

Who Benefits from Scientific Entrepreneurship Training?

Systematic Analysis of Heterogeneous Treatment Effects — First Results

Helena Montoya Calero · May 2026 · Preliminary / Exploratory

The question. Theory-based (SDM) entrepreneurship training works on average—the ATE evidence is robust (Camuffo et al. 2020, 2024; Agarwal et al. 2025). The open question, explicitly identified by McKenzie et al. (*VoxDevLit* 2025) as the field’s primary gap, is: *who benefits more, and why?* This paper provides the first systematic, causally identified answer covering cognitive, venture, and belief-calibration dimensions of heterogeneity.

Data & method. Pooled 6-site RCT; $N = 488$ founders (T&E vs. Pure Control) at baseline; 5 follow-up periods. Outcome: $\Delta\text{SI} = \text{si_overall}_t - \text{si_overall}_0$ (change in Scientific Intensity). ATE: +0.60 SD in period 1, +0.68 SD in period 2 ($p < 0.001$ both). Method: Generic ML—BLP/GATES/CLAN (Chernozhukov, Demirer, Dufflo & Fernández-Val, *Econometrica* 2025). 250 random splits, paper-correct implementation.

Results: BLP scanner across all available moderators (12 variables, 250 splits each).

Moderator	Family	Period 1		Period 2		Verdict
		$\hat{\beta}_2$	p	$\hat{\beta}_2$	p	
illusion_of_control	Cognitive	-0.008	0.658	-0.154	0.449	Null. RF-predictive \neq causal.
idea_aspiration_1	Cognitive	+0.001	0.576	+0.002	0.604	Null.
new_ideas	Cognitive	+0.062	0.601	-0.433	0.349	Null.
future_ev_influence	Cognitive	+0.496	0.084	+0.178	0.516	Borderline P1, site-confounded.
fei_resid [†]	Cognitive (within-site)	+0.243	0.358	+0.028	0.636	Null. Borderline was a site artifact.
prob_unexpevents_6M	Beliefs	-0.803	0.215	+0.138	0.607	Null (trend P1).
prob_unexpevents_2Y	Beliefs	-0.725	0.202	-0.364	0.532	Null (trend P1).
int_startup_phase	Venture	+0.005	0.599	-0.034	0.613	Null.
months_working_1	Venture	+0.010	0.314	-0.001	0.691	Null (closest to signal).
exit_probability	Venture	+0.002	0.537	+0.001	0.643	Null.
idea_breadth_h	Idea scope	+0.003	0.505	-0.004	0.499	Null.
planning_future	Venture	-0.149	0.220	+0.021	0.612	Null (trend P1).

$\hat{\beta}_2$ = HTE coefficient in BLP eq. 3.3. Median p-values over 250 splits (Definition 4.2). All 250/250 valid splits per cell.

[†] fei_resid = residual of future_ev_influence \sim site fixed effects ($R^2 = 0.024$; 97.6% of variance is within-site).

Two findings worth noting.

1. *Illusion of Control: RF-predictive but causally null.* IoC is a top-ranked predictor of ΔSI in Random Forest importance across arms and periods—but is causally inert in BLP ($\hat{\beta}_2 = -0.008$, $p = 0.66$). Predictive divergence across arms does not imply causal moderation. This contrast is a methodological contribution to the program-evaluation literature.

2. *Future Event Influence: site artifact, not a moderator.* The only borderline result from the initial scan ($p = 0.084$) disappears when FEI is residualized on site fixed effects ($p = 0.358$). It was between-site variation masquerading as individual heterogeneity.

Portrait of the high-benefit founder (CLAN analysis, consistent across moderators). Despite null BLP results, CLAN systematically identifies the same profile for the top-quintile CATE group: higher hours worked, working full-time, more months on the idea, and—importantly—*lower* perceived probability of unforeseeable events. The founder who most benefits from SDM training is committed, experienced on the idea, and underestimates environmental uncertainty. SDM may work precisely by correcting that underestimation.

What is missing: the cognitive/person dimension.

The family most directly motivated by theory is absent from the current dataset.

Variable	Construct	Why it matters for this paper
risk_aversion unc_aversion (4 items)	+ Risk & ambiguity aversion	Most theory-grounded cognitive moderator (Drexler et al. 2014; Kotha et al. 2023). Founders who avoid uncertainty may gain most from SDM structure.
learning_orientation_1-6 (6 items)	Mastery vs. performance goal orientation	Directly predicts differential absorption of theory-based training content.
competitiveness_1-7 (7 items)	Competitive orientation	Secondary; completes the cognitive portrait of who benefits.

All 17 items collected in the IBL baseline survey. Exist in raw Qualtrics export per RCT site. Status: requested in variables inventory.

Without these variables, the paper characterizes the venture and belief dimensions comprehensively and delivers a clear null finding with strong identification. With them, it tests the person dimension—the “ready to learn” hypothesis—and becomes a complete foundations paper covering both theoretical families.

Why this paper is worth pursuing.

- **The null result is the result.** Across 12 moderators, 2 periods, and 250 splits each, no single observable pre-treatment characteristic significantly drives differential SDM effects. This is clean, well-powered, and informative: SDM is broadly effective, not targeted to specific sub-types.
- **The CLAN portrait is novel.** No prior paper characterizes the high-benefit founder profile using causal ML in this literature.
- **The IoC finding is publishable on its own.** RF importance \neq causal moderation—a point with direct implications for how practitioners use ML for program targeting.
- **One data request closes the main gap.** The 17 cognitive items transform this from a strong partial paper into a complete foundations paper.

Scripts: 02_analysis/scripts/01_initial_analyses/04_genericml.R Data: ERC_RED_HMC_clean.dta ($N = 6,732$ obs, 6 sites, 3 arms) Method: Chernozhukov et al. (2025, *Econometrica* 93(4)) All results labeled EXPLORATORY.