the best achieved by any affected specification in that draw. Seed 20260901. For comparison the canonical maximum is 0.600 (M-I1), and every canonical term rests on one, two or three enterprises. Source: `Analysis/additional_suggestions_pipeline`, Analysis 3.

D The regression

This appendix reports the regression strand in full. It covers the estimators and variable definitions, the bivariate and multivariate models, the mechanics of the preferred specification, the referral-origin cross-

tabulation, the direction cross-tabulation, and the post-project correlation used as a coding check. The three-basis sensitivity table is in Section 6.

D.1 Estimators, specifications and variables

The models reported here are ordered logistic (proportional-odds) regressions (McCullagh, 1980). We use them because the outcome is a four-step ladder of increasing integration, not integrated, trying, integrated-then-lapsed, and currently integrated, and an ordered logit models the odds of a firm sitting at or above each rung of that ladder as a function of its characteristics. The “proportional-odds” label names the model’s key assumption, which is that a given predictor shifts those odds by the same factor at every rung. Because that assumption can fail, we also estimate a simpler binary logistic model on an “ever-integrated” outcome, which collapses the ladder to a single yes/no (integrated at any point versus never). The binary model discards the ranking but rests on fewer assumptions, and it is the form the preferred specification takes. Both are fitted by maximum likelihood, the standard estimation method for models of this kind (Venables and Ripley, 2002). The predictors are the programme covariates (TRANSFORM phase; broad pillar; referral route), the ordinal Unilever-relationship scores, and supply-chain direction. The relationship scores are buy-in at start, link strength at start, link strength during the project, and local engagement, each on a None–Low–Medium–High scale. Full definitions are in Table 19.

We run each specification in two forms. In the “bivariate” models we enter one predictor at a time and estimate whether that characteristic, on its own, is associated with integration, which is a regression analogue of a simple cross-tabulation. In the “multivariate” models we enter the predictors together and estimate whether each remains associated with integration once the others are held constant. Joint estimation controls for confounding between predictors, for example an apparent pillar effect that reflects supply-chain direction, with which pillar is nearly collinear. Two features of the data warrant caution. The analytical sample is small, and some categories of the referral-route variable contain very few firms. Where a category almost perfectly predicts the outcome, the model suffers quasi-complete separation, which drives the corresponding coefficient towards infinity so that no stable estimate exists. We handle this in three ways, by collapsing sparse categories, by dropping referral route from the preferred specifications, and by re-estimating under Firth’s penalised likelihood, a standard correction that keeps coefficients finite under separation. Given the sample size, we report all regression results as exploratory and suggestive of pattern.

In its ordered form, for enterprise i the ordered model estimates

$$\log \frac{P(Y_i \leq k)}{P(Y_i > k)} = \alpha_k - \mathbf{x}_i' \boldsymbol{\beta}, \quad k = 0, 1, 2$$

where α_k are outcome-specific cut-points, $\mathbf{x}_i$ is a vector of enterprise characteristics, and $\boldsymbol{\beta}$ is common across thresholds. A positive coefficient indicates a higher probability of reaching a greater integration level. We estimate the models by maximum likelihood using `polr` in R (Venables and Ripley, 2002). Four specifications are reported, in ordered and binary form: Model A (T1/T2 and the four Unilever-relationship scores); Model B (A plus pillar and referral route); Model C (A plus supply-chain direction); and Model D (B plus direction). All are estimated on the same $N = 81$ complete cases.

As a robustness check, we re-estimate each specification as a binary logistic regression, collapsing the outcome to ever-integrated (categories 2–3) versus not (categories 0–1). This relaxes the proportional odds assumption and provides a simpler check on direction and significance of effects.

Separately, we test the bivariate association between post-project Unilever link strength and current integration status using Spearman’s rank correlation, appropriate for two ordinal variables.

Table 19 summarises the key variables used in the regression models. The ordinal Unilever-engagement variables (buy-in, link strength at start, link strength during project, and local engagement) are recorded in the database as categorical labels (None, Low, Med, High) and converted to integer scores (0–3) for regression. We treat the resulting scores as continuous regressors, a common simplification given the small sample. We construct broad pillar categories by grouping the programme’s twelve detailed pillars into four theoretically motivated clusters: Circular/Green (recycling, refill, regenerative agriculture, energy), Livelihoods (farmer and retailer livelihoods, future of work), Health & WASH (health, digital health, water, sanitation), and Research/Other. We collapse referral route from the eleven detailed channels the tracker records to three

groups (Global pillar lead, Local/Regional, Other) to mitigate perfect separation in thin cells. Appendix D.7 gives the channels in full.

Table 19: Variable Descriptions

Variable	Scale	Description
<i>Outcome</i>		
Integration status	Ordered (0–3)	Not integrated (0); Trying (1); Integrated, discontinued (2); Integrated (3)
Ever integrated	Binary	1 if categories 2 or 3; 0 if categories 0 or 1
<i>Programme covariates</i>		
T1/T2	Binary	0 = TRANSFORM 1 (2015–2019); 1 = TRANSFORM 2 (2020–present)
Pillar (broad)	Categorical	Circular/Green (ref.); Livelihoods; Health & WASH; Research/Other
Referral route	Categorical	Global pillar lead (ref.); Local/Regional; Other
<i>Unilever link</i>		
UL local buy-in at start	Score (0–3)	The tracker’s four-point rating of the Unilever local team’s buy-in at project commencement, which is a rating of the country business’s support for the project and not of the commercial link (None=0, Low=1, Med=2, High=3)
Strength of UL link at start	Score (0–3)	The tracker’s four-point rating of the Unilever link at project commencement, the team’s assessment of how well the project matched a use inside Unilever (None=0, Low=1, Med=2, High=3)
Prior relationship at entry	Categorical	The programme lead’s 2026 record of whether the enterprise was known to someone in the business, or already working with it, when its first project was approved (no; known, not working together; known, unclear; already working; unknown)
Strength of UL link during project	Score (0–3)	The same rating over project delivery
UL local engagement	Score (0–3)	Unilever local team engagement during project
Strength of UL link after project	Score (0–3)	The same rating after project close (used in correlation analysis only)
<i>Supply chain</i>		
Upstream/Downstream	Binary	1 = Upstream supplier; 0 = Downstream distributor
Referral origin	Categorical	Local-referral-to-global; Global-led; Regional/Other
Time in programme	Months	Start date to end date, or to Mar-2026 cut if ongoing

The three link variables are the tracker’s Strength of UL link ratings, and the naming matters because the tracker holds two columns per window whose titles differ by one word. UL link at start of project is free text describing what the link was—“Refill of FMCG products”, “Collect and process target for flexibles”—while Strength of UL link at start of project is the four-point rating we use. Only the rating enters any model here. The free-text description and the tracker’s value-proposition category record what the match consisted of; Section 5.2 uses the category (Table 22).

D.2 From four link scores to Model P

Entered one at a time, each of the four Unilever link scores is significant (bivariate ordered logits on the $N = 81$ frame; the full tables are in the replication package). Entered together, none of them is. That pattern is what four collinear measures of one construct produce. The tracker records the link at four moments, and once any one of them is in the model the others add nothing. We conclude from the block that the buyer’s assessed link predicts integration, and we cannot conclude which moment of it does.

Collinearity produces the fall in the starting-link coefficient under joint estimation, and a null produced this way is easily misread as evidence of no effect, so we set out the mechanism. Link strength at start and link strength during the project correlate at $\rho = 0.90$. With the during block in the equation, roughly

four-fifths of the variance in the starting measure is accounted for by the other regressors, so its coefficient is identified off the remaining fifth. The point estimate shrinks and the standard error inflates at the same time, and the z -statistic falls from 2.3 to 0.3. The two measures are not separately identified, because the model cannot separate a match at entry from the same match three months later, which reflects when the tracker records its ratings.

D.3 Model P in full

Section 5.2 reports the preferred specification in words. This subsection gives it in full, with the reasoning that selected it, the document-coded re-estimate and the relationship variant.

This subsection estimates the entry result of Section 5.2 by regression and identifies the specification the paper should prefer. Appendix D.2 shows that the four Unilever link scores measure one construct, and that implies a remedy. Two of the four are measured over the window in which integration occurs and cannot be predictors on any reading, so the equation should retain only what is fixed before the project begins. Model P is that equation, with cohort, local buy-in at start and link strength at start as its only predictors. On the enterprises complete on all five predictors ($N = 81$, 12 events) it returns a starting-link coefficient of +1.50 ($p = 0.023$), meaning that each step up the four-point starting-link scale multiplies the odds of integration by about four and a half. The Firth-penalised equivalent is +1.29 with a 95 per cent interval of [+0.27, +2.65] ($p = 0.011$). Adding the during-project block moves the starting-link estimate to +0.30 ($p = 0.787$), because the during-project scores absorb it. A likelihood-ratio test of what the block adds returns $p = 0.364$, and the Akaike criterion narrowly favours the shorter model (66.85 against 68.83). Dropping the block therefore costs nothing measurable, and it leaves a starting-link coefficient that the fuller model does not show. On its own three predictors Model P has 86 complete cases and 13 events. We report it on the smaller set so that the likelihood-ratio test compares the same rows.

A match recorded at entry is not exogenous just because it precedes the project. Enterprises whose projects the buyer's side could already use differ in ways we do not observe, so the coefficient identifies selection, which is what H1 asserts, and it cannot be read as an estimate of what a match would do for an enterprise that lacks one. The starting rating is also the variable on which the sources differ most, and the difference is one of construct. The document condition is coded independently from the application documents, which record whether a relationship with Unilever existed. It agrees with the tracker's match rating on 48 per cent of the 56 enterprises for which we hold both, a Cohen's κ of 0.06, and the disagreement is one-sided, with the tracker recording a match where the documents show no relationship in 24 cases and the reverse in 5. Re-estimating Model P with the document-coded relationship in place of the tracker's match returns +0.52 ($p = 0.440$), the same sign at about a third of the magnitude and not distinguishable from zero. An earlier version of this appendix read the documents as a noisy measurement of the tracker's construct and computed how much of the tracker's gap measurement error alone would leave on the documents. On that reading the eight-point gap on the documents was four times what noise would leave. Since the two records measure different things, the calculation no longer applies and is not reported. Because the documents were coded independently of the team that scored the outcome, the association on the documents cannot be an artefact of that team's scoring. The buyer's own record of a prior relationship, added in 2026, asks the documents' question, and Table 20 gives the agreement between the two, 32 of 46 enterprises (69.6 per cent). Adding that record to Model P beside the match is the relationship variant of Section 5.3, and Table 21 reports it. On the 71 enterprises with a known value, the match coefficient falls from +1.17 to +0.64 when the relationship enters, and the relationship coefficient is +1.00 ($p = 0.197$). The pattern is the same when the unknown values are coded as no relationship, and on the programme's coding without the override. The one enterprise dropped from the known-value rows is an integrator with a relationship whose buy-in cell holds a mis-entered value.

The programme team's category for what each project offered Unilever, which Section 5.2 summarises in two groups, is given in full in Table 22.

D.4 Cohort and tenure

Three further checks concern the cohort, the tenure measure and the outcome coding. On cohort, read from the project the regression keeps for each enterprise, 4 of the 32 first-generation enterprises ever integrated

Table 20: The Enterprises’ Documents against the Buyer’s Record of a Prior Relationship ($N = 46$)

Document-coded link at application	Buyer’s record (prior relationship)		Total
	No	Yes	
No link	23	5	28
Link	9	9	18
Total	32	14	46

Notes: the 46 document-coded enterprises (Section 4.2.2) for which the buyer’s record holds a known value. The two records agree on 32 of 46 (69.6 per cent). Where the documents record a link the buyer’s record does not, the evidence is mostly a previous TRANSFORM project or a partnership described in the proposal; where the buyer’s record has a relationship the documents do not show, the coder’s evidence was thin or the founder’s own history was the link.

Table 21: Model P with the Match and the Prior Relationship at Entry (Firth logit)

	N / events	Match at start $\hat{\beta}$ [95% CI] (p)	Prior relationship $\hat{\beta}$ [95% CI] (p)
<i>Known values on the buyer’s record</i>			
Model P	71 / 12	+1.17 [+0.11, +2.55] (0.029)	—
Model P + prior relationship	71 / 12	+0.64 [−0.60, +2.17] (0.324)	+1.00 [−0.51, +2.69] (0.197)
<i>Unknown and blank coded as no relationship</i>			
Model P	86 / 13	+1.31 [+0.30, +2.67] (0.009)	—
Model P + prior relationship	86 / 13	+0.89 [−0.22, +2.35] (0.121)	+0.85 [−0.53, +2.32] (0.229)
<i>The programme’s coding, without the researcher override</i>			
Model P	71 / 12	+1.17 [+0.11, +2.55] (0.029)	—
Model P + prior relationship	71 / 12	+0.90 [−0.38, +2.52] (0.179)	+0.48 [−1.05, +2.11] (0.543)

Notes: Firth-penalised logit on ever-integration with cohort, declared buy-in at start and the match at start (the tracker’s rating of the Unilever link at the start of the project the regression keeps), with and without the prior relationship at first entry (Table 5). Each pair of rows is estimated on the same enterprises, those complete on the four predictors. Every enterprise with a prior relationship has a match rated High at the start of at least one of its projects (24 of the 25 on the project the regression keeps), so the two measures are nested and the table cannot separate them as two independent gates.

Table 22: Integration by the Kind of Match the Programme Team Recorded ($N = 90$)

What the project offered Unilever	Not ever	Ever integrated	N	% integrated
A route to market	16	6	22	27.3
A sustainable supplier	11	4	15	26.7
Progress on a plastics goal	17	3	20	15.0
A refill model	5	0	5	0.0
A brand-purpose story	6	1	7	14.3
Other (behaviour change, product tests, staff, reputation)	12	0	12	0.0
None recorded	9	0	9	0.0

Notes: the tracker’s “value proposition” field at the enterprise’s first project, the programme team’s retrospective category for what the project offered Unilever, grouped by the authors. Projects offering a commercial function (a route to market or a supplier) integrated at 10 of 37 (27.0 per cent); projects offering progress on a sustainability objective (plastics, refill) at 3 of 25 (12.0 per cent). Cells are small and no test is reported.

(12.5 per cent) and 1 remains currently integrated, and 10 of the 58 second-generation enterprises ever integrated (17.2 per cent) and 6 remain. Fisher exact returns $p = 0.76$ and a penalised logit $\hat{\beta} = +0.32$ ($p = 0.60$). Section 3.1 sets out why the enterprise-level construction reverses the comparison. Thirty-five enterprises carry no entry in the tracker’s integration field on any of their projects, and the programme lead has confirmed that a blank records an enterprise that did not integrate. Reading blanks as zeros is therefore the correct reading of the tracker. We mention it because the alternative changes the result. Restricted to the enterprises Unilever had explicitly assessed, the cohort rates invert to 57.1 and 20.8 per cent. They invert because most of the first-generation enterprises carry a blank against a minority of the second, so an analysis of the assessed enterprises compares the few older enterprises someone had a reason to assess with most of the younger ones. On the full sample, and read from the project the regression keeps, cohort shows no association in either direction. Programme generation is null in every specification, and its sign is positive in the ordered ones.

Table 23: Model P: the starting link with and without the during-project block

	N / events	Starting link $\hat{\beta}$ (p)	Firth $\hat{\beta}$ [95% CI]	Model comparison
Model P, pre-project only	81 / 12	+1.50 (0.023)	+1.29 [+0.27, +2.65]	AIC 66.85
Model P + during-project block	81 / 12	+0.30 (0.787)	—	AIC 68.83; LRT $p = 0.364$
Model P on its own complete cases	86 / 13	+1.52 (0.020)	+1.31 [+0.30, +2.67]	—
Model P with the document-coded link	56	+0.52 (0.440)	—	—

Notes: binary logit on ever-integration with cohort, declared buy-in at start and link strength at start; the during-project block adds link strength during the project and local-team engagement. The first two rows are estimated on the same rows so that the likelihood-ratio test compares like with like. The last row replaces the tracker’s starting link with the relationship condition coded from the enterprises’ own application documents.

On tenure, we find no association between time in the programme, measured from start date to end date or to the March 2026 cut for ongoing projects, and integration. The Spearman correlation is $\rho = 0.11$ ($p = 0.39$) on the 59 enterprises with a computable duration, and the penalised logit coefficient is null ($\hat{\beta} = 0.05$, $p = 0.44$). The measure is imperfect, since ongoing projects are assigned time-so-far, which conflates duration with an unresolved outcome, so we read it as an imprecise estimate that is consistent with a null. It also tests the tenure component of H1 only loosely. The tracker records project start and end dates without a relationship-start date, so tenure at entry is not constructible from it.

D.5 Unilever conduct during the project

Section 5.5 discusses the tracker’s two during-project ratings; Table 24 gives them in full. Both are assessed after the event by the programme team that coded the outcome and neither measures programme effort.

Table 24: Integration by Unilever conduct during the project (portfolio sample)

Level	Link strength during		Local-team engagement	
	Integrated / n	%	Integrated / n	%
None	0 / 13	0.0	1 / 17	5.9
Low	0 / 10	0.0	2 / 14	14.3
Medium	2 / 23	8.7	2 / 15	13.3
High	12 / 44	27.3	8 / 39	20.5

Link strength during the project is graded for all 90 enterprises; local-team engagement is graded for 85, the other five being blank.

D.6 Referral origin, cohort and devolved selection in full

Section 5.4 reports the local-referral result and summarises two checks on it; this subsection gives the checks in full, on the settled 90-enterprise frame.

The tracker records, separately from the referral route, the programme team’s rating of the Unilever local team’s buy-in at project start. Entering local referral and that rating in the same penalised logit, local referral retains its sign but falls to the edge of conventional significance (1.39, $p = 0.056$), while the local buy-in rating is null ($p = 0.62$). The referral route, which records what a market team had already done, predicts integration. The buy-in rating, which records how supportive the team was rated as being, does not, though on the settled frame the joint test no longer clears 5 per cent. We read this as a demand-side result, and the reading is an inference. A referral that began in a market is evidence that somebody in the country business already knew what the enterprise could be used for, which a rating of support at project start does not capture. It is the same thing the interviews describe when programme staff make internal sponsorship the determinant of whether an enterprise is integrated (Section 5.6). That reading also limits what the result can be used for. If the route marks demand that already existed locally, then routing more

enterprises through market teams would not create the demand that makes the route work. The referral route is the one feature under the programme’s own control that appeared to affect integration. On this reading it supports the selection account of H1, since the route records which enterprises a market team already wanted.

The result requires a within-cohort check, because referral decisions were progressively devolved to country and regional panels across the two programme generations, and the check qualifies it. We count each enterprise in the generation in which it entered, which holds fixed the selection regime it was referred under. On that count the narrow local-referral route carries the result among first-generation entrants, 6 of 9 against 2 of 31 ($\hat{\beta} = 3.09$, $p < 0.001$), and none of the five second-generation entrants it brought in has integrated, against 6 of 45 by other routes ($\hat{\beta} = -0.59$, $p = 0.68$). The three local referrals that integrate in the second generation when cohort is read from the project the regression keeps all entered in the first. That is why on that reading the route appears to keep its sign in both generations (3 of 8 against 7 of 50 in the second). The whole within-cohort result rests on 14 enterprises. The broader question, whether devolving selection to local teams improves integration, gets a different answer, and we separate it from the referral result. Pooling the two generations, enterprises selected by a local or regional panel integrate at 20.5 per cent against 9.8 per cent for enterprises selected globally, which looks like a case for devolution. Within each generation, counted by entry in the same way, the gap reverses in the second. In the first generation the locally selected enterprises integrate far more often. In the second, where local selection became the norm and covers 30 of 50 enterprises, they integrate less often, with a penalised logit on that generation alone returning $\hat{\beta} = -1.24$ (95 per cent interval $[-3.15, +0.52]$, $p = 0.16$). The pooled figure is a composition effect. Local selection is concentrated in the younger generation, which has had less time to convert, so the pooled difference reflects cohort. We therefore support the localisation corollary of H1 for the narrow referral route, with the qualification that the support is weak and rests on thin cells and on first-generation entrants. We find no detectable effect of devolved selection as such.

D.7 Every referral channel, and the donor-selected marks

Table 6 collapses the tracker’s referral field into three categories. Lisa Hawkes asked what the underlying channels do, and in particular how the routes each funding partner controlled fared, which the collapse hides. Table 25 gives every channel as the tracker records it, on the same 90 enterprises.

Table 25: Integration by referral channel, as recorded (ever integrated, $N = 90$)

Channel	n	Integrated	%	Marked FCDO	Marked EY
Global — pillar lead	36	3	8.3	7	0
Local referral to global	14	6	42.9	0	0
Regional — FCDO	12	1	8.3	1	0
Regional — Unilever	10	1	10.0	0	0
Social Alpha Plastics Call	6	1	16.7	0	0
Global — regional lead	3	0	0.0	0	0
Regional — EY	3	0	0.0	0	1
Global — procurement referral	2	1	50.0	0	0
Unknown	2	1	50.0	0	0
EY referral to global	1	0	0.0	0	0
Open call	1	0	0.0	1	0
All	90	14	15.6	9	1

Notes: the referral route is a project-level field and an enterprise with several projects can carry more than one; the table reads the modal route, as Table 6 does. Three of the 90 enterprises carry a second route, so the collapse costs one route each for three enterprises and none for the other 87. The last two columns are the tracker’s Notes field, which marks 9 of the 90 “Selected ‘for’ FCDO” and one “Selected ‘for’ EY”; the single “Selected ‘for’ Gates” mark in the sheet falls on a row that is not an enterprise of this frame.

The first observation from the table is that the local route, at 6 of 14, is the only channel above 20 per cent that holds more than two enterprises. The two channels that beat it on rate hold two enterprises each. The second is that the regional routes integrate at about the global pillar lead’s rate. The three routes named

for a regional office, the FCDO, Unilever and EY, return 1 of 12, 1 of 10 and 0 of 3. What distinguishes the local route is that the referral began with someone in a market who wanted the enterprise.

The third concerns the funding partners' own routes, and it is the part of this that has to be stated carefully, because the cells are tiny. Every enterprise on an EY record is a non-integrator. Three were referred by a regional EY team, one was referred by EY to the global panel, and one is marked in the Notes as selected for EY, so the EY routes return 0 of 4 and the mark 0 of 1. The FCDO appears on two records that do not coincide, 12 enterprises whose referral route names the FCDO (13 if a second, non-modal route is counted) and 9 marked in the Notes as selected for it, with an overlap of one. In total, 20 enterprises have an FCDO record of some kind, and one of them integrated, against 13 of the other 70 ($p = 0.18$). These are the same 20 enterprises, and the same single integrator, as the donor-selected robustness check of Section 6, which reaches them from the two tracker fields by a different route. With one event in twenty, none of this is an effect, and we do not report it as one. We take it as corroboration, from the tracker's own selection fields, of Lisa Hawkes's account that some enterprises entered the programme for reasons unrelated to any prospect of integration. The routes a donor controlled produced almost no integrations, and the routes a market team controlled produced nearly half of them.

E The interview evidence

E.1 The interview coding

One researcher coded the nineteen interview transcripts in NVivo, twelve with founders or senior managers of portfolio enterprises and seven with Unilever and programme staff, and a second-coder check is still outstanding. The codebook has fifty codes in thirteen groups, built from the literature of Section 2 and the funders' thematic priorities and extended inductively during coding. The corpus has 415 coded references, 268 in the enterprise transcripts and 147 in the staff transcripts. Table 26 reports every code in three ways, as the number of coded references in the twelve founder transcripts, the number in the seven staff transcripts, and the number of distinct founders whose transcript has the code. References count how often something was said and founders count how many people said it, so the third column shows how widely a theme is shared. A code can reach twenty-seven references and still come from six enterprises. Table 27 reports, for the twelve deep-tier enterprises, the presence in each founder's transcript of the codes that enter the analysis, beside the document-coded conditions and the outcome.

The groups that describe the mechanics of the buyer link (group 2) and the internal organisation of the buyer (group 5) are the largest in the corpus. The codes for a champion inside the buyer, for the loss of that champion, for the buyer's resource-allocation gap and for its restructuring together account for 61 references, and that is the evidence behind the mechanism of Section 5.6. The codes that record integration mode separate mediated supply (4.2) from commercial-platform integration (4.3). Direct supply to Unilever (4.1) is applied to two staff passages and to no founder transcript. The coding separates green credentials that pre-dated the programme (7.1) from greening that emerged through it (7.6), which is the distinction H3 rests on, and three enterprises have both. Two codes designed to record where a founder's account diverges from the tracker's classification of direction (8.3) and of the link at entry (9.3) were never applied.

Interview codes enter the analysis in three ways. First, TIE_PRIOR (Section 4.2.2) is coded 1 where 11.3 or 10.2 is applied to the founder's transcript, or by the one researcher ruling recorded in Table 13, which codes Refill C 1 on the founder's prior employment inside the buyer. The columns below are the NVivo export and show the codes as exported, so Refill C is 0 in the 11.3 column. Second, the alternative outcome definitions of Section 5.8 are hand-coded from the same transcripts. Third, Table 27 sets the mechanism codes against the configurational results. Of the seven deep-tier enterprises reached by no specification, four have the resource-allocation-gap code (5.1) and three the internal-champion code (2.3). Throughout, a 1 records that the code was applied to the founder's own transcript and a 0 that it was not, under the convention of Appendix C. Staff transcripts are not used for the per-enterprise matrix.

Table 26: Interview coding: references per code and the number of founders raising it, twelve enterprise and seven programme-staff transcripts (NVivo, Final Clean Coding v3)

Code	Founder refs	Staff refs	Founders
<i>1. Barriers (SME-side)</i>			
1.1 Resource and tech gaps	3	0	3
1.2 Compliance and certification costs	0	1	0
1.3 Power imbalance and price pressure	6	0	4
1.4 Operational / readiness gap	7	10	4
1.5 Supplier registration burden	5	2	2
1.6 SME internal change during project	2	0	1
<i>2. Mechanisms of the buyer link</i>			
2.1 Repeat business / reinvestment	1	0	1
2.2 Knowledge / capability transfer	20	11	8
2.3 Internal champion at Unilever	11	2	5
2.4 Project-lead continuity	1	3	1
2.5 UL project-lead active support drove integration	7	10	3
2.6 Brand signalling	9	4	7
2.7 Project-lead discontinuity / champion loss	3	2	3
<i>3. Enabling conditions</i>			
3.1 Scale pathway	6	2	3
3.2 Matching contribution	2	1	2
3.3 Policy Environment	2	4	1
<i>4. Integration modes</i>			
4.1 Direct supply to Unilever	0	2	0
4.2 Mediated supply via intermediary	5	1	5
4.3 Commercial platform integration	8	2	5
4.4 Wider market access beyond Unilever	5	3	4
<i>5. Inside the buyer</i>			
5.1 Resource-allocation gap	17	10	6
5.2 Internal restructuring / staff churn	11	5	4
5.3 Sustainability pillar changes	1	7	1
5.4 Procurement demand present / mature	3	9	3
5.5 Procurement demand nascent / absent	7	6	5
<i>6. Programme design and additionality</i>			
6.1 FCDO–Unilever objective misalignment / tension	0	10	0
6.2 Post-grant transition / funding gap	6	1	3
6.3 External shocks	6	1	4
6.4 Catalytic additionality	21	7	11
6.5 FCDO–Unilever objective alignment / complementarity	3	12	3
<i>7. Greening</i>			
7.1 Green selected	7	3	7
7.2 Green value / claim capture	5	4	4
7.3 Consumer behaviour / adoption constraint	4	2	2

continued

Code	Founder refs	Staff refs	Founders
7.4 Aggregation / smallholder fragmentation	7	0	4
7.5 Sustainability time-horizon asymmetry	1	2	1
7.6 Green built	4	1	4
8. <i>Chain direction</i>			
8.1 Upstream (input supplier)	5	0	5
8.2 Downstream (distributor/seller)	5	0	4
9. <i>Link at entry</i>			
9.1 Strong / active link	4	0	3
9.2 Weak / nominal link	6	0	4
10. <i>Founder origin</i>			
10.1 Founder: Global South	8	0	8
10.2 Founder: Global North	3	0	3
11. <i>Founder profile</i>			
11.2 Foreign-educated	1	0	1
11.3 Internationally networked	6	0	6
12. <i>Stage at engagement</i>			
12.1 Idea / prototype	3	0	3
12.2 Pilot	1	0	1
12.3 Scaling	4	0	4
12.4 Mature	4	0	3
13. <i>Source type</i>			
13.1 SME founder	12	0	12
13.2 Unilever / programme staff	0	7	0

Founder refs and Staff refs are coded references in the twelve enterprise and seven programme- or buyer-staff transcripts; Founders is the number of the twelve enterprise transcripts carrying the code at all; a code is counted once per transcript there however often it appears. Codes 1.7, 8.3, 9.3, 10.3, 11.1 and 11.4 exist in the code-book and were never applied. The Recycling B transcript carries 21 of the founder references.

E.2 The interview evidence by literature-derived mechanism

This part of the appendix gives the exhibits behind the interview evidence used in Sections 5.5 to 5.10 and 6. The source is the audited NVivo coding of the nineteen interviews that attach to the case analysis, twelve with founders or senior managers of the deep-tier enterprises and seven with Unilever and programme staff, read against the literature domains of Section 2. Three conventions govern every count. First, the founder and staff groups are counted separately, because they are two perspectives on the same relationships. Second, an absent code records that the transcript does not mention the mechanism, under the convention of Appendix C. Third, where a case is assigned to a domain, that assignment is a synthesis across cases and is labelled as such.

E.2.1 Which mechanisms are evidenced in which cases

Figure 1, in Section 5.5, shows, for each of the twelve interviewed enterprises, which of the fourteen mechanisms most central to the hypotheses and to programme value are evidenced in the founder's transcript. Reading down the columns gives the spread of each mechanism. Reading across the rows gives each enterprise's profile, and the row labels show the tracker's outcome so that the two can be set against one another. The buyer-governance block (live commercial demand, a named champion, a procurement or resourcing gap and restructuring) is the densest, and its codes are present among integrators and non-integrators alike,

Table 27: Interview codes present in each founder transcript, deep tier ($N = 12$), beside the document-coded conditions and the outcome

Firm	7.1	7.6	4.2	4.3	2.3	2.7	5.1	5.2	5.5	11.3	GREEN	DOWN	INT
Refill A	1	0	0	1	1	0	1	0	0	0	1	1	0
Platform A	0	0	0	1	1	1	1	1	1	1	0	1	1
Agri A	1	0	1	0	0	0	0	0	1	1	0	0	1
Platform B	0	1	0	1	1	1	1	1	0	1	0	0	0
Recycling B	1	1	0	0	0	0	0	0	1	0	1	0	0
Recycling A	1	1	1	0	1	1	0	0	1	1	1	0	1
Agri B	0	0	1	0	0	0	0	1	0	1	1	0	0
Platform C	0	0	0	1	0	0	1	0	0	0	0	1	0
Refill B	1	0	0	1	0	0	1	0	0	1	1	1	1
Agri C	0	0	1	0	0	0	0	1	0	0	0	0	1
Refill C	1	0	0	0	0	0	1	0	1	0	1	1	0
Agri D	1	1	1	0	1	0	0	0	0	0	1	0	0
Present in inclN (INT)	7 0.60	4 0.20	5 0.60	5 0.40	5 0.40	3 0.40	6 0.40	4 0.40	5 0.60	6 0.80	7 0.40	5 0.40	5 —

7.1 green credentials pre-dated the programme; 7.6 greening emerged through it; 4.2 mediated supply; 4.3 commercial platform; 2.3 internal champion; 2.7 champion lost; 5.1 buyer resource-allocation gap; 5.2 buyer restructuring or churn; 5.5 procurement demand nascent or absent; 11.3 internationally networked founder (TIE_PRIOR also codes 10.2). A 1 records that the code was applied to the founder's own transcript; a 0 records silence, not established absence. GREEN and DOWN are the entry-stage document codes of Table 13; INT the tracker outcome with the Agri C override. 7.1 agrees with GREEN on 10 of 12 (Agri A: interview yes, documents no; Agri B: the reverse). The modes agree with DOWN on 9 of the 10 firms carrying one (Platform B: platform in interview, upstream on entry-stage documents; Refill C and Recycling B carry neither). 4.1 direct supply is applied to no founder transcript.

which is consistent with the sufficiency null of Section 5.5. The founder-network column is filled for four of the five integrators and two of the seven that did not integrate, which is the TIE_PRIOR result of Section 5.2. The strict-additionality column is sparse everywhere, four cells in twelve, while the capability-transfer column is filled most often among the enterprises that did not integrate. The grid extended to all twenty-eight mechanisms in the coding-to-literature mapping is roughly half grey, which is why no claim in the paper rests on inferring absence. That grid is in the replication package.

E.2.2 Founder and staff evidence by mechanism

Table 28 gives the count of founder and staff interviews carrying each mechanism, and a reading of which group carries it. The two groups converge on the buyer-side mechanisms the sufficiency null needs. They diverge on two themes that founders are not positioned to observe, tension between the funder's and the buyer's objectives, and shifts in the buyer's sustainability priorities (Section 6).

Table 28: Founder and programme/buyer staff evidence by mechanism. Counts are interviews carrying the code: founder interviews out of 12, staff interviews out of 7. The two pools are counted separately and never pooled. A count of zero records silence, not established absence.

Domain	Mechanism	Founder (of 12)	Staff (of 7)	Reading
Selection / treatment	Prior Unilever link at entry	3	0	founder-weighted
	Founder international network	6	0	founder-weighted
	Capability transfer	8	5	both pools
Supplier readiness	Operational readiness gap	4	6	both pools
	Resource or technology gap	3	0	founder-weighted
	Compliance / certification cost	0	1	thin in both
	Vendor registration burden	2	2	thin in both
	Power / price pressure	4	0	founder-weighted
Buyer governance	Live commercial demand	3	7	both pools
	Named champion in buyer	5	2	founder-weighted
	Project-lead active support	3	6	both pools
	Procurement / resourcing gap	6	4	both pools
	Internal restructuring	4	3	both pools
	Sustainability priority shift	1	3	staff-weighted
Chain position	Upstream input supplier	5	0	founder-weighted
	Direct supply to buyer	0	2	thin in both
	Platform / channel role	5	2	founder-weighted
Environmental upgrading	Green at entry	7	2	founder-weighted
	Green built in relationship	4	1	founder-weighted
	Green value capture	4	2	founder-weighted
	Sustainability time horizon	1	2	thin in both
Durability / programme value	Repeat / reinvested business	1	0	thin in both
	Champion loss	3	2	founder-weighted
	Brand signalling / credibility	7	3	both pools
	Wider-market access	4	3	both pools
	Post-grant funding gap	3	1	founder-weighted
	External shock	4	1	founder-weighted

E.2.3 Where the quotations come from

Because the paper relies on quoted material that a reader cannot check against the transcripts, this subsection reports its provenance. We matched every quotation in the paper back to its source. The ninety quoted passages divide into thirty-five drawn from the interview transcripts and fifty-five quoted from documents (application forms, milestone and closeout reports, the tracker’s own fields and the programme’s partner guide) or used as terminology instead of speech. The thirty-five interview quotations come from thirteen different people, nine of the twelve interviewed founders and four of the seven programme and buyer staff. Table 29 gives the distribution.

Table 29: Provenance of the interview quotations in the paper

Source	Speakers quoted	Quotations
Refill enterprises	2	9
Downstream distribution platforms	3	6
Upstream agricultural suppliers	2	4
Upstream recycling enterprises	2	3
<i>Enterprises</i>	<i>9</i>	<i>22</i>
Senior programme managers	2	5
Regional programme managers	1	5
Thematic and country-market staff	1	3
<i>Programme and buyer staff</i>	<i>4</i>	<i>13</i>
<i>All interview quotations</i>	<i>13</i>	<i>35</i>
Notes: the interviewed pool is twelve enterprises and seven programme and buyer staff. Three enterprises and three staff are quoted nowhere in the paper. The largest single contribution is seven quotations, from one refill founder, five of them in Section 5.6; the second largest is five, from a regional programme manager. No quotation in the paper is attributed to a named individual, and no persistent code identifies a speaker across quotations, for the reason given below. Source: an audit of every quoted passage in the compiled paper against the interview corpus and the project document set, 11 September 2026; the working table is filed with the analysis.		

The quotations are spread across all four business-model groups in the case design and across three of the four staff roles, so the argument does not rest on a single account. The densest passage in the paper, the account of why a well-disposed counterpart could not move a decision in Section 5.6, draws its sixteen quotations from nine different people, six founders and three staff, with no more than five from any one of them. The distribution is nonetheless uneven, and one refill founder supplies a fifth of all quoted interview material. The four quotations in the sentence about the local team’s incentives are that same founder speaking continuously, and a reader should treat that passage as one account, corroborated elsewhere by the coding counts instead of by further quotation. Three founders and three staff are not quoted, because their transcripts yielded no compact, quotable formulations. The mechanism counts reported above are taken from all nineteen transcripts, quoted or not.

We identify speakers in the text by role descriptors, such as a refill founder, a downstream distribution platform, an upstream agricultural supplier, or a senior or regional programme manager, and never by a persistent code. Each descriptor covers at least two enterprises on the published case list, so it does not identify one. A persistent code would allow a reader to assemble several quotations into a profile and match it against the case list in a single step. Where recurrence matters to an argument we say that the same founder is speaking, within a paragraph only.

E.2.4 Green credentials at entry and during the relationship

Table 30 sets the interview code for a green characteristic pre-dating the programme against the code for one added, raised or required through it, case by case. It is the evidence for the finding of Section 5.7 that green characteristics were mainly selected at entry and sometimes deepened during the relationship.

Table 30: Green characteristics pre-dating the relationship (NVivo 7.1) against green characteristics added, raised or required through it (7.6), 12 interviewed enterprises. A blank records silence in the transcript, not established absence; the one case of affirmative evidence that nothing was added is noted.

Enterprise	Tracker outcome	Green at entry	Green built	Verdict
Platform B	Not integrated		•	Built only (weak)
Platform A	Integrated			Neither coded
Agri B	Not integrated			Neither coded
Platform C	Not integrated			Neither coded
Agri C	Integrated			Neither coded
Recycling B	Not integrated	•	•	Selected and deepened
Recycling A	Integrated but later discontinued	•	•	Selected and deepened
Agri D	Trying to integrate	•	•	Selected and deepened
Refill A	Not integrated	•		Selected only
Agri A	Integrated but later discontinued	•		Selected only
Refill B	Integrated	•		Selected only
Refill C	Not integrated	•		Selected only
Total		7 of 12	4 of 12	

One upstream agricultural supplier gives affirmative evidence that nothing was added: asked directly, the founder said “no...I don’t think there’s any certifications that we got for having participated in TRANSFORM.” The single built-only case is weak: the capability was self-initiated in response to a logistics constraint rather than required by the buyer, was operational rather than a standard or certification, and did not outlive the operation. No commercial premium follows in any group. Agricultural certifications appear to be under-captured by the green constructs; this is stated as a measurement limitation and no condition has been recoded.

E.2.5 The cases the configurations do not reach

Table 31 describes each case that behaves otherwise than the configurational analysis or the hypotheses would lead one to expect, with the domain of its binding constraint and what the condition set misses. Table 32 compares the mechanisms evidenced among the cases a reported configuration reaches with those it does not, first over every specification and then over the seven document-derived specifications only. The second pair is the one Sections 5.6 and 5.10 use, because two of the reported specifications are built on TIE_PRIOR and a case can be reached because its founder had an international network.

Table 31: Cases that do not behave as the configurational analysis or the hypotheses would lead one to expect. The domain of the binding constraint is a synthesis across cases, not a coded variable.

Enterprise	Tied to	Domain of binding constraint	Why deviant	What the conditions miss
Refill A	H1 and outcome definition	Outcome definition	Carries the most favourable selection profile of the seven non-integrating cases – a senior named champion, live commercial demand and vendor registration – yet is coded not integrated.	The outcome variable cannot distinguish a registration state from an evidenced transaction.
Platform B	H2 and programme value	Buyer governance	The only non-integrating case meeting the strict catalytic-additionality criterion: a platform that would not exist without the programme, and no commercial conversion.	No condition records whether a commercial sign-off route exists, or whether technical access is permitted.
Agri B	H2 and chain direction	Supplier-side and structural	Non-integration is not explained by buyer-side conditions: no champion, resource-gap or restructuring code appears anywhere in this transcript.	No condition measures upstream feasibility – input certification, aggregation cost, agronomic risk – which is where this constraint sits.
Platform C	H2	Buyer governance and supplier-side disengagement	Buyer-side support existed but could not convert, and the enterprise itself moved away from the Unilever-linked model.	No condition records the existence of a sign-off process, nor the SME’s own strategic redirection.
Refill C	H2	Buyer governance	Formal commercial standing existed and was declined: the buyer holds a first right of refusal over the categories and did not exercise it.	Conditions record whether a relationship exists, not whether the person holding it can commit the buyer.
Agri D	H2, green corollary	External and unfinished process	The only non-integrating case with continuous, non-departed champions; also the cleanest instance of buyer-requested environmental upgrading.	A cross-section taken mid-relationship cannot distinguish an unfinished process from a blocked one.
Recycling B	H2	Unclassified / no binding-constraint domain coded	Seventh QCA-unreached non-integrator (twelfth SME interview, added 4 Sep 2026): an aspiring PCR/packaging supplier with no Unilever transaction at interview.	The transcript does not carry the specific mechanism coding needed to locate the binding constraint; this is not a claim that no constraint existed.
Agri A	Outcome definition; H2, green corollary	Outcome definition	Coded integrated by the tracker while the founder denies meaningful integration; also gives rare affirmative evidence that no green capability was added.	The outcome threshold counts a single intermediary-mediated season as integration.

Table 32: Mechanisms among configurationally covered and uncovered cases, before and after removing the interview-coded specifications

Mechanism	All specifications		Document-derived only	
	Covered	Uncovered	Covered	Uncovered
Founder international network	4/5	2/6	3/4	3/7
Strict catalytic additionality	3/5	1/6	2/4	2/7
Capability transfer	2/5	5/6	1/4	6/7
Procurement / resourcing gap	2/5	4/6	1/4	5/7
Operational readiness gap	3/5	1/6	3/4	1/7
Wider-market access	0/5	3/6	0/4	3/7
Live commercial demand	1/5	2/6	1/4	2/7
Named champion in buyer	2/5	3/6	2/4	3/7
Green selected at entry	3/5	3/6	2/4	4/7

The left pair defines coverage using every reported specification, including M-I1 and M-I2, which are built on TIE_PRIOR — the condition that operationalises a founder’s prior international network. A case can therefore be covered *because* it has such a tie, so the founder-network row on the left is partly mechanical. The right pair uses only the seven document-derived specifications. Refill B is covered on the left and not on the right, which is the whole of the difference. Counts are enterprises whose founder transcript carries the mechanism; the balance is transcript silence, never read as absence. Source: Analysis 7A.

E.2.6 How the interviewed enterprises differ from the rest of the portfolio

Figure 3 compares the ten interviewed enterprises that carry a tracker row with the other eighty on the portfolio frame, as standardised differences. They differ substantially on two dimensions, ever-integrated and a Medium or High starting link, and they under-represent the locally originated referral route. On chain position, multi-project status, circular pillar, cohort and region they are unremarkable. The sample is purposive by design, so these gaps describe what the interview evidence covers, and no test is reported (Section 4.2.1).

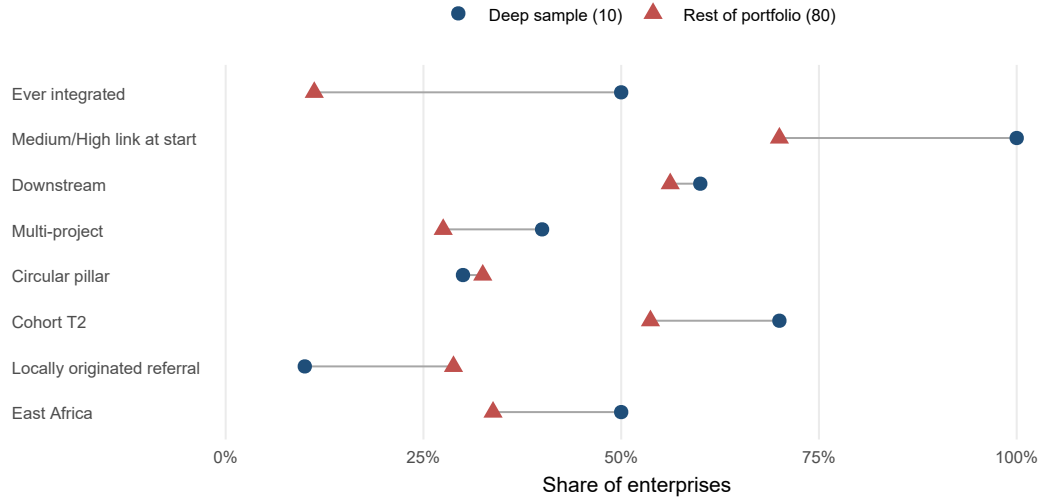


Figure 3: Composition of the interviewed cases against the rest of the portfolio

Notes: standardised differences between the ten interviewed enterprises with a tracker row and the other eighty enterprises of the portfolio sample. The interviewed cases are purposive, so the gaps describe coverage and not a sampling failure.

F Results reported in full

This appendix carries in full four results the main text reports in a paragraph. They are the green comparison behind the credential corollary of H2, chain direction, the monitoring-data comparison of what integration is worth, and what the enterprises’ own closeout reports record about the relationships the tracker counts.

F.1 The green comparison in full

Section 5.7 reports the credential corollary of H2 on three measures of green, and this subsection gives what the text summarises. The GREEN condition is satisfied by any one of a verifiable certification, a stated founding mission or membership of a recognised green sector, which are not equivalent. Appendix C separates certification (CERTIFIED, held by 2 of 12 interviewed and 2 of 56 document-coded enterprises) from sector membership (GREEN_SECTOR, 5 of 12 and 23 of 56), and neither enters a sufficient term or changes the null. The interviewed cases were also selected on circular-economy business models, while the condition is coded on credentials at entry, and the two differ in two of the twelve (Section 4.2.1). The document-coded contrast of 13.3 against 23.1 per cent appears only when conditions are coded from application-stage documents, which record what the enterprises claimed about themselves at entry, and it was flat on the full record (Appendix B). On the portfolio, two keyword readings of the tracker’s pillar field give a contrast of opposite sign, and the whole difference is whether off-grid energy access counts as green. Eleven enterprises separate the two readings and every one is an Energy project, which is why the text uses the programme’s own pillar labels and reports each family on its own (Table 33).

Table 7 in the text reports the environmental-against-social contrast as a single split, which makes the social side a residual category although the social pillars are the larger half of the portfolio. Table 33 therefore gives every family its own row, on the same 90 enterprises, with the energy pillar separate because where energy falls is what divides the two keyword readings of this field.

Table 33: Integration by the programme’s pillar family (ever integrated, $N = 90$)

Pillar family	Enterprises	Ever integrated	% integrated
Environmental	33	5	15.2
Health and WASH	21	2	9.5
Livelihoods and future of work	29	6	20.7
Energy	7	1	14.3
All	90	14	15.6

Notes: the programme’s “reclassified pillars for integration analysis” field, grouped into families and covering all 90 enterprises with no residual category. Environmental is Recycling, Refill and Regen Ag; health and WASH is health, digital health, water and sanitation; livelihoods is farmer livelihoods, retailer livelihoods and the future of work. One label (“Digital Health”) is a misspelling of another and is folded into it, which moves three enterprises. Fisher’s exact test across the four families returns $p = 0.80$ (simulated, 2×10^5 replicates). The environmental family against the other three is 5 of 33 against 9 of 57 ($p = 1.00$); the two social families pooled against the rest is 8 of 50 against 6 of 40 ($p = 1.00$).

No family is distinguishable from any other (Fisher’s exact test in the notes to Table 33). The spread runs from 9.5 per cent in health and WASH to 20.7 per cent in livelihoods, and cells of this size cannot distinguish that spread from chance. The two families the paper calls social are on either side of the environmental rate. Neither social family has a rate that differs from the environmental rate, so the green null of Section 5.7 does not arise from pooling a heterogeneous social side. The highest-rate family is a social one, which is the reverse of the ordering a green premium predicts. We report the whole table because a single environmental-against-rest split invites the reading that the residual is a single category.

The interviews add detail to the green result of Section 5.7. Seven of the twelve interviewed enterprises arrived with a green characteristic that pre-dates the programme, and three of those seven had it deepened through the relationship (Appendix E.2, Table 30). In the clearest case the buyer’s own supply-chain staff asked whether the crop could “get certified under any of the sustainability programs so that we can buy it,” and the enterprise approached Rainforest Alliance. No order has yet followed. This is the mechanism Archer and Elliott (2021) document in the same buyer’s Kenyan tea chain, observed here at the moment of request. In the other case an upstream recycler attributes a fair-trade certification route to the relationship, the

one such claim in the corpus corroborated from the buyer’s side. Against these, one upstream agricultural supplier gives the clearest negative evidence in the corpus, answering directly that no certification was added because of the programme, having listed several held beforehand. Green characteristics in this portfolio are therefore mainly selected at entry and sometimes deepened during the relationship, once at the buyer’s explicit request, and neither group shows a higher rate of commercial integration. One measurement limit applies to this result. The green constructs appear to under-capture agricultural certification, since two upstream agricultural suppliers holding organic, fair-for-life, GlobalGAP or ISO certifications at entry are coded not green in the document-coded condition. Both are integrated or formerly integrated, so the bias runs against the green category on the cases most likely to sustain a commercial relationship, and it cannot produce the null we report. No condition has been recoded on the strength of interview evidence collected for a different purpose.

Selecting on green credentials has two possible costs, which should be kept apart. Integration is scarce across the whole portfolio, so the first cost falls on what the programme can hope to achieve. Where the enterprises’ own credentials are the measure there may be a second, because selecting on them concentrates the portfolio in the part of the value chain where the buyer’s procurement function is least developed. Type and value-chain position are nearly collinear here, so the green contrast and the direction result below cannot be distinguished in this sample.

F.2 Chain direction

Section 2.7 sets chain direction aside; we give the detail here, on the portfolio sample. The condition is described in Appendix A.

Chain direction is nearly collinear with enterprise type, so we report it here as a robustness result. The unconditional contrast is flat, at 17.9 per cent upstream against 13.7 per cent downstream on the binary measure in the portfolio sample, with a bivariate ordered-logit coefficient of +0.08 ($p = 0.88$). Supply-chain direction is nonetheless the one coefficient in the multivariate models of Appendix D (full tables in the replication package) that matters under joint estimation. Adding it gives a positive downstream coefficient that is marginal in the ordered specification ($\hat{\beta} = +1.32$, $p = 0.052$) and not significant in the binary ($+1.09$, $p = 0.164$). The result is conditional. Downstream position predicts a higher rung on the integration ladder among enterprises with comparable Unilever link ratings, while the unconditional contrast, reported above, is flat. The variables it is conditional on include two measured during the project, which cannot serve as predictors, and conditioning on them may itself generate the coefficient. We show in Section 6 how the coefficient behaves across sample bases, without resolving between the two readings. With those qualifications, the sign is the opposite of the backward-linkage prediction, and we read it as setting-specific. The upstream inputs here are nascent recycled and regenerative materials for which Unilever’s procurement function is still forming.

The interviewed cases’ upstream pattern has weakened to the point where we no longer rely on it. UL_LINK_START appears in no deep solution, and ~DOWNSTREAM is present in three of the five integrated enterprises. The sample was selected on circular-economy business models, which are overwhelmingly upstream, so its upstream cases are integrated close to by construction. No larger sample reproduces the pattern, and the larger samples do not agree with one another. The portfolio gives 17.1 per cent upstream against 12.3 downstream in the regression reading and 17.9 against 13.7 in the configurational reading, the document-coded sample gives 15.2 against 30.0, and the regression gives a negative conditional upstream coefficient. Both document-coded tiers agree on every condition cell across the ten enterprises they share, so the divergence arises from case selection, and the reading of direction (Section 7) rests on the regression.

F.3 Integration and reported impact

This subsection reports the exercise behind Section 5.9. The monitoring workbook is a single cross-section of cumulative totals to June 2026, so nothing in this subsection is a before-and-after comparison. Every estimate contrasts enterprises that integrated with enterprises that did not, observed once. The enterprises reported the totals during and at the close of their projects, and for many of the integrated enterprises that was before the relationship the tracker records had formed. The timing is therefore inconsistent with a reading in which integration produced the jobs. If anything it allows the reverse, that the enterprises already

creating jobs at scale were the ones the buyer went on to trade with. The prior expectation is also a null. The employment and productivity gains in the linkage literature are estimated on firms that become direct, first-tier suppliers to a multinational, and Alfaro-Ureña et al. (2022) find none for second-tier suppliers. Section 5.8 shows that, among the relationships we can observe in detail, what this portfolio calls integration is seldom first-tier supply. The question is therefore whether the data show a gain anyway and, if they do, whether it can be told apart from the buyer’s selection.

Two of the paper’s results make that comparison worth attempting. Integration requires a match at entry, a project the buyer’s side could use (Section 5.2), and among enterprises with one no supplier characteristic we observe is sufficient for it (Section 5.5). They motivate restricting the comparison to the enterprises with a match at entry, which we call the gate-passing enterprises, on the reasoning that what separates them is less likely to be supplier quality. The restriction does not establish that the remaining variation is ignorable. The configurational null shows that our conditions do not sort the integrated from the rest, and the interviews identify buyer-side determinants we cannot observe. What follows is therefore a descriptive sensitivity analysis of how far the association moves as the comparison is tightened. It is not an identification strategy. The monitoring workbook matches 86 enterprises of the portfolio sample, 14 of them integrated. Of these, 64 passed the entry gate, and they contain all 14 integrations. Table 9 (Section 5.9) reports quasi-Poisson rate ratios on cumulative totals as the design tightens, from the raw cross-section, then with cohort, chain direction and application-stage scale held fixed, then restricted to the gate-passing enterprises, then on coarsened exact matching strata over the same three covariates, and finally with pillar family added.

The ladder itself is Table 9 in Section 5.9. An exposure-adjusted specification was estimated and dropped. A project duration is computable for only 43 of the 86, and on the 32 enterprises carrying both it returned an interval of [2.61, 393.76].

The beneficiaries rate ratio is indistinguishable from one once the first control enters and remains so at every later rung, so there is nothing to report for that outcome. The jobs rate ratio persists across the rungs of the ladder. It does not shrink as covariates are added, under the restriction to gate-passing enterprises or under matching. Two placebo comparisons, contrasts that ought to show a gap if the jobs result were only selection, support it. The five enterprises the buyer engaged and never bought from, the tracker’s “trying to integrate” category, report no more jobs than the rest (rate ratio 0.49, $p = 0.557$). Among non-integrators a High starting link, which Section 5.2 shows does not predict integration any better than Medium, does not predict jobs either (1.05, $p = 0.936$). With fourteen treated enterprises the model-based p -value cannot be trusted, so we test the jobs estimate by permuting the integrated label ten thousand times within the gate-passing set. That returns $p = 0.072$ against a model-based 0.004, an order of magnitude apart, and the permutation value is the one we report.

Integrated enterprises therefore report more jobs, and the association holds under every restriction in the ladder. We nonetheless do not read it as a return to integration, for three reasons, and it does not hold under a fourth restriction, taken up at the end of this subsection. The first reason is the magnitude of the estimate. Alfaro-Ureña et al. (2022) estimate a 26 per cent employment gain from a first multinational sale, a rate ratio of 1.26, on a matched panel with pre-event trends in the cleanest setting the literature has. Ours is 15.32, twelve times theirs, in a portfolio that contains almost none of the first-tier relationships through which their gains run. An estimate that large in that setting is more plausibly measuring differences between the enterprises than an effect of integration. The second reason is composition. Within the gate-passing enterprises the median integrated firm reports 17.5 direct jobs against 9.5 for the median non-integrated one, a ratio of 1.84, against a ratio of means of 8.93 and an adjusted rate ratio of 15.32. A handful of large enterprises therefore accounts for the estimate, and the underlying distribution is mostly zeros, with five of the fourteen integrated enterprises and twenty of the fifty gate-passing non-integrated ones reporting no direct jobs. The third reason is that the estimate is sensitive to which enterprises are in it. Leaving out one pillar family at a time moves the gate-restricted rate ratio between 9.78 and 46.41, a wider range than the whole ladder of Table 9 produces (Appendix F.3.3). The gate restriction also shows how much of the raw contrast was selection. Across the whole matched sample the median integrated enterprise reports more than four times the jobs of the median non-integrated one (the raw medians below), and confined to enterprises whose project the buyer had already judged it could use the multiple falls to 1.84.

The programme’s own attribution measure, “individuals employed as a result of TRANSFORM funding”, is identical to total employment for every enterprise across all 122 rows of the source workbook. Every job an enterprise reports is therefore credited to the programme, and the programme’s claim for itself cannot

be separated from the enterprise’s headcount. We record that as a property of the monitoring instrument. The totals are also unadjusted for exposure. Cumulative counts grow with time, first-generation enterprises have had roughly a decade, and a duration is computable for fewer than half the sample, so the adjusted specification was too imprecise to report (above).

Integrated enterprises report more jobs than non-integrated ones, and once the entry gate and observable scale are held fixed that gap cannot be separated from the buyer’s selection. The magnitude is implausibly large to interpret as an effect of integration, it rests on a few large enterprises, and the totals were mostly reported before the relationships formed. The association may also run through what the enterprise was funded to do, since the monitoring instrument records different outcomes for different families. A digital health platform reports beneficiaries and a livelihoods enterprise reports jobs. Cutting Table 9 by pillar family bears this out. What each family reports differs before any enterprise integrates anything. Of the 32 environmental enterprises, 25 report a positive jobs total, against 7 of the 21 in health and WASH, and the median environmental enterprise reports 16.5 direct jobs against a median of zero in health and WASH. The jobs association is not confined to one family. Within the environmental family the gate-restricted rate ratio is 2.83, within livelihoods 23.33 and within health and WASH 50.90, so it is present, and weakest, in the family that holds most of the portfolio, but each within-family interval spans an order of magnitude either side of its estimate. What can be estimated with more precision is the pooled ladder with family held fixed, and it moves the result. Adding pillar family takes the cross-sectional rate ratio from 10.66 to 9.40 and the gate-restricted one from 15.32 to 11.42, and the permutation test on the gate-restricted estimate goes from $p = 0.07$ to $p = 0.13$. Dropping one family at a time moves it further, from 9.78 without the environmental enterprises to 46.41 without the livelihoods ones (Appendix F.3.3). We therefore report the jobs association as one that holds across the design ladder of Table 9 and does not hold once pillar family is added. With family in the model it is no longer distinguishable from chance on the test we prefer, and the composition of the portfolio by pillar accounts for part of the association the design was meant to isolate. Given how seldom what this portfolio calls integration is first-tier supply, that is what the linkage literature predicts. The count is small, and we cannot show that integrated enterprises have better outcomes because they integrated (Section 5.9).

F.3.1 Raw medians, funding, and lapsed against active relationships

Ninety-one of the 120 enterprises in the monitoring dataset join the portfolio sample, of which 86 have an integration status and 14 are integrated. Eleven enterprises with monitoring data have no tracker row, a difference in coverage between two of the programme’s own instruments that bounds any portfolio claim about how impact and integration relate. Integrated enterprises are larger on every dimension the data measure, and both distributions are severely right-skewed. The median integrated enterprise reports 18 direct jobs against 4, 20,676 beneficiaries against 662, and GBP 345,706 against 14,237 on the monitoring workbook’s field “Total funding leveraged (GBP)”. The second monitoring workbook resolves that field into private capital and grants raised from other funders after the TRANSFORM grant (Appendix F.3.2), and it is not the grant, the match funding or the follow-on funding the tracker records. Of the non-integrated enterprises, 31 of the 72 report zero on it, and 3 of the 14 integrated enterprises do. We do not condition on funding, for three reasons. The field is a different quantity from programme spending, and funding leveraged is plausibly a consequence of integration (Wilcoxon $p = 0.089$). A per-pound model also forces the elasticity of the outcome on funding to equal one where the freely estimated elasticity is near zero. Conditioning on log funding as a free covariate leaves the estimates essentially unchanged. The binary outcome also collapses a distinction the tracker draws, and it gives the only within-integrated variation the data contain. Of the integrated enterprises with monitoring data, six still held the Unilever relationship in March 2026 and seven had lost it. If the impact gap reflects selection, the two groups should look alike and both should exceed the never-integrated. If integration produces impact, the enterprises that retained it should be ahead. The estimates run the other way. Against never-integrated enterprises the lapsed group’s rate ratio is 6.15 for beneficiaries and 17.62 for jobs, the still-active group’s 1.53 and 4.39, and the currently integrated are not distinguishable from the never-integrated on beneficiaries. Seven and six enterprises cannot support an inference, and the comparison is confounded in a knowable way, since lapsed relationships integrated earlier and have had longer to accumulate cumulative totals. The one comparison that holds selection into integration fixed therefore does not show currently integrated enterprises ahead, which a stepping-stone

account would require.

F.3.2 Capital raised and revenue during the grant

Section 5.9 asks whether integrated enterprises report more jobs and beneficiaries. A second monitoring workbook, received in September 2026, lets us ask whether they were larger during the grant on two measures the first workbook does not carry. These are capital raised from funders other than TRANSFORM, split into private capital and grants, and the enterprise’s revenue over the grant period. Both are reported quarterly from October 2021 to June 2026 and summed here. Of the 90 portfolio enterprises, 76 match the workbook by name, 13 of the 14 integrated among them. Table 34 reports the comparison on three samples, the matched portfolio, the enterprises that passed the entry gate, and the gate-passing second-generation enterprises, and then Firth logits on the gate-passers with cohort held fixed.

Four properties of the workbook decide how the table is read, and the notes to the table restate them. First, the revenue sheet’s own total is not enterprise revenue. The programme tags as “sales” every monetary line an enterprise reports, which includes the income of the waste pickers, farmers and agents it works with, the savings of its customers, the wages of staff hired for the project and, for one enterprise, revenue earned by Unilever, and it adds totals to their own components. We rebuild revenue from the indicator lines, keeping only the enterprise’s own sales. Second, revenue is reported only from October 2021 and only for the grant period, so a first-generation enterprise appears only if it held a later project. Of the 8 integrated first-generation enterprises, 5 do, and 3 of the 32 non-integrated do, which is why the revenue comparison is read within the second generation and why the cohort term in a revenue model measures that coverage. Third, two enterprises’ revenue lines are sales to Unilever itself, so for them revenue is the outcome, and the logit is repeated without them. Fourth, the first-generation capital figures are a single cumulative entry for 2016–2021 carried over from the previous monitoring provider, with genuine quarterly reports thereafter. One enterprise reports own revenue of GBP 4.6 million over five quarters, an order of magnitude above every other and inconsistent with its grant and its employment. It is treated as missing and has been queried with the programme.

Integrated enterprises raised more private capital across the matched portfolio, a median of GBP 279,450 against zero ($p = 0.05$), and more of them raised any (54 against 30 per cent). Grants do not differ on any sample, and own revenue does not differ once the two Unilever-sales lines are set aside. The private-capital gap belongs to the first generation. Restricted to the gate-passers it is no longer distinguishable from chance, and among gate-passing second-generation enterprises a third of each group raised any private capital and the medians are zero on both sides. In the Firth logits the odds ratio on log private capital is 1.04 and on log grants 1.01, and the cohort term absorbs what the raw comparison showed. The timing points the same way. Of the GBP 29 million of private capital the integrated first-generation enterprises report, GBP 7.8 million is in the carried-over 2016–2021 entry and GBP 21 million was reported from 2022 onwards, after their projects had closed, most of it by one enterprise. Capital of that kind follows a relationship with a global buyer, or an enterprise’s survival. It was not a condition for either, and among the enterprises the buyer could use it does not separate the ones it traded with from the rest.

F.3.3 By pillar family

Section 5.9 reports that the jobs association does not hold once pillar family is added. This subsection gives the working, because the reason is partly a property of the measuring instrument.

The first panel of Table 35 reports what each family reports, integrated and non-integrated together. A quarter of the health and WASH enterprises report any direct jobs; three-quarters of the environmental ones do. The median health and WASH enterprise reports zero jobs and zero beneficiaries. The monitoring instrument records people for enterprises funded to reach people and staff for enterprises funded to process material, so a zero in the jobs column for a sanitation enterprise reflects what the instrument asks of it.

Panel B is reported because it was asked for. It is informative about the sign of the association within families more than about its size. Three of the four families return a rate ratio above one, so the association is not confined to a single family. The within-family magnitudes are reported but imprecise. The fourth family holds a single gate-passing enterprise and cannot be fitted, the health and WASH estimate rests on two integrated enterprises in eight, and every interval that can be computed spans at least an order of magnitude. Panel C asks the same question of the pooled model, with family held fixed. Family absorbs

about a quarter of the gate-restricted estimate and rather more of the covariate-adjusted one, and the randomisation-inference p moves from 0.067 to 0.125. Panel D shows why the paper does not rely on the residual. The estimate ranges from 9.78 to 46.41 depending on which family is left out, which is a wider range than the design ladder of Table 9 produces across all four of its rungs.

In Section 5.9 we draw a weaker conclusion from this than Table 9 on its own would support, for the reason given there. The jobs association holds under every restriction of the design and does not hold once pillar family is added. Because the instrument records different quantities for different families, we cannot tell whether family is a confounder or whether family determines which outcome is recorded, and on neither reading do we read the association as a return to integration.

F.4 What the closeout reports record

The outcome records whether a commercial relationship existed. The closeout documents add what kind, and they let us code it for twelve of the fourteen integrated enterprises. Nine have an endline report and three more a final milestone report, and only Recycling J and the parent organisation have neither. Table 36 sets out the coding. None of the twelve records direct first-tier supply to Unilever under a contract. One documents supply through an intermediary. Five are commercial platforms in which the enterprise sells Unilever product onward to its own customers. Six record no payment flow with Unilever in either direction, the enterprise collecting or processing material in an arrangement that contributes to a corporate commitment without constituting a sale.

The coding is close to a restatement of chain direction. The five platforms are all downstream and five of the six no-payment-flow cases are upstream, the exception being WASH B, whose arrangement is an advertising one. What makes it worth reporting separately is that the buyer’s own records agree with it. The tracker has a value-proposition field, recorded by the programme team against each project and independent of the enterprises’ reporting. It classifies every one of the five platforms as strengthening routes to market, and the intermediary and material-flow cases as sustainable suppliers or as a contribution to Unilever’s goals.

Agri A is the one case in which the two sources together record supply reaching a Unilever product. Its final report records deliveries to a named third-party processor on three dates at a stated purchase price, and concludes that growing the crop for that processor is not commercially sustainable. The link to Unilever is recorded only in the tracker, where the programme team notes that the crop was being localised for a Unilever brand and that the enterprise “tried to supply . . . Unilever in Kenya but could not get yields/finances right”. This is the portfolio’s one documented instance of supply reaching a Unilever product, through a co-packer, and it failed on unit economics.

Two of the interviewed enterprises bear on the definitions, and we state what their reports record. The strict definition counts Platform A as integrated, and the returns literature would not. Its final report, written at project close in March 2024, describes the relationship more exactly than either the tracker or the guide. The pilot in its home market ran through Unilever’s largest key distributor there instead of with Unilever. The Unilever leadership team had asked the enterprise to demonstrate the model’s cost-effectiveness before operations grew. The report records commercial agreements for the enterprise’s technology and services as “in the process of closing” in two other markets. Where Unilever appears as a counterparty in that document it is as a prospective customer for a platform, which is what the endline coding above records. The four-country agreements the guide records are a later state of the world than the report describes. The strict definition excludes Recycling A, and the programme presents it as a supplier partnership. Recycling A’s endline report names the brands the enterprise supplies post-consumer plastic to, and Unilever is not among them. It records the termination of one of those brand contracts during the project, and it places Unilever and the programme in a single closing paragraph on attribution, as the funder without which the project would not have happened. The report’s own account of the relationship at project close is that TRANSFORM paid for the facility and other companies bought the material.

That case also explains why these reports appear to capture commercial detail poorly. The enterprise’s counterparty is frequently an intermediary. Agri A reports the processor, Recycling A’s founder describes the flow running through a converter, and Platform A’s pilot ran through Unilever’s key distributor. The reports are accurate about who the enterprise transacts with, and that party is often a step removed from the brand owner. An integration measure recorded at the brand owner will therefore omit some relationships that exist, and it will include some relationships that the enterprise does not experience as commercial.

Two of the twelve are coded from documentary silence instead of documented absence, and the distinction matters. Agri C is the one enterprise in the study with a first-tier supply relationship to Unilever, confirmed by the programme lead and pre-dating the programme. Unilever appears exactly once in its final report, as one of the three TRANSFORM partners, and the supply relationship is not mentioned. Recycling F names Unilever three times, each inside a list of partners and resources in a business-model diagram, and records its material as sold to authorised end destinations. Neither report denies a relationship, and each is silent about one. A report that does not mention a relationship is therefore weak evidence that none existed, as Agri C shows. Its supply relationship is confirmed by the programme lead, and its report does not mention it.

Whether the kind of relationship predicts whether it survives cannot be answered here. Two of the five platforms and three of the six no-payment-flow cases were still integrated in March 2026, which on the eleven enterprises with a position on the tracker’s ladder is no difference (Fisher exact $p = 1.00$).

F.5 Governance form at the Unilever interface

The coding above records what kind of relationship the closeout reports describe. In this subsection we code the form the relationship takes and what determines it, which is the prior question in the literature. Gereffi et al. (2005) answer with three properties of the transaction, which are the complexity of the information that must pass between the parties, the extent to which that information can be codified into a specification, and whether the supply base can already meet the requirement. They derive five forms from them, market, modular, relational, captive and hierarchy. Section 5.5 reports the result, and here we give the coding.

We code the three determinants for the twelve interviewed enterprises at two moments, from the application-stage documents (entry) and from the endline or latest milestone report (close). All twelve have a document at both moments. The unit of coding is the transaction with Unilever, as proposed at entry and as documented at close. Complexity is 1 where the transaction requires transferring detailed product or process specifications or bespoke adaptation between the parties, and 0 where it is an exchange of standard finished goods. Codifiability is 1 where those requirements can be written into an existing or newly created specification, grade, standard, procedure, certification or numeric benchmark. Capability is 1 where the enterprise can already meet the requirement of that transaction at that moment. The form is then derived from the framework’s own table instead of being assigned, and every one of the seventy-two cells has a quoted passage and a named source document.⁶ All twenty-four rows fall on one of the five forms, and none lands in a combination the framework marks as unlikely.

The first thing to note is that the forms are spread across all five categories at both moments, so the coding is not a relabelling of a binary. The second is that the distribution moves toward modular, from one enterprise at entry to four at close, and away from captive and hierarchy, from six to four. Six of the twelve change form. The third is which determinant changes. Capability changes in four of the twelve and codifiability and complexity in two each, and the capability changes are the substantive ones. Refill A reached ISO 9001 certification with laboratory tests showing no deviation between refilled and packaged product, Recycling B established a validated ten-tonne-a-month line and an independently tested hundred-per-cent recycled pouch, and Agri B demonstrated that farmers could meet a written specification, against Platform A, which could not complete the systems integration inside the grant period.

The fourth is the one Section 5.5 reports. Of the four enterprises at the modular form at close (the form the framework associates with arm’s-length supply by a capable supplier working to a codified specification), one records a Unilever transaction in its closeout document. That one, Refill A, buys through Unilever’s distributors instead of from Unilever, which is why the programme’s own integration measure scores it 0 while the goods flow is real. Recycling B names Unilever nowhere in its milestone report. Agri B’s report records that internal changes at Unilever delayed the planned offtaker introductions. Agri C’s report does not mention its supply relationship, as discussed above. The same asymmetry appears in the three enterprises with a live goods flow at close. Refill A, Platform C and Refill C all buy Unilever product through distributors or wholesalers instead of from Unilever, and all three are scored 0 on integration. We conclude that the three determinants describe whether a transaction could be conducted at arm’s length and leave undetermined

⁶Analysis/qca_pipeline/registry/gvc_governance.csv and 40_gvc_governance.R, which checks each derived form against the published table and each cell for its evidence passage. The protocol, the marginal calls and the checks are in notes/GVC_GOVERNANCE_20260911.md.

whether the buyer conducts it. A goods flow through distributors and an integrated supply relationship are different things, and the determinants describe neither directly.

Three checks bound the reading. They ask whether a determinant duplicates an existing condition, whether codifiability duplicates CODIFIED, and whether the forms can serve as a condition. No determinant is a relabelling of a condition in Appendix C, but entry-stage codifiability comes close to two of them. It agrees with PRIORITY_FIT on eleven of twelve ($\phi = 0.84$) and runs against a prior international tie at $\phi = -0.71$, either of which is high enough that the two could not enter a specification together. No other pairing reaches $\phi = 0.7$, and the closest at close is codifiability against PRIORITY_FIT at $\phi = 0.53$. The new codifiability differs from the existing CODIFIED condition, and the two agree on six of twelve at entry and four of twelve at close. The difference arises because CODIFIED records whether a specification existed at application, while codifiability records whether one could be written, which admits benchmarks the buyer sets and standards the enterprise writes itself. Bundling the five forms into the binary the framework's own axis suggests, market and modular against the three coordinated forms, does not produce a usable condition. At entry it is satisfied by four of the twelve with a sufficiency consistency of 0.50, and at close by seven with a consistency of 0.29. The close-stage version is in any case ineligible for a design coded on entry-stage evidence only (Section 4.2.3). We therefore report the coding descriptively and do not enter it into any specification.

Four caveats apply. Recycling B's close document is a third milestone report, so its modular coding describes a capability reached in a live project. Agri C's entry complexity is the coding most sensitive to where the transaction boundary is drawn. It is coded 0 because the product information is carried by third-party organic and fair-trade certification instead of being transferred between the parties, and drawing the boundary the other way makes the enterprise modular at entry as well as close and removes one of the six form changes. Agri B's entry capability is the second marginal call. It is coded 0 because the enterprise's record was in training and platform transactions instead of in delivering to a processor specification, and coding it 1 makes the entry form relational instead of hierarchy. Finally, the transaction column and the integration outcome measure different things. The outcome is an ever-integrated code and this column records the state at close, which is why Agri A and Recycling A appear as integrated enterprises with no transaction in their closeout reports.

Table 34: Capital raised and revenue during the grant, integrated against not

	<i>n</i> int.	<i>n</i> not	Integrated	Not integrated	<i>p</i>
<i>Portfolio, in the workbook</i>					
Any private capital (share)	13	63	0.54	0.30	0.12
Any grant (share)	13	63	0.77	0.59	0.35
Private capital, GBP (median)	13	63	279,450	0	0.05
Grants, GBP (median)	13	63	30,000	13,474	0.81
Own revenue, GBP (median)	11	36	56,397	24,414	0.35
<i>Gate-passers</i>					
Any private capital (share)	13	46	0.54	0.35	0.33
Any grant (share)	13	46	0.77	0.57	0.22
Private capital, GBP (median)	13	46	279,450	0	0.11
Grants, GBP (median)	13	46	30,000	9,110	0.64
Own revenue, GBP (median)	11	28	56,397	23,152	0.28
<i>Gate-passers, second generation</i>					
Any private capital (share)	6	35	0.33	0.29	1.00
Any grant (share)	6	35	0.67	0.46	0.41
Private capital, GBP (median)	6	35	0	0	0.70
Grants, GBP (median)	6	35	12,002	0	0.71
Own revenue, GBP (median)	6	27	62,022	22,949	0.21
<i>Firth logits on the gate-passing enterprises: odds ratio [95% interval], <i>p</i></i>					
Private capital (log) + cohort	<i>N</i> = 59, 13 events		1.04 [0.95, 1.15], <i>p</i> = 0.38		
Grants (log) + cohort	<i>N</i> = 59, 13 events		1.01 [0.89, 1.14], <i>p</i> = 0.93		
Own revenue (log) + cohort	<i>N</i> = 39, 11 events		1.15 [0.94, 1.69], <i>p</i> = 0.20		
Own revenue (log) + cohort, no Unilever-sales lines	<i>N</i> = 37, 9 events		1.10 [0.91, 1.55], <i>p</i> = 0.37		
Own revenue (log), second generation	<i>N</i> = 33, 6 events		1.16 [0.95, 1.84], <i>p</i> = 0.18		

Notes: from the second monitoring workbook (LSE_stats_TRANSFORM_Sept2026.xlsx), matched by name to 76 of the 90 portfolio enterprises, 13 of the 14 integrated among them. Private capital and grants are funds raised from sources other than TRANSFORM after the TRANSFORM grant, summed over quarterly reports from October 2021 to June 2026; for first-generation enterprises the figure is a single cumulative entry for 2016–2021 carried over from the previous monitoring provider, with later quarters reported as they occurred. Own revenue is the enterprise’s own sales over the grant period, which the programme attributes to the grant; the workbook’s revenue total also sums income earned by the people the enterprise works with, savings to its customers, the wages of project staff and in one case revenue earned by Unilever, and adds totals to their own components, so it is rebuilt here from the indicator lines. Revenue is reported only from October 2021, so first-generation enterprises appear only if they held a later project (5 of 8 integrated, 3 of 32 not), which is why the revenue rows are read within the second generation. Two enterprises’ revenue lines are sales to Unilever itself and are removed in the fourth logit. One enterprise’s revenue is set to missing as implausible and queried with the programme. Gate-passers have a Medium or High Unilever link at start. Shares are compared by Fisher’s exact test and amounts by the Wilcoxon rank-sum test; the logits are Firth-penalised with profile-likelihood intervals, cohort entered as a first-generation indicator.

Table 35: The jobs association within and across pillar families

<i>Panel A. What each family reports (all enterprises, integrated and not)</i>				
Family	<i>n</i>	Integrated	Median jobs	Reporting jobs > 0
Environmental	32	5	16.5	25 of 32
Health and WASH	21	2	0.0	7 of 21
Livelihoods and future of work	28	6	2.5	14 of 28
Energy	5	1	0.0	2 of 5
<i>Panel B. Gate-restricted rate ratio for direct jobs, within family</i>				
Family	<i>n</i>	Integrated	Rate ratio [95% interval]	
Environmental	31	5	2.83 [0.77, 10.45]	
Health and WASH	8	2	50.90 [1.42, 1825.92]	
Livelihoods and future of work	24	6	23.33 [2.17, 250.46]	
Energy	1	1	not estimable	
<i>Panel C. The pooled ladder, with and without family</i>				
Design	<i>n</i>	Integrated	Without family	With family
Cross-section	86	14	10.66 [4.11, 27.67]	9.40 [3.90, 22.64]
+ cohort, direction, scale	55	10	18.85 [5.55, 64.02]	10.75 [2.86, 40.43]
Gate-restricted	40	10	15.32 [2.78, 84.57]	11.42 [1.95, 66.76]
<i>Panel D. Gate-restricted estimate, dropping one family at a time</i>				
Dropped	<i>n</i>	Integrated	Rate ratio [95% interval]	
(none)	86	14	15.32 [2.78, 84.57]	
Environmental	54	9	9.78 [0.73, 130.81]	
Health and WASH	65	12	15.33 [2.55, 91.95]	
Livelihoods	58	8	46.41 [5.80, 371.54]	
Energy	81	13	14.78 [2.98, 73.32]	

Notes: quasi-Poisson rate ratios for direct jobs, integrated against not, on the 86 enterprises the monitoring workbook matches to the portfolio sample. Panels B and D are gate-restricted and adjusted for cohort (the generation of entry, as in Table 9), chain direction and application-stage scale where the cells allow it; Panel B's within-family models carry no covariates, since a family holds too few enterprises to fit them. Families follow Appendix F.1. Randomisation-inference p for the gate-restricted jobs estimate, permuting the integrated label ten thousand times: 0.067 without family and 0.125 with it. The 0.067 and the 0.072 reported in Section 5.9 are the same test under different seeds.

Table 36: What the closeout reports record about the kind of relationship

Enterprise	Mode	Evidence	Source
<i>Direct first-tier supply under a contract (0)</i>			
— none			
<i>Supply through an intermediary (1)</i>			
Agri A	intermediary	documented	milestone
<i>Commercial platform: sells Unilever product onward (5)</i>			
Platform A	platform	documented	endline
Health B	platform	documented	endline
Platform I	platform	evidenced	milestone
Platform K	platform	documented	endline
Platform L	platform	documented	endline
<i>No payment flow with Unilever in either direction (6)</i>			
Recycling F	no payment flow	silent	milestone
WASH B	no payment flow	documented	endline
Recycling A	no payment flow	documented	endline
Recycling L	no payment flow	documented	endline
Agri C	no payment flow	silent	endline
Recycling Y	no payment flow	documented	endline
Notes: the twelve of the fourteen integrated enterprises that have a closeout document. “Documented” means the report states the arrangement; “evidenced” that a single passage establishes the direction of the flow without a contract or a volume; “silent” that the report does not mention a Unilever relationship at all, which is not the same as recording its absence. Recycling J and the parent organisation have neither an endline nor a milestone report and are not coded.			

Table 37: Governance form at the Unilever interface, interviewed enterprises

Enterprise	Form at entry	Form at close	Transaction at close
Refill A	captive	modular	yes, through a distributor
Platform A	relational	captive	no, in due diligence
Agri A	hierarchy	hierarchy	no, a third-party processor
Platform B	relational	relational	no payment flow
Recycling B	captive	modular	no, Unilever unmentioned
Recycling A	modular	market	no, other brands named
Agri B	hierarchy	modular	no, introductions delayed
Platform C	market	market	yes, through a wholesaler
Refill B	captive	captive	partial, match funding unpaid
Agri C	market	modular	silent
Refill C	market	market	indirect, through wholesalers
Agri D	captive	captive	no, mapping not effected
Notes: the form is derived from three binary determinants by the table in Gereffi et al. (2005): low complexity with high codifiability and high capability gives market; high complexity with both high gives modular; high complexity with low codifiability and high capability gives relational; high complexity with high codifiability and low capability gives captive; and all three unfavourable gives hierarchy. Entry is the application-stage documents, close the endline or latest milestone report. The transaction column records what the closeout document establishes and is coded separately from the form, because six of the twelve have no Unilever transaction at close and the framework governs transactions.			

This material has been funded by UK International Development from the UK government; however, the views expressed do not necessarily reflect the UK government's official policies.



theigc.org
