Heterogeneous Effects of Theory-Based Entrepreneurship Training

Opening a Research Program

Helena Montoya Calero · May 2026

The research program. Theory-based entrepreneurship training works, on average. Camuffo et al. (2020, 2024) and Agarwal et al. (2025) establish this across eight RCTs and multiple countries. The natural next question is: *for whom does it work, and why?* This paper opens a systematic investigation into heterogeneous treatment effects (HTE) in theory-based training.

The gap. Camuffo, Gambardella & Jannace (2025, CEPR DP20300) provide the empirical starting point: using the same pooled RCT dataset (8 RCTs), they detect significant HTE along a vector of basic demographic covariates (gender, work experience, education, months operational, RCT location). The evidence that heterogeneity *exists* is clear. What remains unknown is *why* that is, which cognitive and behavioral traits of entrepreneurs, and which structural features of their ventures, generate differential responses to the same training. McKenzie, Woodruff et al. (*VoxDevLit*, 2025) identify this as the primary open question in the field: "There is a lot of heterogeneity in both samples and results, so there is less guidance on which groups benefit most." No paper has yet addressed this systematically for theory-based training.

The framework. I organize moderators into three categories—*person*, *venture*, and *dosage*—each tied to observable IBL baseline (and, where harmonized, follow-up) fields and to specific papers that motivate *which* gaps remain open after those papers.

Category	IBL measures (examples)	Literature anchor	What is still open / contribution
Person-level	Risk/ambiguity tasks (`risk_aversion*`, `unc_aversion*`); illusion of control; learning orientation and competitiveness; motivation and growth aspiration; attitudes toward unforeseeable events (`future_ev_*`, `future_events`); endowment items; education and experience (already in ERC).	Novelli & Spina (2024, *SMJ*) call for theory quality, confidence, and team composition as moderators not fully tested. Agarwal et al. (2025, *Org. Sci.*) move theory from mechanism to measurement but leave individual-level HTE thin. Drexler, Fischer & Schoar (2014, *AEJ Applied*) motivate cognitive *fit* of training content. Morris et al. (2023, *SEJ*) link mindset/self-efficacy to action. Kotha et al. (2023, *SMJ*): ambition and HC moderate growth training (adjacent treatment).	Little causal evidence ties *risk/ambiguity preferences* and *epistemic orientation* to the *scientific-approach* treatment itself; pooling ERC lets us pre-specify a short cognitive battery plus interactions.
Firm-level	Interview scientific-intensity block (`int_*`, `si_*`) plus survey stage proxies (`advanced_stage`, `months_working_1`, `full_time`, pivot/exit probabilities, `idea_aspiration*`); scope of alternatives (`idea_breadth`, `idea_distant`, `new_ideas*`); team mental models (`tms`); optional: valuation scenarios (`distribution_*`) and challenge/self-efficacy grids (`challenges*`) where harmonized.	Novelli & Spina (2024, *SMJ*): HTE by business-model development—effects can flip sign across maturity. Camuffo et al. (2024, *SMJ*): focused pivots and termination. Agarwal et al. (2025): coordinated business-model changes when theory is explicit. Santamaria et al. (2024, *SMJ*): demand-side information search (design contrast).	The field rarely uses *pre-treatment* measures of idea-space breadth, team cognition, and belief calibration alongside the same intervention pool; harmonizing ERC baselines addresses "how to measure comparable BM state across sites."
Dosage / uptake	`attendance`, `num_interv`, `has_interv` (ERC); where available, self-reported use of methods (`absorption_method` on treated-firm follow-up surveys).	Karlan & Valdivia (2011, *ReStat*): ITT vs. actual exposure. McKenzie et al. (*VoxDevLit*, 2025): field needs clearer guidance on who benefits—exposure is part of that story.	Linking dosage and uptake to the mechanism-level Z vector separates "assigned" from "received" training and connects to Chernozhukov et al. (2025, *Econometrica*) style discovery (BLP/GATES) without overfitting small-N sites.

These categories map onto whether the founder is *ready to learn* (person), whether the venture *needs* structured theory-based reasoning at baseline (venture), and whether the curriculum was *actually absorbed* (dosage). Follow-up-only fields (e.g. additional signal-intensity items on later waves) enter as controls or secondary mechanisms, not as pre-treatment HTE unless lagged by design.

First step: exploration. The starting point is an exploratory pass using the `GenericML` package (Chernozhukov, Demirer, Duflo & Fernández-Val, *Econometrica* 2025) and random forests over an expanded Z vector—baseline survey items above, plus baseline interview scores already in ERC—for the five pooled RCTs with harmonized baselines. Variable importance and BLP/GATES/CLAN objects summarize heterogeneity; a small set of theory-first interactions (Novelli-style maturity $\times$ treatment; risk/ambiguity $\times$ treatment) provides confirmatory estimates with multiplicity control.

Contributions. This paper reframes the heterogeneity question from "does X moderate the effect?" to "which *families* of mechanisms—beliefs, venture state, uptake—matter for theory-based training, and what does the pooled ERC data say?" It connects demographic HTE (Camuffo, Gambardella & Jannace, 2025) to mechanism-level moderators emphasized by Novelli & Spina and Agarwal et al., addresses the McKenzie et al. (*VoxDevLit*) call for sharper guidance on who benefits, and pre-structures measurement for future experiments that build heterogeneity in by design.